Universidad Carlos III de Madrid

Master on Telematics Engineering
Academic Course 2015-2016

Master Thesis

"Auditing Methodology to Asses the Quality of Online Display Advertising Campaigns"

Author: Patricia Callejo Pinardo
Director: Rubén Cuevas Rumín

Leganés, 6th of September 2016

Keywords: **Online Advertising, Auditing methodology, Fraud detection**

Summary: The reported lack of transparency of the online advertising market may seriously affect the interests of advertisers. In this paper, we present a novel methodology that allows advertisers to independently assess the quality of display advertising campaigns. This methodology also serves to audit the accuracy and completeness of reports delivered by the vendor responsible for running a campaign. We have applied our methodology in 8 display ad campaigns configured in Google AdWords, which overall produced 160K ad impressions displayed in more than 7K publishers. Our results reveal that AdWords seems to provide incomplete information to advertisers. In the case of our campaigns we found that: (i) AdWords did not report 57% of publishers where ad impressions from our campaigns were delivered, (ii) AdWords reports a significantly larger fraction of impressions delivered to contextually meaningful publishers than our auditing methodology (iii) higher CPM investment does not provide guarantees to get impressions delivered to more popular publishers, (iv) AdWords does not offer default control of *frequency cap* (i.e., the number of impressions of the same ad delivered to a user), (v) around 10% ad impressions delivered in two of the ad campaigns were related to fraud. These findings should contribute to open a debate between advertisers and Ad Tech vendors to standardize the utilization of independent auditing methodologies as the one presented in this work.

Auditing Methodology to Asses the Quality of Online Display Advertising Campaigns

Patricia Callejo
Universidad Carlos III de Madrid
IMDEA Networks
patricia.callejo@imdea.org

ABSTRACT

The reported lack of transparency of the online advertising market may seriously affect the interests of advertisers. In this paper, we present a novel methodology that allows advertisers to independently assess the quality of display advertising campaigns. This methodology also serves to audit the accuracy and completeness of reports delivered by the vendor responsible for running a campaign. We have applied our methodology in 8 display ad campaigns configured in Google AdWords, which overall produced 160K ad impressions displayed in more than 7K publishers. Our results reveal that AdWords seems to provide incomplete information to advertisers. In the case of our campaigns we found that: (i) AdWords did not report 57% of publishers where ad impressions from our campaigns were delivered, (ii) AdWords reports a significantly larger fraction of impressions delivered to contextually meaningful publishers than our auditing methodology (iii) higher CPM investment does not provide guarantees to get impressions delivered to more popular publishers, (iv) AdWords does not offer default control of *frequency cap* (i.e., the number of impressions of the same ad delivered to a user), (v) around 10% ad impressions delivered in two of the ad campaigns were related to fraud. These findings should contribute to open a debate between advertisers and Ad Tech vendors to standardize the utilization of independent auditing methodologies as the one presented in this work.

1. INTRODUCTION

Many Ad Tech companies make the argument that online advertisements are an effective form of advertising, and that such advertisements provide a plausible alternative to TV and other forms of traditional advertising. As a result, online advertising attracted a total investment of $125B in 2014 and it is expected to attract $240B in 2019, with an annual growth rate of 12.1% over this period [8]. At the same time, a credible body of evidence supporting the assumption on the alleged effectiveness of online advertising in comparison to other forms of advertising, is largely missing. It may very well be that other forms of media present in major advertisers' media-mix, such as TV, are far more effective than average online advertisements. The three main arguments on behalf of online advertising, that it is more accessible, lower priced per unit, and easily executable at any scale, are also the factors that have contributed to an opaque, and poorly understood fragmented market place. With thousands of vendor companies, helping advertisers place ads on dozens of millions sites, to target over 3 billion Internet users, the online advertising ecosystem is far from transparent. Without transparency, it is not possible to truly establish if online advertising indeed is as effective as a form of advertising as investment in it would suggest.

In particular, the implicit opacity of this market forces advertisers to rely in reports and metrics provided by different vendors such as Ad Networks, Demand Side Platforms (DSPs) or Agency partners to assess the quality of their advertising campaigns. Some recent works have shown that, protected by this opacity, some vendors are providing inaccurate information to advertisers about their advertising campaigns [26]. These findings urge to define methodologies that allow advertisers to independently assess the quality of their online advertising campaigns as well as auditing the reports received from vendors.

In this thesis, we present a lightweight and scalable methodology to audit the performance of display advertising campaigns. In essence, we propose to inject a light *JavaScript* code in a display ad. This code collects relevant information associated to each impression of such ad including the time the ad was exposed to the user, IP address and User-Agent receiving the impression, URL where the impression was shown, user interactions (mouse movements or clicks on the ad), etc. In addition, the *JavaScript* code sends this information to a central server where it is stored.

The processing of this collected information would allow advertisers to evaluate important quality aspects of their campaigns including: (i) the occurrence of potential *Brand Safety* violation episodes, (ii) the popularity and *contextual* relevance of publishers where ad impressions were delivered, (iii) the quality of delivered impressions as measured by de-facto standard metrics such as *viewability* or *frequency cap* and (iv) the exposition of the ad campaign to fraud.

