Ad Blockers: Global Prevalence and Impact

Matthew Malloy[+], Mark McNamara[+], Aaron Cahn[+], Paul Barford[+†]
{mmalloy,mmcnamara,acahn,pbarford}@comscore.com
[+]comScore, Inc. [†]University of Wisconsin - Madison

ABSTRACT

Ad blockers are a formidable threat to the vitality of the online advertising eco-system. Understanding their prevalence and impact is challenging due to the massive scale and diversity of the eco-system. In this paper, we utilize unique data gathering assets to assess the prevalence and impact of ad blockers from an Internet-wide perspective. Our study is based on *(i)* a 2 million person world-wide user panel that provides ground truth for ad blocker installations and *(ii)* telemetry from large number of publisher web pages and ads served to publishers. We describe a novel method for assessing the prevalence of ad blocker installations that is based on Mixture Proportion Estimation. We apply this method to nearly 2 trillion web transactions collected over the period of 1 month (February 2016), to derive ad blocker prevalence estimates for desktop systems in diverse geographic areas and for diverse demographic groups. Next, using deployment estimates we consider the impact of ad blockers on users and on publisher sites. Specifically, we report on the reduction of ads shown to users with ad blockers installed and show that even though a user may have an ad blocker installed, they are still exposed to a significant number of ads. We also characterize the impact of ad blockers across different categories of publisher sites including those that may be participating in whitelisting [9].

1. INTRODUCTION

An *ad blocker*, as the name suggests, offers the capability to prevent ads from being delivered to a user's browser. The stated intent of entities that have developed ad blockers is to enable users to surf the web without annoying ads. While the definition of "annoying" is somewhat unclear, what is clear is that these capabilities pose a significant threat to the digital media ecosystem, which has delivered a wide range of transformative services that have been funded by online ads.

Ad blockers are typically implemented as plugins or browser extensions that when installed, attempt to intercept and eliminate outgoing ad requests from a base web page [1]. They use a variety of mechanisms to identify ad requests. One of the most common and effective mechanisms is to compare the URLs in the embedded requests to a blacklist(s) of URLs of ad servers and advertising platforms. If there is a match, the blocker will prevent the request from being transmitted. While a number of the most popular ad blockers are open source and free to users, authors of these systems are now monetizing their efforts by offering to whitelist certain advertisers and publishers [9].

While ad blockers have been available for over a decade, they have been receiving significant attention in the popular media over the past year. Given this attention, we seek answers to several simple questions: what is the prevalence of ad blocker installs in the internet? what is the behavior of ad blockers when installed? and what is the impact to publishers? Answers to these questions will help to clarify the broader conversation about ad blockers and inform the digital media ecosystem about how it can evolve.

There are two significant challenges in addressing questions that focus on internet-wide population estimation. First, it is difficult to assemble data sets on browser configurations that enable such estimates. Second, since it is impossible to have ground truth for *all* browser configurations deployed on all user systems, estimates based on a smaller population of users is required. Care must be taken in any such estimate to remove bias from the sample population.

In this paper we report results of our study of the prevalence and impact of ad blockers in the internet. Our analyses is based on unique data assets that enable us to assess ad blocker deployment and impact in a comprehensive fashion that goes well beyond standard reports on blocker downloads (*e.g.,* Ad Block Plus [5]). The data assets include *census* information from nearly 2 trillion web transactions collected over a period of a month. This is complemented by data collected over the same period of time from a 2M person user panel (distributed across the internet) that ex-

IMC 2016, November 14-16, 2016, Santa Monica, CA, USA

DOI: http://dx.doi.org/10.1145/2987443.2987460

[1]While the focus of this paper is the web, VPN-based ad blockers are also available for mobile apps.

poses all details of browser configurations. This *panel* data, with ground truth information about ad blocker installs, is the starting point for our study.

We posit that users who opt-in to participate in a worldwide user panel that is specifically designed to track web browsing and ad consumption are less likely to install and enable an ad blocker. We develop a novel technique based on Mixture Proportion Estimation (MPE) to quantify this effect. In particular, MPE enables the proportions of subpopulations in a mixture to be estimated from samples. Our technique is based on the assumption that users with ad blockers will see fewer ads than those without ad blockers. We begin by separating panelists into two groups – those with and those without ad blockers – and then quantify the ad display rates for those groups. The ad display rates allow us to infer the prevalence of ad blocking in the census data. Theoretical guarantees in the form of confidence ranges and statistical significance levels are provided. The MPE technique employed in this paper enables, to the best of our knowledge, the largest scale assessment of ad blocker deployment to date.

