Block the blocker: Studying the effects of Anti Ad-blocking

Rohit Gupta, Rohit Panda

Master of Science, Informatics, Technical University of Munich
`rohit.gupta@tum.de, rohit.panda@tum.de`

Abstract. Advertisements generate huge chunks of revenues for websites and online businesses. Ad-blocker and tracker blocking programs have gained momentum in the last few years with massive debates raging on privacy concerns and improving user experience online. Acceptable Ads programme and Anti Ad-blockers are primary elements emerging in recent years that combat ad-blockers.

In our seminar paper, we discuss at length data collection of top websites in the world, Germany, DACH region and news category. We generate feature based A/B testing metrics and employ classifier evaluations on them along with then analysing the result. Our paper also discusses how Anti Ad-blockers impact the economic, legal and ethical usage in Germany alongwith the recent changes in GDPR while taking a look at Acceptable ads programme and Whitelisting.

Keywords: Advertisements, Ad-blocker, Privacy, Anti Ad-blocker

1   Introduction

1.1   Motivation

Advertisement is a technique used by product and service companies that detail the specific selling points of their products by means of which it becomes known to the masses. Marketers usually place ads at calculated locations for their products such that they are able to draw the attention of potential consumers and persuade them to buy it.

Online advertising or online marketing is a technique of delivering promotional messages to various consumers. It is what drives the economy of most of the World Wide Web businesses. Most modern websites, in general, tend to monetize their user visits. They include certain spaces across their websites aimed at advertisers to come and put their promotional content. An unsaid agreement exists between the website owner and the advertisers on displaying only genuine promotional content and not include malwares or even involve scamming a user.

Ad-blockers have emerged as tools that blocks such advertisements to improve users' web-browsing experience, maintaining privacy, and recently protecting themselves against malware. This has impacted businesses that rely on revenues from advertisements.

Anti Ad-blockers have emerged as an increasingly popular solution. Anti Ad-blockers detect if Ad-blockers are running and use several techniques such as simply notifying the user that the tool interferes with the content and the user-experience or other times when the message blocks the user from accessing the content until they have turned off the Ad-blocker. In extreme cases the goal is to circumvent the tool completely.

Interactive Advertising Bureau (IAB) have been actively working on this topic and have also made a script to "DEAL (Detect, Explain, Ask, Limit)" with such ad-blockers public. They call this "Ad Block Detection Code Access Request" and their source code is available on Github [3]:

https://github.com/InteractiveAdvertisingBureau/AdBlockDetection

1.2 Roadmap

In our paper we propose to study approaches to:

- Study the usage of these Anti Ad-blocker scripts and their mechanism.
- Collect and analyse top websites in the world using Alexa Website Rankings.
- Economic impact of Anti Ad-blockers.
- Legality and ethics of Anti Ad-blocking.
- Impact of GDPR on Anti Ad-blocking.
- Alternatives to Anti Ad-blocking such as Whitelisting and Acceptable Ads programme and also taking a look at Anti Ad-block killers.

This paper is structured as follows:

Section 2: In this section, we discuss the background related to how Anti Ad-blockers work and also take a look at the related work on this topic.

Section 3: In this section, we discuss how we collect top websites from different sources.

Section 4: In this section, we discuss how we process the data that we collected including methodology used for our results analysis.

Section 5: In this section, we discuss the results for the methodology used in the previous section.

Section 6: In this section, we analyze the impact of Anti Ad-blockers such as the economic, legality, ethical aspects including how recent changes in GDPR affects Anti Ad-blockers.

Section 7: In this section, we discuss alternatives to Anti Ad-blocking such as Acceptable Ads and Whitelists and also take a look briefly at Anti Ad-block killers.

Section 8: We conclude our discussion of our work and provide the roadmap ahead.

2   Background and Related Work

2.1   Background

Online advertisements or online marketing involves putting out promotional content on the Internet for users. Largely, the economy on the World Wide Web

hangs on the fate of online advertisements. Advertisements are put out at different locations on websites such that users are enticed to look at them and react if they are interested in such promotions.