We have tested the proposed methodology in 8 different campaigns set up using Google AdWords. Overall these campaigns delivered around 160K ad impressions across more

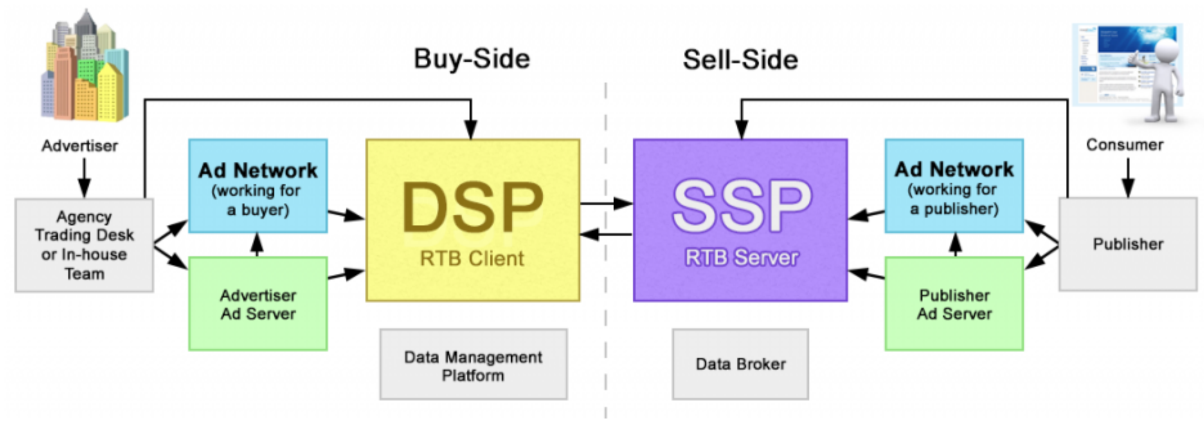

Figure 1: Diagram of the current scenario in the online advertising ecosystem.

than 7K publishers. The obtained results indicate that the information reported by AdWords to advertisers is incomplete. In particular, our auditing methodology reveals the following insights: (i) AdWords did not report 57% of the publishers where ads from our campaigns were delivered. Without the complete list of publishers, an advertiser cannot properly assess whether a *Brand Safety* violation episode occurred; (ii) AdWords reports a significantly larger fraction of ads delivered to publishers whose theme matches the *context* defined by the campaign's keywords than our methodology; (iii) We configure campaigns with Cost-Per-Mille (CPM) investment ranging between 0,01€ and 0,30€ and conclude that, contrary to our expectation, a higher investment does not lead to impressions delivered to more popular publishers; (iv) AdWords does not impose any default *frequency cap*. This leads to hundreds of cases in our campaigns where a user receives the same ad more than 100 times with inter-arrival times between two consecutive ad impressions lower than 1 minute; (v) $\sim$10% impressions are served to IP addresses belonging to Cloud Providers in two of the campaigns. This type of impressions are commonly considered fraudulent.

In summary, this thesis contributes a novel methodology whose application in a real use case provides solid evidences about the inconsistency of reporting from vendors in the online advertising market and how this may affect the interests of advertisers.

2. BACKGROUND

2.1 Overview of online advertising ecosystem

The digital advertising ecosystem is quite complex nowadays. A few years ago if a digital media buyer wanted to run a branded display campaign they would go to an Ad Network that had tags running on a long list of websites. The buyer would usually procure a list of these sites and pay a CPM to run their advertising on these sites. But in the current ecosystem, buyers are purchasing by the impression which means that CPM's have increased but rather than buying bulk page views, advertisers are now able to buy large targeted datasets through programmatic exchanges. Figure 1 show the scenario of the whole online advertising ecosystem. We are going to define the different parts of the ecosystem that form the current scenario of online advertising, in order to understand how the ecosystem works, which it is actually called programmatic advertising:

- Supply Side platform (SSP): Supply Side platforms offer an intermediary service for publishers to help them optimize their ad inventory, i.e., how publishers supply inventory to buyers. This take place through automated systems such as programmatic buying and real-time bidding. They tend to offer technology platforms that help publishers display their inventory to ad buyers, as well as other services such as setting up private marketplaces between an ad buyer and a publisher to create a closed bidding process for premium inventory. *E.g: Rubicon, AdMeld*

- Demand Side platform (DSP): Demand Side platforms work like Supply Side platforms, in this case are used by advertisers who are demanding the ad inventory. They assist to connect ad buyers and agency trading desks to multiple ad exchanges, providing them with aggregated audiences selected from the ad exchanges. *E.g: MediaMath, Efficent Frontier*

- Ad Network: The Ad Networks are advertising companies that aggregate the ad space supply from publishers and matching it with advertiser demand. They sell inventory on a huge scale, offering publishers a ready constant supply of ads via a central ad server *E.g: Google, Videology*

- Ad exchanges: These exchanges enable the bidding and auctioning of inventory. Publishers offer their own inventory to the exchange, making it available to potential advertisers, either direct or through Supply Side platforms or Ad Networks. That inventory is collected in a "central" market, making it easier for advertisers to buy highly-targeted audiences. Exchanges are increasingly becoming the default link between Ad Networks and the rest of the buy-side of the ecosystem as programmatic and real-time bidding technology. *E.g: Doubleclick, RightMedia, Microsoft Media Network.*

- Ad servers: These servers provide the technology that

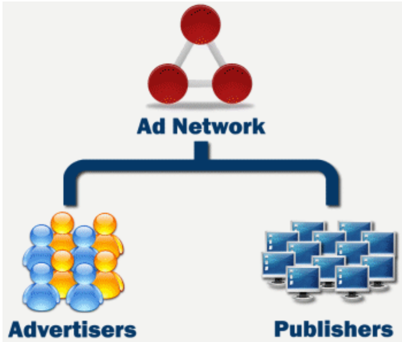

Figure 2: Diagram of the scenario used in this study.

actually delivers an ad to the user who has been targeted. In other words, they place advertisements on the publishers websites. *E.g: Doubleclick, Mediaplex, Bloom*
- Agency Trading Desks: Agency Trading Desks (ATDs) are centralized management platform used by ad agencies specialized in buying online ad inventory through programmatic buying and audience buying. These ATDs are the people who effectively control how an online ad budget is spent if it is being traded through automated systems. They attempt to help clients in order to improve their advertising performance and receive increased value from their display advertising. Moreover, they measure results of their campaigns and report audience insight to their clients. *E.g: Vivaki, Accuen, Xaxis*

All these companies explained above are the most important parts that make up the chain of online advertising between advertiser and publisher, there are also related intermediaries that provide extra services designed to enhance the system but they are less important.