2. DATA

In this section we provide an overview of the two sources of data used in this study.

Panel. The comScore panel consists of 2 million users worldwide who voluntarily install monitoring software in exchange for various benefits, including cloud storage, anti-virus software, tree planting, and other cash prizes [2]. The panel provides measurements on panelist's web browsing behavior and internet use. When a panelist registers, they voluntary provide their geographic location and demographic information including age, sex, household income, etc.

The panel monitoring software is also able to observe installations of software on panelist's computer. This enables enumeration of web browser configurations including whether or not ad block software packages are installed. We do this by using search queries to build a list of popular ad blockers (*e.g.,* Ad Block Plus, Ad Block, etc.) for three major browsers (*e.g.,* Internet Explorer, Google Chrome, and Mozilla Firefox) and then search for these names in the configuration data. The current lists include 10 ad blockers for Internet Explorer, 20 ad blockers for Google Chrome, and 15 ad blockers for Mozilla Firefox. While we make no claims that each list is complete, we argue that each include most if not all of the widely reported blockers (which we quantify in Section 4). We refer to the subset of panel users that have an ad blocker installed as the *panel percent ad block*.

As with any widely recruited panel, bias can be introduced when some populations find the incentivized recruitment more attractive than others. This is true of the comScore panel and the populations of internet users that do/don't use ad block software. Our intuition is that an individual who voluntarily installs panel software on their machine is likely to be less concerned with privacy – a trait that could correlate with ad block installation. As we show, the panel is indeed biased *away from ad block usage*. One of the key aspects of our work is correcting this bias using the broader census via MPE. For the purposes of estimating ad block usage in the general internet population, the panel provides a data set consisting of tuples of *(i)* a comScore browser cookie and *(ii)* a binary label indicating the presence or absence of ad blocking software. Details of the MPE approach are provided in Section 3.

Census. The comScore census network is one of the most widely deployed internet census networks in the world. The census network collects information daily on over 20 billion page views across half a million top level domains. In addition, the census network collects data on over 2 billion ad deliveries daily. This data is collected via JavaScript tags deployed on publisher pages and JavaScript tags deployed with advertisements. In both settings, a client machine executes the tag locally and reports information directly to our data warehouse. The information includes a cookie identifier, a timestamp, and the type of tag (*e.g.,* page or ad).

Estimation of the prevalence of ad block usage across the census network relies on counts of page and ad tags associated with individual cookies over the course of a month. In particular, for each cookie present on the census network, the count of page views and the count of ad deliveries is tallied over the reporting period. The data gathered from the census network are tuples consisting of *(i)* a comScore browser cookie *(ii)* the count of tagged ads delivered to that browser cookie and *(iii)* the count of tagged page views delivered to that browser cookie. The data set under consideration is reduced to ensure a longitudinal view of each cookie; specifically, a cookie is excluded from the study if *(i)* it has fewer than 200 pageviews or *(ii)* has not existed on the census network for a minimum of 30 days.

Of the remaining cookies, a subset correspond to comScore panelists. This subset of cookies can be labeled as associated or not associated with ad block software by comparing against the panel data set described above. Ultimately this defines three disjoint populations of cookies: *(i)* the set of labeled cookies associated with ad block software, denoted $\mathcal{S}_{\mathrm{block}}$, *(ii)* the set of labeled cookies known to not be associated with ad block software, denoted $\mathcal{S}_{\mathrm{ads}}$, and *(iii)* the set unlabeled cookies, denoted $\mathcal{S}$.