This has led to a stark increase in security and personal concerns and issues. Users argue that they do not want such annoying advertisements on websites because it hampers their area of viewing actual content seamlessly and without obstructions. Some users also say that they could be used to deploy malwares and scams on the web.

Ad-blocker tools are countering this problem by introducing a technique of removing such undesirable advertisements. Some of the famous Ad-blocker tools include Adblock Plus or Ghostery. As of July 2018, more than 10 million users use Adblock Plus on Google Chrome [6] which shows that users are actively seeking to filter out or limit such advertisement promotions.

In a measurement study "Annoyed Users: Ads and Ad-Block Usage in the Wild" by Pujol et. al. [7], they observe that 22% of active users use Adblock Plus on their browsers.

The popularity of Ad-blockers in Germany. According to a "2017 Global Adblock Report" published by PageFair [10], Germany ranked highest in Ad-block installations when it came to Ad-blockers penetration data for desktop (Figure 1).

Adblock penetration per online capita, %
PageFair

	Desktop	Mobile	Overall
U. States	18%	1%	18%
China	1%	13%	13%
U. Kingdom	16%	1%	16%
Japan	3%	–	3%
Germany	29%	1%	29%
Australia	20%	2%	20%
Canada	24%	–	25%
France	11%	1%	11%
Brazil	6%	1%	6%
South Korea	4%	–	4%
Russia	6%	3%	6%
Italy	17%	1%	17%
Netherlands	17%	2%	17%
Spain	19%	–	19%
Mexico	8%	–	9%
Sweden	27%	–	27%
Indonesia	8%	58%	58%
Denmark	25%	2%	25%
India	1%	28%	28%
Argentina	14%	–	14%
Finland	23%	–	23%

Fig. 1. Top Ad Markets (ad spend) statistics taken from a PageFair report titled "2017 Global Adblock Report".

These statistics indicate a tremendous potential for European markets since there is no clear single solution for mobile Ad-block usage.

On the other hand, Online-Vermarkterkreis (OVK) and Bundesverband Digitale Wirtschaft (BVDW) published a report titled "Zentrale Adblocker-Rate

des OVK" [11] in December 2016, that shows a decline in the "incidence rates" related to ad-blockers from the first quarter to the third quarter of 2016 (Figure 2).
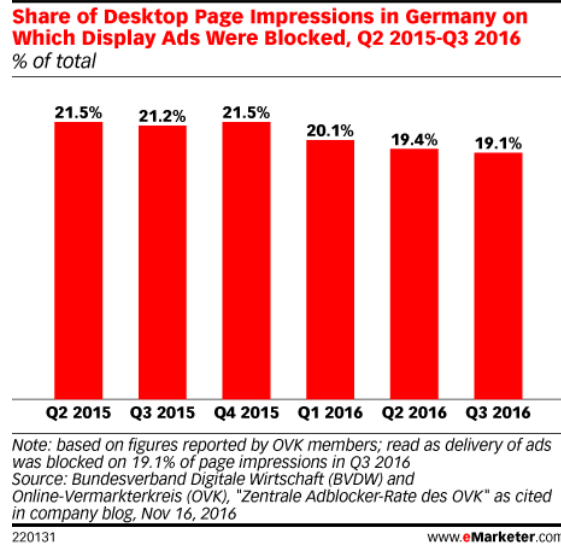

Fig. 2. Share of Desktop Page Impressions in Germany on Which Display Ads Were Blocked according to a report by BVDW and OVK [11]

The critical difference to be noted in the two sources cited above is that PageFair looked at the number of penetration or installations of Ad-blockers whereas BVDW & OVK focused on the actual number of websites that were affected by such Ad-blockers.

Working of Ad-blockers. Ad-blocker tools work mostly on removing browser page elements. They look for certain HTML, DOM or CSS elements in the web page and process them to be removed. Ad-blockers usually work on a set of rules that indicate which such elements must be removed. These rules are part of "filter lists" such as EasyList [12]. There are also user privacy protection lists such as EasyPrivacy [13] which remove trackers. Another technique that Ad-blockers use is web request blocking, where these tools remove URLs that correspond to any publisher.