For the shake of clarity, in this study, we consider a simplified model of this ecosystem formed by an advertiser, the publishers and an Ad Network that selects the publishers where the ads from the advertiser are placed. Figure 2 presents this model.

2.2 Assessment of the quality of online advertising campaign

As we have just explained, the current online advertising ecosystem is quite complex [14]. We find several intermediaries (Media Agencies, Trading Desks, Demand Side Platforms -DSPs-, Ad Exchanges, Ad Networks, etc), which interact among them in order to display ads from advertisers into the inventory provided by publishers. These intermediaries jointly operate as a black box that prevents advertisers to independently assess the quality of their advertising campaigns. Instead, advertisers have to trust the reports from their Ad Network, DSP or Agency Partner to perform such assessment. It is expected to encounter some discrepancies in these reports, however, the results in this thesis suggest that the extend to which misreporting takes place may be significant.

Some of the most important aspects to consider in order to assess the quality of advertising campaigns are:
- Brand Safety: One of the “golden rules” for an advertising campaign is to NOT affect the advertiser’s brand reputation. This occurs when the ads of an advertiser are placed in websites with wrong *Context*. For instance an ad of a toys’ brand displayed in a porn website. Hence, in order to evaluate potential episodes of *brand safety* violations, an advertiser must know the complete list of publishers where ads from its campaign were displayed.
- Context: Advertisers are in general interested in displaying their ads in publishers whose theme aligns with the topic of the ad. For instance, a hotel ad is better placed in websites related to holidays or travel agencies than in websites related to job agencies or hospitals. This allows to minimize episodes of *brand safety* violations and also results in more efficient campaigns. Note that recent forms of online advertising, such as Online Behavioural Advertising (OBA) [12, 9, 17], selects the ad to show based on the profile of the user and thus, they may give less relevance to *Context*.
- Publishers’ quality: Although quality could be measured in different ways, in digital marketing, a *publisher’s quality* is typically defined by its popularity. In particular, advertisers pay higher CPM (cost per thousand impressions) and CPC (cost per click) for impressions placed (or clicks occurring) in popular publishers. Indeed, the term *premium inventory* is used to describe inventory from popular websites. Hence, the quality of a campaign is positively impacted by ads placed in popular publishers.
- Impressions’ quality: The de-facto standard metric to monetize an impression is the *viewability* [15]. Based on this metric, an impression is considered to be of good quality (and thus monetized) if the user watches (at least) 50% of it during (at least) 1 second. Another important metric to measure the quality of impressions is the *frequency cap*, which defines a limit for the number of impressions of the same ad that should be shown to the same user in a given period of time. Microsoft Advertising published a study that analyzed the use of different *frequency cap* values across 38 different advertisers [18]. The study concludes that the conversion rate rarely increases for a *frequency cap* beyond 10 impressions. The results demonstrate that usually the conversion rate rarely increase for a frequency cap beyond 10 impressions and they do not even consider cases using a frequency cap higher than 20.
- Fraud indicators: A fraudulent impression is defined as that one that is not visualized by a human or is forced into the web browser of the user by malicious practices. Fraud is one of the major problems in online advertising and it is estimated that it costs the U.S. Ad Tech industry more than $8B annually [16]. Then, indicators of the fraction of impressions of a given campaign associated to fraudulent activity are very relevant to asses the overall quality of a display ad campaign.
- Conversion Ratio: This is the fraction of ad impressions that leads to a conversion. A conversion can be defined as a

product purchase in an e-commerce site, a click in an ad, etc. In particular, an extensively used metric is the Click Through Ratio (CTR), which measures the fraction of ad impressions in a campaign that generated a click.

3. METHODOLOGY

We have designed a methodology focused in HTML5 display ads, which are expected to become the de-facto standard in display advertising [6]. HTML5 allows creating ads using web technologies such as CSS or *JavaScript*. We leverage this opportunity to inject a simple *JavaScript* code into HTML5 display ads. This code collects information associated to the impressions of each ad, including: i) the URL of the webpage where the ad was displayed. Note that the domain part of the URL reveals the publisher; ii) the IP address[1] and the User-Agent receiving the ad impression; iii) the timestamp of the ad impression; iv) user interactions with the ad. In particular, we collect mouse movements over the ad as well as click events. All this information is collected locally by the *JavaScript* code and transferred to a central server where it is properly stored in a database. Note that in addition to the previous metrics we also compute the *exposition time* of an ad impression as the time difference between the establishment and the closure of the connection, measured at the central server.

We implement the described methodology employing widely used and lightweight technologies to guarantee efficiency, scalability and robustness. In particular, we use: (i) *plain JavaScript* for the code inserted in the ad; (ii) *the WebSocket protocol* [22] for transferring the information from the ad impression to the central server. Note that the information is transferred in the form of a string; (iii) *Node.js JavaScript library* [11] to parse and process the information received in the central server; (iv) *MySQL and Python* to store and process the collected datasets.

We notice that similar methodologies, using code inserted in Flash display ads, have been used in research to perform network measurements experiments [30, 31].

3.1 Limitations and Validation

We have tested our methodology in a lab controlled environment and confirmed its capacity to retrieve all the data described above. However, our methodology is expected to run in operational network environments and thus it is subject to different errors. Then, we cannot guarantee to retrieve information from every ad impression. Errors happening in the browser, the network, our server or in the connection establishment process would result in the affected ad impression(s) not being logged in our central server.