Ad Ratio Statistic. The ad ratio statistic, defined as the number of ads delivered divided by the number of page views, is computed on a cookie by cookie basis:

$$\textit{ad ratio} = \frac{\text{count of ad deliveries}}{\text{count of pageviews}} \tag{1}$$

This statistic, *in aggregate*, is closely tied to ad block usage since it acts as an estimate of the number of ads delivered to a user per unit of internet browsing. However, this statistic alone is not sufficient to infer if a user has or does not have an ad blocker installed. First, in some cases users may browse pages that do not deliver ads, and will have an ad statistic equal to zero regardless of whether or not they have ad block installed. Second, it is often the case that users with an ad

[2]The panel is 100% opt-in with thorough disclosure on data privacy. The privacy policy can be found at http://www.comscore.com/About-comScore/Privacy-Policy.

blocker installed, either through disabling of the ad blocker or whitelisting of ads [9], are shown some number of ads and possibly many ads.

Ultimately the inherent restrictions of the data on hand gives rise to two challenges: *(i)* the panel is biased away from ad block usage and *(ii)* census data is insufficient to reliably classify individual cookies as associated with ad blocking software. We address both of these challenges via the MPE approach described below.

Cookies and Users. There are a number of important details and nuances pertaining to the precise definition of an ad block user, and the relationship between users and cookies. While the panel allows association of a user with ad blocking software, it does not imply that the user actually employed ad blocking software for the entire reporting period. Instead, it simply indicates the presence of ad blocking software at some point during the reporting period. A user may install ad block software and immediately disable it entirely or more commonly install ad block software and disable it on a subset of sites; in both cases the user would still be considered an *ad block user*.

Definition - *Ad block user*: a person with one or more ad blocking programs installed on their primary computer at some point during the reporting period.

Second, when studying ad block prevalence in the census network, ad blocking is associated directly with browser cookies, not users. In this study, there is nearly a one to one correspondence between a user and a cookie. Using the subset of cookies associated with the panel, the estimated number of cookies per person was 1.03, sufficiently close to 1 to be omitted.

3. MIXTURE PROPORTION ESTIMATION

Mixture proportion estimation (MPE) [8] is a technique for finding the proportions of classes in unlabeled data sets. While there are many variations, the basic problem setup can be captured as follows. Let $1, \ldots, k$ denote the classes, and let $P_1, \ldots, P_k$ be known, estimated, or otherwise restricted class conditional probability distributions over a feature space. Given unlabeled data with a probability distribution P, and the estimates of the class conditional distributions, find the relative proportion of each class in the unlabeled data. In short, given $P_1, \ldots, P_k$, and P, find proportions $\pi_1, \ldots, \pi_k$, such that

$$P = \sum_{i=1}^{k} \pi_i \, P_i. \tag{2}$$

MPE is a powerful approach as it circumvents the need for data classification. The proportions of the classes are inferred from the data in aggregate, permitting success even when classification error rates are prohibitively high. In contrast, a manifest and domineering approach to this problem is to first classify each data point, and then find the proportions of each class directly from the inferred classification. This approach is inherently limited by the error rates of the classifier; in the noisy, feature limited setting studied here, where classification would be based on the *ad ratio* statistic alone, this approach can fail entirely.

In the context of estimation of ad blocking deployment, the MPE approach relies on the distribution of the ad ratio statistic associated with the three populations of cookies: $\mathcal{S}_{\text{ads}}$, $\mathcal{S}_{\text{block}}$ and $\mathcal{S}$, corresponding to labeled cookies without ad blocking software, labeled cookies with ad blocking software, and unlabeled cookies. These populations define three histograms over the *ad ratio* statistic:

1. $\widehat{P}_{\text{ads}} \in \mathbb{R}^m$ – normalized histogram over the ad ratio statistic for cookies known to have one or more ad blockers installed
2. $\widehat{P}_{\text{block}} \in \mathbb{R}^m$ – normalized histogram over the ad ratio statistic for cookies not associated with one or more ad blockers,
3. $\widehat{P} \in \mathbb{R}^m$ – normalized histogram of unlabeled comScore cookies from the comScore census network.

Histograms are generated so that each of the m bins contains approximately the same number of cookies for $\widehat{P}$, *i.e.,* the boundaries of the bins are such that $\widehat{P} \approx \left[\frac{1}{m}, \ldots, \frac{1}{m}\right]$. With the three histograms as input, a single variable optimization is run to find the mixture proportion such that the labeled histograms *best align* with the unlabeled histogram according to (2). The procedure is detailed in Algorithm 1.