How Anti Ad-blockers work. Anti Ad-blocker scripts detect the presence of Ad-blockers and displays appropriate messages such as asking users to turn off Ad-blocker tools or not allowing users to view content. One such script has been published by IAB that "DEAL"s (Detect, Explain, Ask, Limit) with such Ad-blockers [3].

Anti Ad-blocker scripts are known to perform couple of operations. The first is to detect any Ad-block tool being used and the second is to notify the user
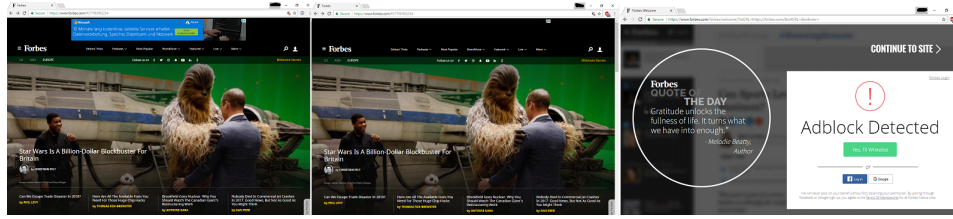
Fig. 3. Web page evolution for Forbes shows from left to right web page with ads, with Ad-blocker enabled and with an Anti Ad-blocker.
Source: `www.forbes.com`

to either disable Ad-block or to whitelist their website. These scripts may range from using first-party domains that check only aesthetic attributes such as height or width of ads to more complicated scripts from third-party domains that provide baits such as time delays, continuous detection or even use cookies to track Ad-block detection. It is a big challenge to detect such scripts since they are obfuscated and hidden deep into the system.
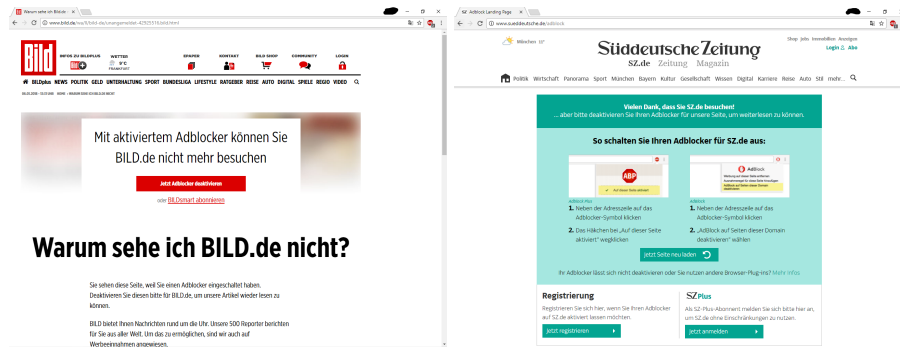

Fig. 4. German websites using Ad-blocker detection scripts.

Figure 4 shows examples of German websites that show Ad-block detection responses. This is done by employing Anti Ad-blocking tools or scripts such as the one by IAB which detect the presence of such Ad-blockers.