In addition, we have validated that the delayed error expected in the measurement of the exposition time ranges in the order of 159 ms (median), which guarantees a sufficient accuracy for measuring the upper bound of the viewability metric.

Finally, we have evaluated the scalability of our server-side infrastructure. For that purpose, we have calculated the load that the system can handle in *WebSocket* connections. In particular, the conducted experiments confirm that an instance of our server in a Virtual Machine with 2 VCPUs and 4GB RAM is able to correctly receive and process up to 1k request per second. Note that this include open-close the *WebSocket* connection and process-storage the data in the database. This confirm the good scalability properties of our solution.

The described methodology is directly applicable in ad formats that support *JavaScript* in a native manner, such as HTML5 ads. In other ad formats, such as images or video, this methodology would only work if the Ad Network allows to add a tracking pixel. This is a quite extended practice for conversion and user tracking. If this option is available, and it allows JavaScript code, we could add our code inside the tracking pixel.

Moreover, there exists a widely extended security policy referred to as *Same-Origin* policy [13]. Following this policy a code running as part of an iFrame cannot track the activity in other parts of the webpage different from this iFrame. Since most Ad Networks insert ad impressions in a single or a double iFrame, in general, this policy would prevent our methodology from collecting some information such as the referrer (i.e., the previous publisher from which the user reached the current one) or events (e.g., mouse movements) happening in other parts of the webpage different than the iFrame enclosing our ad. It also prevents us from collecting information about the position of the iFrame in the webpage. Then, we cannot assess if the ad (or part of it) was shown in the visible part of the screen. Therefore, our methodology can only measure an upper bound of the *viewability* metric presented in Section 2. This is, whether the ad was displayed more than 1 sec, but without knowing if (at least) 50% of it was shown.

4. REAL USECASE

4.1 Ad Network and Datasets

We have applied our auditing methodology to campaigns configured in Google AdWords[2], which uses Google Display Network (GDN) to deliver display ads. We have selected this Ad Network due to the following two reasons: First, GDN is one of the most important Ad Networks worldwide. It spans over 2 million publishers that reach over 90% of Internet users [2]; Second, GDN allows to run low budget campaigns, starting at few dollars. Then, using AdWords/GDN, we can test our methodology while respecting our budget restrictions. Although our intention was testing our methodology in other Ad Networks, we could not do it

[1] Note that we use the IP address to extract meta-data information such as the user's provider. Afterwards, we anonymize the IP using hashing techniques.

[2] We have reported the obtained results to Google and we will incorporate their feedback into future versions of this thesis.

Campaign ID	# Impressions	# Publishers	Start date	End date	CPM	Targeted Keywords	Targeted Location
Research-010	5117	350	29 March	31 March	0.10 €	Research	Spain
Research-020	42399	1777	29 March	31 March	0.20 €	Research	Spain
Football-010	33730	1086	02 April	03 April	0.10 €	Football	Spain
Football-030	24461	1367	02 April	03 April	0.30 €	Football	Spain
Russia	4096	274	29 March	31 March	0.01 €	Research	Russia
USA	1178	136	29 March	31 March	0.01 €	Research	United States
General-005	8810	580	15 February	23 February	0.05 €	Universities, Research, Telematics	Spain
General-010	42357	1549	18 February	23 February	0.10 €	Universities, Research, Telematics	Spain

Table 1: Description of the 8 AdWords campaigns used to test our auditing methodology.

because those ones meaningful for our study requested an initial investment in the order of few thousands dollars before running any campaign. Unfortunately, such requirement exceeded our budget capacity.

To test our methodology we have run 8 different display advertising campaigns using Google AdWords. Overall we registered around 160K ad impressions distributed in approximately 7K publishers. Note that we set-up campaigns with different duration, different CPM values as well as different targeted keywords and geographical locations. This diversity aims at providing guarantees that the observed results are not due to a specific campaign set-up. Table 1 summarizes the main properties of each campaign.

4.2 Results

In this subsection, we prove the validity of our methodology to first, perform the quality assessment for our eight display ad campaigns and, second, audit the ad campaign reports from AdWords. To this end we study the different quality aspects presented in Section 2: *Brand Safety*, *Context*, *Quality of Publishers*, *Quality of Impressions* and *Fraud Indicators*. Our campaigns were configured based on CPM and then we did not obtain a sufficiently large number of clicks to evaluate the *Click Through Ratio*. We leave this for future work.

Note that the results presented in the rest of this section, excepting for the case of *Brand Safety* and *Context*, are obtained from the analysis of our datasets without considering the information available in AdWords reports.

4.2.1 Brand Safety

As discussed in Section 2, an advertiser must know every publisher where ad impressions are displayed in order to properly assess potential *Brand Safety* violations. For each one of the 8 ad campaigns, we have compared the list of publishers where ad impressions were displayed as reported by our methodology vs. AdWords. Figure 3 shows a Venn diagram representing the total number of publishers exclusively reported by AdWords (in green), exclusively reported by our methodology (in red) and those reported by both (in yellow). In particular, the figure presents results for a specific campaign (*General-005*) as well as the aggregate results across all campaigns. The aggregate results reveal that AdWords did not report 57% of the publishers where ads

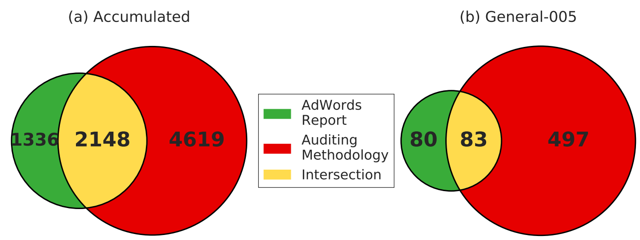

Figure 3: Venn diagram showing the number of publishers exclusively reported by our auditing methodology (red), exclusively reported by AdWords (green) and reported by both (yellow) for all our campaigns and campaign *General-005*.

from our campaigns were delivered[3]. This number can increase for individual campaigns up to 75%, as in the case of *General-005*.