Algorithm 1 MPE for Ad Block Prevalence

1: *Input:* ad ratio statistics for cookies in $\mathcal{S}_{\text{ads}}, \mathcal{S}_{\text{block}}, \mathcal{S}$.
2: Compute bin edges $a_1, \ldots a_{m-1}$ such that histogram of ad ratio for $\mathcal{S}$ satisfies

$$\widehat{P} \approx \left[\frac{1}{m}, \ldots, \frac{1}{m}\right]$$

3: Generate histograms $\widehat{P}_{\text{block}}$, $\widehat{P}_{\text{ads}}$, and $\widehat{P}$ by binning data according to $[0, a_1), [a_1, a_2) \ldots [a_{m-1}, \infty)$
4: Solve optimization:

$$\pi^* = \arg \min_{\pi \in [0,1]} f\left(\widehat{P}, \, \pi \, \widehat{P}_{\text{block}} + (1-\pi)\widehat{P}_{\text{ads}}\right)$$

5: *Output:* estimate of proportion of users with ad block software π^* and significance level (p-value) derived from χ^2 test

The algorithm returns the estimate of the proportion of ad block users, denoted π^*. We refer to this value as the *MPE percent adblock*. As equality in (2) is rarely if ever satisfied for any mixture proportion, the mixture proportion that results in the minimum objective function, $f(\cdot)$, between the histograms is returned by the optimization.

We employ three objective functions: the canonical L_1 and L_2 norms and the χ^2 statistic, given by:

$$f(P, Q) = n \sum_{i=1}^{m} \frac{(Q(i) - P(i))^2}{P(i)} \tag{3}$$

where n is the number of samples in histogram Q.

Validation via Significance Testing. The underlying hypothesis of MPE is that unlabeled data are well represented by a combination of the labeled datasets according to (2). After the optimization is run, this hypothesis can be confirmed or rejected by asking, in the context of goodness of fit (GoF) testing, how well does the best mixture distribution, $\pi^* \widehat{P}_{\text{block}} + (1-\pi^*)\widehat{P}_{\text{ads}}$, match the data, $\widehat{P}$? If the results of the GoF test indicate that the histograms aren't well matched the original hypothesis is rejected, and the estimated proportions are invalid.

The canonical GoF test for categorical data is Pearson's Chi-squared test [1]. In general, the Chi-squared test takes the χ^2 statistic and the number of samples associated with the histogram under test as input, and outputs a p-value. Note that in our setting, a high p-value is good, as it indicates the data is well represented by the mixture distribution. As both histograms are empirical, we consider the larger dataset, $\widehat{P}$, to represent the theoretical frequencies, which is a standard approach in GoF testing. When the χ^2 statistic is used as the objective, the GoF test is baked into the optimization and MPE approach is equivalent to finding the mixture proportions that result in the maximum p-value (*i.e.,* the least statistically significant outcome, the outcome that best matches the data). We use Chi-squared tests to confirm the validity of the MPE approach on the various datasets.

4. RESULTS

This section presents estimates on the prevalence and impact of ad blockers in the internet. More precisely, our MPE approach was used to generate ad block percentages for key geographies as shown in Table 1. From these percentages, the *MPE projection factor* for each of the key geographies was determined by dividing the *MPE percent ad block*, with the χ^2 objective, by the *panel percent ad block*. The MPE approach can be applied to any arbitrary "breakout" (subset of the overall population) by multiplying the projection factors by panel results for a target breakout. Due to space considerations, we limit our analysis to several key breakouts including *(i)* geographic, *(ii)* demographic, and *(iii)* publisher, which highlights the impact of ad blockers in terms of potential revenue loss.

To motivate and validate our MPE-based approach for this problem we provide a visual representation for the underlying distributions for $\hat{P}$, $\hat{P}_{\text{ads}}$, and $\hat{P}_{\text{block}}$ for the US in Figure 1. Alongside these three histograms (*e.g.,* the bottom right histogram), we provide the histogram generated by using the α value that minimized the objective function in our MPE algorithm. Visually, the success of the method is dictated by how closely the histogram generated by the mixture of $\hat{P}_{\text{ads}}$ and $\hat{P}_{\text{block}}$ matches the histogram of $\hat{P}$. Figure 1 shows that the two histograms match quite closely.