Part of the impressions reported by AdWords are associated to *"anonymous.google"*. They correspond to impressions served through Google Ad Exchange to publishers that want to preserve their anonymity[4] [5]. Then, someone may argue that the not reported publishers correspond to those associated to *"anonymous.google"*. Our results show that this is an invalid argument. For instance, in *General-005*, AdWords registers only 425 impressions whose associated publisher is labelled as *"anonymous.google"*, however, 497 publishers identified by our methodology were not reported by AdWords. Then, even if these 425 impressions had been distributed across 425 publishers, still 72 publishers had not been reported by AdWords, in this specific campaign.

Moreover, the fact that these publishers are not reported by AdWords means that their associated impressions have not been charged to the advertiser. This may be, for instance, because these impressions did not fulfil the *viewability* metric requirements. Even if these impressions have not ever been viewed by a user (e.g., they were displayed in the non-visible part of the screen) the potential risk for *Brand Safety* remains. The Ad Network may have displayed ads in a potentially harmful publisher in the past and there are no means to avoid the ad network doing it again in the future before

[3] Note that our methodology did not log 16.5% of the publishers.

[4] Note that advertisers can configure their campaigns to exclude anonymous publishers [7].

Campaign ID	Auditing Methodology (% impressions)	AdWords Report (% impressions)
Research-010	2.50%	2.66 %
Research-020	3.75%	3.05 %
Football-010	64.12%	100 %
Football-030	46.66%	100 %
Russia	4.10%	7 %
USA	6.28%	10.73 %
General-005	4.96%	7.36 %
General-010	6.63%	56.65 %

Table 2: Fraction of impressions delivered to contextually meaningful publishers as reported by AdWords vs. our auditing methodology.

an actual visualization leading to a *Brand Safety* violation may occur. Instead, if the advertiser would have access to the complete list of publishers, it could identify those potentially harmful and blacklist them avoiding possible *Brand Safety* violation episodes in the future. Finally, note that we have reviewed the Terms of Service from AdWords as well as the AdWords and related Google support pages and the references to not reported publishers, which we have found, relate exclusively to *"anonymous.google"*.

4.2.2 Context

AdWords support guidelines indicate that campaigns configured based on keywords, as it is the case of our campaigns, would follow a *contextual* strategy. This is, AdWords tries to display ads in publishers whose theme is related to the targeted keyword(s), and thus contextually meaningful for the campaign. In addition, AdWords may use other factors to determine if a publisher is contextually relevant to the campaign such as the recent browsing history of a user [1]. We have leveraged our auditing methodology to assess whether the context of a publisher is relevant to the keywords defined for a given campaign. In particular, we have extracted the keywords and topics that AdWords assigns to each publisher where any of the logged impressions by our methodology has been displayed. Then, we consider a publisher contextually meaningful if 1) any of its keywords matches any of the campaign's keywords or 2) any of the publisher's topics is semantically similar to any of the keywords of the campaign. For this purpose we use the Leacock-Chodorow semantic similarity as described in [17].

Table 2 shows the fraction of impressions delivered to contextually meaningful publishers, as reported by AdWords vs. our auditing methodology, for our 8 campaigns. AdWords reports a notably higher rate of ads delivered to contextually meaningful publishers compared to our methodology in most campaigns. This difference may be explained by impressions delivered by AdWords using different factors than the publisher's theme (e.g., previous browsing history of the user). Unfortunately, assessing the contextual appropriateness of such impressions is not possible because either they are due to factors not revealed by Google or would require to know the recent browsing history of the user.

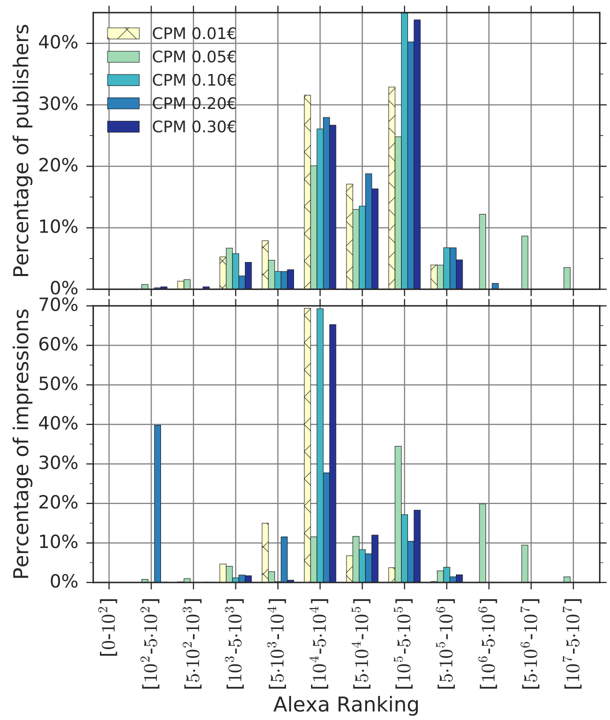

Figure 4: Distribution of publishers (top) and ad impressions (down) across the Alexa Ranking for 5 campaigns configured with different CPM investment.

4.2.3 Quality of Publishers

As indicated in Section 2, the quality of publishers is typically associated to its popularity, and thus, campaigns configured with a higher CPM are expected to get their ads displayed in more popular publishers. We measure the popularity of a publisher using its position in the Alexa ranking [3], which is a common method to assess the popularity of a publisher in both academia and industry.

Figure 4 shows the distribution of publishers and impressions across the Alexa ranking for 5 of our campaigns with CPMs ranging between 0,01€ and 0,30€. Specifically, we have defined logarithmic buckets and computed the fraction of publishers and impressions that fall in each bucket for each campaign. The results indicate that, contrary to our expectation, higher CPMs do not lead to increase the fraction of popular publishers or impression in such publishers. Indeed, the campaign with a CPM equal to 0,01€ seems to be the one of those achieving better performance with roughly 46% publishers and 89% impressions accumulated in the Alexa Top 50K sites. Instead, the campaign configured with a CPM of 0,30€ , representing a 30× investment increase, presents just 35% publishers and 68% impressions in the Alexa Top 50K. This observation may lead advertisers to consider whether higher investments are really necessary.

4.2.4 Quality of Impressions

In this section we evaluate the quality of impressions of our 8 campaigns using the two metrics described in Section

Campaign ID	View $\geq 1s$
Research-010	56.18 %
Research-020	52.21 %
Football-010	79.89 %
Football-030	82.80 %
Russia	62.69 %
USA	71.13 %
General-005	75.13 %
General-010	55.03 %

Table 3: Upper bound *viewability* (i.e., fraction of ad impressions displayed $\geq$ 1s) for each ad campaign.

2, *viewability* and *frequency cap*.

Viewability: Table 3 presents the fraction of impressions that fulfils the upper bound of the *viewability* that we can measure with our methodology. The values range between 52% and 85% across campaigns. Interestingly, the two campaigns presenting the highest fraction of "viewable" impressions are the ones targeting "football", whereas other campaigns targeting other keywords (e.g., research) achieve a significant lower viewability. We conjecture that the targeted audience is an important factor that modulates ads *viewablity*.

Frequency Cap: Our goal in this case is to assess whether AdWords implements any default control in the *frequency cap*. Note that AdWords is used by a large number of customers without expertise in digital marketing, which may not configure a *frequency cap* in their campaigns. Therefore, it would be desirable that AdWords (or any other ad network) defines a default *frequency cap* on behalf of their customers. Research studies in the literature [18] have shown that a *frequency cap* over 10 does not lead to better conversion ratios. Based on this, 10 seems to be a good reference value.

Figure 5 presents a scatter plot in loglog scale where the x-axis shows the number of impressions of a specific ad delivered to a user and the y-axis represents the median inter-arrival time between two consecutive impressions of that ad shown to the user. The figure presents aggregate results for all our campaigns. Note that we define a user as the combination of IP and User-Agent, so that two users behind a NAT using different User-Agents will be considered separately.

The results indicate that AdWords does not seem to use any default *frequency cap*. Indeed, 1720 (176) users receive more than 10 (100) impressions from the same ad. In addition, we observe that in many of these cases the inter-arrival time between impressions is rather small (below 1 min). In particular, there are extreme cases in which users receive hundreds of impressions with an inter-arrival time below 20 seconds. These observations suggest that unskilled or careless advertisers may experience inefficiencies in their campaigns performance due to the absence of a reasonable *frequency cap*.

Finally, note that AdWords does not report information about users and thus in order to assess *viewability* or *frequency cap* aspects our methodology becomes essential.

4.2.5 Fraud Identification

Fraud is one of the main problems in online advertising

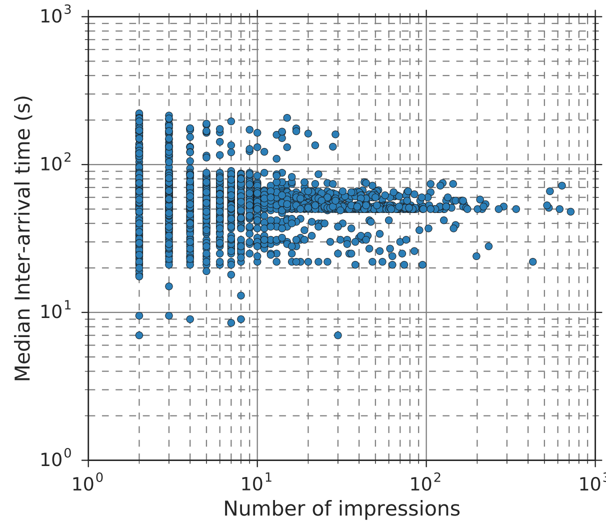

Figure 5: Number of ad impressions of a specific ad Vs. median inter-arrival time between impressions for each user considering all our campaigns.

leading to annual losses over $8B to US Ad Tech industry [8]. Indeed, identifying, preventing and mitigating fraud is a complex and still unsolved problem which has attracted the attention of the research community [19, 20, 21, 26, 29, 32]. In this subsection we show an example of how our auditing methodology can be used to identify some type of ad fraud.

A common technique used by fraudsters consists in installing a bot in a server hosted in a Cloud Provider[5]. This bot visits websites owned by the fraudster, receiving ads and eventually clicking in some of them. This activity reports benefits to the fraudster based on CPM and CPC monetization schemes. Then, it is commonly accepted in the Ad Tech industry that views (or clicks) coming from IPs hosted in Cloud Providers are fraudulent.

Our methodology collects the IP addresses receiving ad impressions from a given campaign. Then, we can identify which of the collected IPs belong to Cloud Providers. In particular, we use the following methodology for this purpose: First, we use MaxMind service [10] to map each IP address in our dataset to its associated Provider. Second, we leverage a Botlab's public list [4] including more than 130M IPs from the top 100 Cloud Providers worldwide to identify the IPs in our dataset present on it. Finally, for the remaining IPs, we have manually verified the website of its associated Provider to assess whether it was a Cloud Provider or not.