4.1 Geographic Breakout

Our geographic analysis of ad blocker prevalence considers the US, the UK, Germany, France, and Canada. These countries were selected because they are all large digital advertising markets. Table 1 shows results of the MPE approach using L_1, L_2, and the χ^2 statistic as the objective

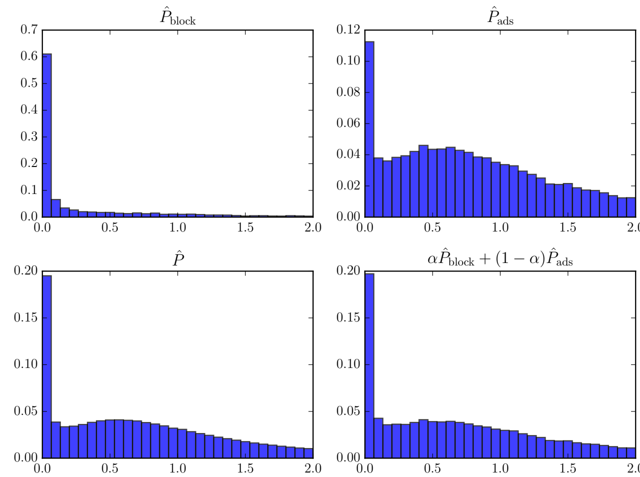

Figure 1: The underlying distribution of the ad ratio statistic associated with the three populations of cookies from Section 3: $\hat{P}_{\text{block}}$, $\hat{P}_{\text{ads}}$, and $\hat{P}$. The bottom right histogram is the mixture combination of $\hat{P}_{\text{ads}}$ and $\hat{P}_{\text{block}}$ utilizing the MPE approach. Visually, the success of the method is dictated by how closely the histogram generated by the mixture of $\hat{P}_{\text{ads}}$ and $\hat{P}_{\text{block}}$ matches the histogram of $\hat{P}$.

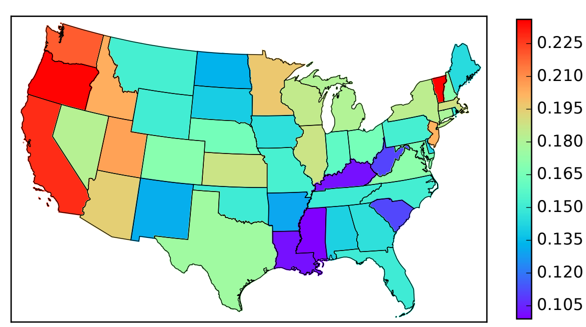

Figure 2: A heat map of ad blocker penetration on a state by state in the US. Vermont has the highest ad blocker penetration at 23.6% and Mississippi has the lowest at 9.9%.

functions. Ad block penetration in the US is 18% and varies between 16% and 37% for other countries.

Within the US, we consider ad blocker installations on state by state basis. Figure 2 quantifies the ad block penetration rates in a heat map. We find that ad blocker penetration is greatest in Vermont (23.6%) and lowest in Mississippi (9.9%).

4.2 Demographic Breakout

Figure 3 provides ad blocker penetration rates for key demographic categories in Germany, the UK, and the US. Ad blocker penetration is most prevalent among males 18-34. This finding is consistent across all geographic areas with Germany at 49%, the US at 29%, and the UK at 29%. The 18-34 age group is also consistently (across key geos) the most prevalent ad blocker group among females with Germany at 43%, the UK at 22%, and the US at 20%.

	Feb-16							
Geo	L_1	L_2	χ^2	p	95% Confidence	n_{block}	n_{ad}	n_{census}
US	18%	18%	17%	0.10	15.7% - 18.6%	6,788	52,368	49,406,827
UK	16%	16%	17%	0.88	11.5% - 23.5%	2,200	11,952	8,660,037
Germany	32%	32%	37%	0.56	28.4% - 46.6%	1,114	2,142	3,174,325
France	29%	29%	32%	0.89	22.5% - 42.5%	1,133	3,016	3,949,981
Canada	22%	22%	24%	0.52	18.5% - 30.5%	1,666	6,033	5,376,049

Table 1: The percentage of users with an ad blocker installed (the *MPE percent ad block*) in key geographies for the month of February. Results from using L_1, L_2, and χ^2 statistical distance with p-value as the objective function are shown. A high p-value indicates success of MPE as it implies the resulting mixture distribution is *not* statistically significant with respect to the unlabeled data. The 95% confidence values indicate the range for which the corresponding mixture distribution has a p-value greater than 0.05. The underlying size n of each data set used in generating $\widehat{P}_{\text{ads}}$, $\widehat{P}_{\text{block}}$, and $\widehat{P}$ are also provided.

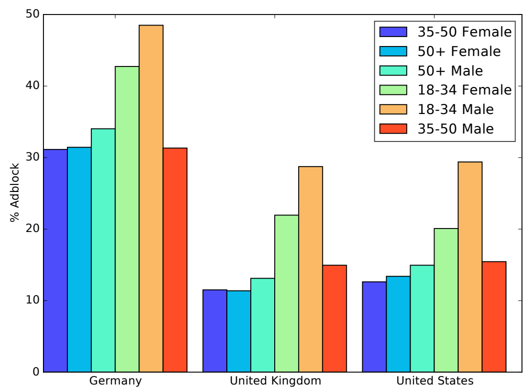

Figure 3: Ad block penetration rates among key user demographic categories for Germany, the UK, and the US.

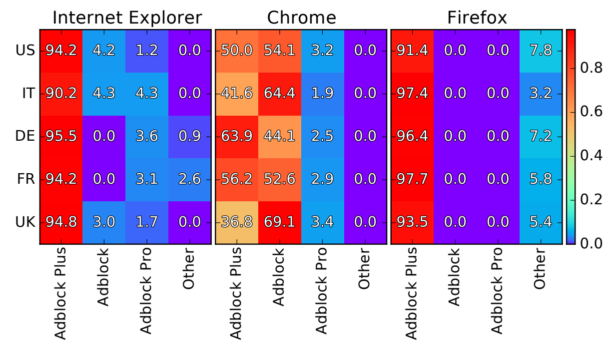

Figure 4: A heat map showing the market share of the top three ad block offerings across three major browsers (Internet Explorer, Google Chrome, and Mozilla Firefox). Results are further stratified across key geographies.

4.3 Ad Block Market Share Analysis

As indicted in Section 2, there are a number of different ad blockers available and in use today. A number of these report their total installations. Thus, it is of interest to investigate their relative market share, and our data and analytic method allows us to estimate the prevalence of specific browsers/ad blocker deployments. To quantify this, we focused on three major browsers (*e.g.,* Internet Explorer, Google Chrome, and Mozilla Firefox). Figure 4 is a heat map of the relative market share of the top three ad block offerings (as well as a catch all *Other* category) across three major browsers. It is clear from Figure 4 the market is dominated by Adblock Plus for both Firefox (95.2% averaged across geos) and Internet Explorer (93.8% averaged across geos). For Google Chrome, the market share is distributed fairly evenly between Adblock Plus (49.7% averaged across geos) and Adblock (56.9% averaged across geos). Note, the values for a particular geo/browser pair will not sum to 1 as a single panelist may have more than one ad block offering installed.

4.4 Publisher Breakout

The analysis in subsections 4.1 and 4.2 on ad blocker penetration across key geographies and key demographics highlights the difference in ad block use among different population segments. For instance, ad blocker penetration skews toward young males. These users inherently carry bias in the sites they are likely to visit. Thus, the percentage of users with an ad blocker installed can be markedly different from site to site. To quantify this behavior and it's associated impact, we calculate the percentage of users with an ad blocker installed across *(i)* ten publishers and *(ii)* the major publisher segments. The ten publishers are a random selection of those with large audiences that illustrate the scope of impact of ad blockers. Publisher names have been removed to preserve anonymity.

Feb-16	
Publisher segment	% Ad Block users
Automotive	18.82%
Entertainment	20.21%
Games	22.30%
Lifestyles	20.56%
News/Information	20.21%
Portals	17.77%
Search/Navigation	18.12%
Sports	20.91%
Technology	20.04%
XXX Adult	24.74%

Table 3: The percent of users with an ad blocker installed by publisher segment.