Table 4 presents the results of applying the previous methodology in each of our campaigns. Specifically, it shows: (i) the fraction of IPs located in Cloud Providers, (ii) the portion of ad impressions delivered to those IPs and, (iii) the fraction of publishers that served impressions to those IPs. We observe that none of our campaigns is fraud-free and the fraction of impressions and exposed publishers is non-

[5]Note that for simplicity we use the term Cloud Providers to refer to both Cloud and Hosting providers.

Campaign ID	% of Cloud Providers IPs	% of Impressions delivered to Cloud IPs	% of Publishers showing ads to Cloud IPs
Research-010	3.39 %	4.42 %	8.62 %
Research-020	2.36 %	2.88 %	8.73%
Football-010	7.61 %	8.6 %	23.55%
Football-030	11.08 %	10.95 %	23.13%
Russia	0.52 %	0.27 %	2.58%
USA	1.03 %	0.68 %	5.56%
General-005	0.54 %	0.55 %	3.94%
General-010	0.42 %	0.58 %	2.59%

Table 4: Statistics on the volume of fraud activity from Cloud Providers IPs in each campaign.

negligible in some cases. Specifically, "Football" campaigns present roughly 10% of the impressions delivered to Cloud Providers' IPs and 23% of publishers are exposed to such impressions. For these particular campaigns we have verified that AdWords initially charged us for more than 1K impressions delivered to Cloud Providers' IPs. Later, we got a refund from AdWords. However, AdWords did not give details on the reasons for such refund and then we cannot assess if the previous impressions were part of it.

Finally, note that AdWords do not provide detailed information about the impressions and publishers exposed to fraud, and thus an advertiser cannot assess its exposition to the analyzed type of fraud without a methodology like ours.

5. RELATED WORK

The research community has contributed in studying the online advertising ecosystem in order to make the market more trusted. Stone-Gross et al. [29] describe how online ad exchanges work and prove that the complexity of the ad exchanges give the fraudulent users or fraudsters an opportunity to generate revenue by developing malware that impersonates user activities.

In terms of quality assessment of online advertising campaign, most of the previous works focus on the fraud problem, specifically as click fraud since the fraudulent activity is associated to fake clicks on ads. Metwally et al. [27] present an early study in which they use the IP address as the parameter to detect coalition of fraudulent users or fraudsters. In a more recent work, Li et al. [25] propose to analyze the paths of ad's redirects and the nodes found in the content delivery path to identify malicious advertisement activities. Moreover, Neal et al. [28] measure several advertising platforms and present the ratio of the Ad-Clicks obtained considered human beings against those considered ad-bot or automated computer programs. Haddadi [24] presents a type of ads for fight against online click fraud, called Bluff Ads which are a set of ads that are designed to be detected and clicked only by machines, or poorly trained click-fraud work force.

Related with the Click Traffic, Zhang et al. [32] evaluate the quality of purchased traffic, measuring mouse movements, timing properties, correlation with blacklists, which shows evidence of fake traffic in the behaviours analyzed.

Furthermore, Marciel et al. [26] present a recent work that shows the evidences of fake views in video content portals like Youtube. In the same field of video portals, Chen et al. [19] present a study focused on the detection of fake views caused by robots using a technique that combines the IP entropy and video entropy.

In the field on Online Behavioural Advertising Guha et al. [23] present measurements for targeted advertising, including metrics to the high levels of noise inherent in ad distribution, network effects like load-balancing and timing effects. This methodology applies chiefly to text-ads and ads on online social networking pages. Following this topic, there is a more recent work by Carrascosa et al. [17] who present a large analysis of users profiles that shows the influence of ad targeting based on the user profiles.

Our work is different in focus to these previous works since we are worried about the lack of transparency of the online advertising market, which affects the interest of advertisers. To the best of the authors knowledge, there is not previous work addressing, in a holistic manner, the assessment of quality in online advertising campaigns. As we have seen, most existing works focus on a single angel which is ad fraud. However, as we have shown in this study there are some other aspects to consider in addition to fraud.

6. CONCLUSION

This thesis introduces a novel auditing methodology that proposes to inject a very light javascript code in the ads to monitor the performance of advertisers online campaigns through several quality indicators.The main contributions of the methodology are the following: (i) Allow advertisers knowing all the websites where their ads were displayed in order to guarantee their brand safety. (ii) Measure the context quality of ad campaigns by evaluating the semantic similarity between the target keywords of the campaign and the keywords (or tags) associated to the publishers where the ads are displayed. (iii) Evaluate the quality of the publishers reached in our campaigns using their popularity (i.e., alexa ranking) as a reference of their quality. (iv) Measure the quality of the impressions in terms of viewability (for any selected criterion) and frequency-cap. (v) Create blacklists of IP addresses that may be used by fraudsters to monetize fake visits to websites.

Overall, this thesis illustrates the lack of transparency and accurate information that advertisers are suffering from in the current online advertising ecosystem. This avoids advertisers to accurately assess the efficiency and quality of their online campaigns, and even more important, they lack the required information to take decisions and actions to protect their *Brand Safety*. These results should encourage advertisers to request the Ad Tech industry to standardise the use of independent measurements methodologies, as the one presented in this work, which would allow them to independently assess the quality of their online advertising campaigns as well as auditing the reporting activity of different vendors such as Ad Networks and DSPs.

7. REFERENCES

[1] About contextual targeting. Google Support. https://support.google.com/adwords/answer/2404186. (Date last accessed 25-August-2016).

[2] About the Google Display Network. Google Support. https://support.google.com/adwords/answer/2404190?hl=en. (Date last accessed 25-August-2016).

[3] Alexa Ranking. http://www.alexa.com/topsites. (Date last accessed 25-August-2016).

[4] Botlab.io Deny-hosting IP List. https://github.com/botlabio/deny-hosting-IP. (Date last accessed 25-August-2016).

[5] Differences between Ad Exchange and AdSense. Google Support. https://support.google.com/adxseller/answer/4599464?hl=en. (Date last accessed 25-August-2016).