Table 2 provides the results for ten publishers. The percentage of users with an ad blocker installed ranges from 25.27% for Publisher H to 17.95% for Publisher B. Note, Publishers A, B, D, E, and F appear in some form on Adblock Plus's whitelist while Publishers C, G, H, I do not.

To further elucidate user ad block install behavior, we ran the same analysis across publisher segments. The re-

Publisher	Feb-16					
	% Ad block users	% Ad requests blocked	AV_{ad}	AV_{block}	Ad blocker exposure rate	Potential revenue loss
Publisher A	19.52%	18.99%	0.23	0.08	0.34	$1,550,138
Publisher B	17.95%	5.17%	0.95	0.57	0.60	$508,534
Publisher C	21.09%	5.82%	1.95	1.49	0.76	$3,904,207
Publisher D	18.47%	7.76%	1.33	0.72	0.54	$1,575,406
Publisher E	21.96%	14.63%	0.40	0.17	0.42	$183,531
Publisher F	18.82%	8.06%	0.69	0.31	0.44	$190,625
Publisher G	21.43%	16.21%	1.81	0.55	0.30	$195,651
Publisher H	25.27%	16.07%	2.10	0.76	0.36	$170,779
Publisher I	23.00%	14.42%	0.70	0.22	0.31	$121,581

Table 2: Publisher breakout. Publisher names have been anonymized. The table shows the percent of ad block users and the percentage of ad requests blocked. AV_{ads} and AV_{ads} are the number of ads shown per page view to ad block users and non-users. The *ad blocker exposure rate* is the number of ads delivered to an ad block user per ad delivered to a non-blocker. Lastly, estimated potential revenue lost due to ad block usage is shown.

sults of this analysis are shown in Table 3. The percentage of users with an ad blocker installed are highest on XXX Adult (24.74%) and lowest on Portals (17.77%). Notably, ad blocker install rates among users visiting Sports, Technology, Entertainment, and News/Information segments is quite similar. These segments represent sites where users consume various forms of media and are therefore likely to be confronted with advertisements that disrupt the consumption of this media.

4.5 Ad Blocker Impact Analysis

It is clear from the previous subsection that a significant proportion of *users* employ ad blockers. However, only considering users ignores two important factors: *(i)* ad block users (could) still be exposed to a significant number of ads due to whitelisting and disabling and *(ii)* different classes of users have different browsing behaviors.

To capture these effects, we considered the *ad blocker exposure rate*, which is interpreted as the number of ads shown to an ad block user per ad shown to a non-block user. Table 3 and Table 2 show the *ad blocker exposure rate*, computed as $AV_{\text{block}}/AV_{\text{ads}}$, where AV_{block} and AV_{ad} are the number of ads per page view shown to ad block users and non-users.

Additionally, we calculated the percentage of ad requests blocked and the potential revenue lost for ten publishers. The revenue lost assumes a modest $1 CPM (cost per thousand ads shown, selected arbitrarily). The calculation of the ad requests blocked is outlined here:

1. compute X, the percentage of ad block users, as *panel percent ad block* on publisher A multiplied by the *MPE projection factor* for the geo of interest
2. multiply X by the total number of users on publisher A, which gives an estimate of *the number of ad block users* on publisher A, denoted Y
3. from panel data, compute the average page views per ad block user for publisher A, and multiply by Y to get Z, *the number of page views from ad block users* on publisher A
4. from panel data, compute the average number of ads shown per page view to ad block users (AV_{block}) and the average number of ads shown per page view to non-block users (AV_{ad})
5. compute blocked impressions as $Z\times(AV_{\text{ads}}-AV_{\text{block}})$

Feb-16	
Country	Ad Blocker Exposure Rate
Canada	0.46
France	0.35
Germany	0.51
Italy	0.45
UK	0.49
US	0.60

Table 4: *Ad blocker exposure rate*. The ad blocker exposure rate is the number of ads delivered to an *ad block user* per ad delivered to a non-*ad block user*.