[6] Display and Mobile Advertising Creative Format Guidelines. IAB, 2015. http://www.iab.com/wp-content/uploads/2015/11/IAB_Display_Mobile_Creative_Guidelines_HTML5_2015.pdf. (Date last accessed 25-August-2016).

[7] Exclude Anonymous sites. Google Support. https://support.google.com/adxbuyer/answer/159152?hl=en. (Date last accessed 25-August-2016).

[8] Global entertainment and media outlook 2015-2019. PwC, Ovum. http://www.pwc.com/gx/en/global-entertainment-media-outlook/assets/2015/internet-advertising-key-insights-1-advertising-segment.pdf. (Date last accessed 25-August-2016).

[9] IAB Europe EU Framework for Online Behavioural Advertising. IAB Europe. http://www.iabeurope.eu/files/9613/6984/1480/2012-12-11_iab_europe_oba_framework.pdf. (Date last accessed 25-August-2016).

[10] MaxMind GeoIP Legacy ISP Database. https://www.maxmind.com/. (Date last accessed 25-August-2016).

[11] Node.js. https://nodejs.org/. (Date last accessed 25-August-2016).

[12] Online Behavioural Advertising. Advertising Standars Authority UK. https://www.asa.org.uk/Consumers/What-we-cover/Online-behavioral-advertising.aspx. (Date last accessed 25-August-2016).

[13] Same-Origin Policy. https://developer.mozilla.org/en-US/docs/Web/Security/Same-origin_policy. (Date last accessed 25-August-2016).

[14] The Online Advertising Ecosystem Explained. Digital Ad Blog. http://digitaladblog.com/2015/02/19/online-advertising-ecosystem-explained/. (Date last accessed 25-August-2016).

[15] Viewable Ad Impression Measurement Guidelines. Media Rating Council and IAB. http://www.mediaratingcouncil.org/063014%20Viewable%20Ad%20Impression%20Guideline_Final.pdf. (Date last accessed 25-August-2016).

[16] What Is An Untrustworthy Supply Chain Costing The U.S. Digital Advertising Industry? . IAB. http://www.iab.com/insights/what-is-an-untrustworthy-supply-chain-costing-the-u-s-digital-advertising-industry/. (Date last accessed 25-August-2016).

[17] J. M. Carrascosa, J. Mikians, R. Cuevas, V. Erramilli, and N. Laoutaris. I Always Feel Like Somebodys Watching Me. Measuring Online Behavioural Advertising. In *Proceedings of the 11th ACM International Conference on emerging Networking EXperiments and Technologies*, CoNEXT'15, 2015.

[18] J. Chandler-Pepelnjak and Y.-B. Song. Optimal Frequency: The impact of frequency on conversion rates. Microsoft Advertising Institute, 2009. https://advertising.microsoft.com/wwdocs/user/en-us/researchlibrary/researchreport/OptimalFrequency.pdf. (Date last accessed 25-August-2016).

[19] L. Chen, Y. Zhou, and D. M. Chiu. Analysis and Detection of Fake Views in Online Video Services. *ACM Transactions on Multimedia Computing Communications and Applications (TOMM)*, 2015.

[20] V. Dave, S. Guha, and Y. Zhang. Measuring and Fingerprinting Click-spam in Ad Networks. In *Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '12, 2012.

[21] V. Dave, S. Guha, and Y. Zhang. ViceROI: Catching Click-spam in Search Ad Networks. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer Communications Security*, CCS '13, 2013.

[22] I. Fette and A. Melnikov. The websocket protocol. RFC 6455, RFC Editor, December 2011. http://www.rfc-editor.org/rfc/rfc6455.txt.

[23] S. Guha, B. Cheng, and P. Francis. Challenges in measuring online advertising systems. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pages 81–87. ACM, 2010.

[24] H. Haddadi. Fighting online click-fraud using bluff ads. *ACM SIGCOMM Computer Communication Review*, 40(2):21–25, 2010.

[25] Z. Li, K. Zhang, Y. Xie, F. Yu, and X. Wang. Knowing your enemy: understanding and detecting malicious web advertising. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 674–686. ACM, 2012.

[26] M. Marciel, R. Cuevas, A. Banchs, R. González, S. Traverso, M. Ahmed, and A. Azcorra. Understanding the Detection of View Fraud in Video Content Portals. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, 2016.

[27] A. Metwally, D. Agrawal, and A. El Abbadi. Detectives: detecting coalition hit inflation attacks in advertising networks streams. In *Proceedings of the 16th international conference on World Wide Web*, pages 241–250. ACM, 2007.

[28] A. Neal, S. Kouwenhoven, and O. B. SA. Quantifying online advertising fraud: Ad-click bots vs humans. Technical report, tech. rep., Oxford Bio Chronometrics, 2015.

[29] B. Stone-Gross, R. Stevens, A. Zarras, R. Kemmerer, C. Kruegel, and G. Vigna. Understanding Fraudulent Activities in Online Ad Exchanges. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, 2011.

[30] S. Zander, L. L. Andrew, G. Armitage, G. Huston, and G. Michaelson. Investigating the IPv6 Teredo Tunnelling Capability and Performance of Internet Clients. *ACM SIGCOMM Computer Communications Review (CCR)*, 2012.

[31] S. Zander, L. L. Andrew, G. Armitage, G. Huston, and G. Michaelson. Mitigating Sampling Error when Measuring Internet Client IPv6 Capabilities. In *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, IMC '12, 2012.

[32] Q. Zhang, T. Ristenpart, S. Savage, and G. M. Voelker. Got Traffic?: An Evaluation of Click Traffic Providers. In *Proceedings of the 2011 Joint WICOW/AIRWeb Workshop on Web Quality*, WebQuality '11, 2011.