The potential revenue lost R_{lost} using a $1 CPM (cost per thousand ads shown, selected arbitrarily) for the ten publishers is found in Table 2. The intermediary values for AV_{ad} and AV_{block} are also found in Table 2. Even with a modest $1 CPM, it is clear that ad blockers have a significant impact on revenue lost for publishers (*e.g.,* $3.9M/mo. for Publisher C to $120K/mo. for Publisher I).

5. RELATED WORK

A study of general aspects of the ad serving ecosystem is reported in [2]. That paper highlights the impact of user-targeting in online advertising as well as the broad range of ads that are delivered to different types of publishers. Nath presents a similar study, which is focused on advertising in the mobile app space [4]. More specific to ad blockers, Walls *et al.* [9] provide a comprehensive analysis of Adblock Plus's Acceptable Ads (*i.e.,* whitelisting) program. This study directly informs our work with respect to the issue of whitelisting by ad blockers. In one of the few academic studies on the topic, Pujol *et al.* [7] use passive measurements on a residential broadband network to infer ad block users and classify the impact of ad blockers on HTTP traffic. Our work is based on different datasets and employs different methods to infer ad blocker use. Recently, Post and Sekharan investigated the capabilities of 3 popular ad blocker via source code analysis [6]. Finally, there continue to be many reports in the popular press on ad blocker prevalence. Most of these are based on data provided (*e.g.,* [3]) by entities that work with publishers to deploy a page-based detector. Our work provides a complementary perspective based on instrumentation deployed widely throughout the internet.

6. SUMMARY AND CONCLUSIONS

In this paper we used two unique data sets to estimate the prevalence and impact of ad blockers in the internet. We use mixture proportion estimation to remove bias from a world-wide user panel that provides ground truth on ad blocker installs.

Our results show that in the US, 18% of users have ad blockers installed; up to 37% of users have ad blockers installed in other key geographies. Males 18-34 are the most likely users of ad blockers in all geographies studied, with 49% using ad blocking software in Germany. Users with the highest income levels are the most likely users of ad blockers in all geographies studied. Results show that Adblock Plus is the most widely used ad blocker. Ad blockers are most prevalent on Chrome, followed by Firefox. Users with ad blockers see roughly half as many ads that users without ad blockers see. Estimated monthly revenue lost due to ad blockers on 10 large publisher sites varies between \$3.9M and \$120K (assumes \$1 CPM).

While the the results in this paper help further the understanding of ad block prevalence and impact, we would be remiss to not mention two inherent limitations. First, internet-wide population or behavior estimation can never be shown to be completely accurate due to scale, complexity and dynamics. However, we argue that our methodology that includes confidence ranges, enables results to be more effectively judged and interpreted. Second, our reliance on proprietary datasets by definition limits the repeatability of our experiments. In future work, we intend to address these limitations by continuing to refine our analytic capabilities, and to consider how ad blocker prevalence might be measured in a way that does not rely on proprietary data.

7. REFERENCES

[1] A. Agresti and M. Kateri. *Categorical Data Analysis*. Springer, 2011.

[2] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan. Adscape: Harvesting and Analyzing Online Display Ads. In *Proceedings of the 23rd World Wide Web Conference*, Seoul, Korea, April 2014.

[3] S. Blanchfield. PageFair, 2016. https://pagefair.com.

[4] S. Nath. MAdScope: Characterizing Mobile In-App Targeted Ads. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, Florence, Italy, May 2015.

[5] W. Palant. Adblock plus, 2016. http://adblockplus.org.

[6] E. Post and C. Sekharan. Comparative Study and Evaluation of Online Ad-Blockers. In *Proceedings of the 2nd International Conference on Information Science and Security*, Seoul, Korea, December 2015.

[7] E. Pujol, O. Hohlfeld, and A. Feldmann. Annoyed Users: Ads and Ad-Block Usage in the Wild. In *Proceedings of the ACM Internet Measurement Conference*, Tokyo, Japan, October 2015.

[8] C. Scott. A Rate of Convergence for Mixture Proportion Estimation, with Application to Learning from Noisy Labels. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, San Diego, CA, May 2015.

[9] R. Walls, E. Kilmer, N. Lageman, and P. McDaniel. Measuring the Impact and Perception of Acceptable Advertisements. In *Proceedings of the ACM Internet Measurement Conference*, Tokyo, Japan, October 2015.