Impact of Anti Ad-blockers. "The Ad Wars: Retrospective Measurement and Analysis of Anti-Adblock Filter Lists" is a paper published jointly by scientists at University of Iowa and University of California-Riverside. In this paper, Iqbal et. al. scan Alexta Top 5K websites for Anti Ad-block tools such as Anti Ad-block Killer List [16] and Combined EasyList for Anti Ad-block scripts. The results as shown in (Figure 5) taken from their paper, showed a gradual increase in Anti Ad-blocking scripts, which indicate that the online advertising industry reacting to the immediate losses.
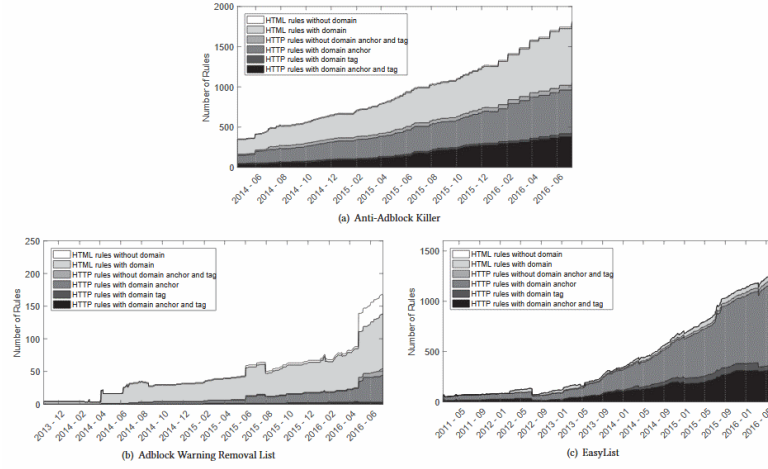
Fig. 5. Anti Ad-blocking scripts with their gradual rise in usage taken from the paper by Iqbal et. al. [15]

2.2   Related Work

Our paper discusses in length the Anti Ad-blocking techniques and their effects adopted mainly by Nithyanand et al. [4] and Haris et al. [1]. Nithyanand et al. [4] in their paper "Adblocking and Counter-Blocking: A Slice of the Arms Race", found that 6.7% from the top 5K Alexa websites deploy Anti Ad-blockers. They also consider all websites that include Anti Ad-block scripts in their analysis.

Haris et al. [1] in their paper "Detecting Anti Ad-blockers in the Wild", noted that 686 websites in the Alexa top 100K use Anti Ad-blockers on their home page. They also notice different notifications ranging from asking users to disable Ad-blockers to paying a subscription fee.

Garimella et al. [2] in their paper "Ad-blocking: A Study on Performance, Privacy and Counter-measures", discuss extensively multiple Ad-Blocker tools and compare their performance and privacy facets.

3   Data Collection

As part of our data collection, we collected top websites from Amazon Alexa Website Rankings: *https://www.alexa.com/topsites*. Publicly available data is restricted to only 50 per category out of the 500 available ranked websites. The ranking is done based on the average daily visitors to a site and also over the pageviews for that site for the last month. We subscribed to a 7-day free trial to collect top websites for our different category use cases. Our data sets include top 500 news websites, top 500 websites in Germany and a unique subset of top websites in the DACH (500 each for Germany, Austria and Switzerland) region.

To pull this data, we had to generate an access key with an Amazon AWS account which grants policy rights to AWIS (Amazon Web Information Services) located here: *https://aws.amazon.com/awis/*. Furthermore, we had to subscribe to the 7-day Amazon Alexa free trial so that we can pull the information from Alexa using API calls. Listing 1 shows an HTTP request using query parameters used to pull data using this API. We used Java to pull this information. We made use of the sample code provided by Amazon AWIS (*https://aws.amazon.com/awis/getting-started/*) and modified it to our needs to pull the site names.

```
1  https://awis.amazonaws.com/api?Action=CategoryListings&Count=20
2          &Descriptions=True&Path=Top%2FNews
3          &Recursive=False&ResponseGroup=Listings
4       &SortBy=Popularity&Start=1
```

Listing 1: An HTTP request to AWIS to pull top news sites

A maximum of 20 results can be pulled at each HTTP request to the host *"awis.amazonaws.com"* with a starting count of counter = 1 (Start). This counter is incremented by 20 for every request, till all the data has been pulled in. The code uses SHA-256 hashing algorithm to create an Authorization that includes the Access Key obtained earlier.

The response contains a list of websites sorted according to the parameters specified. A sample response is listed in Listing 2. The response is then parsed using an XML parser to obtain each website name and is stored in a file. For generating the unique subset of all the websites in the DACH region, we created a HashSet that does not allow duplicates to be entered into the set. This is done so that common websites is only taken into consideration once.

These websites are then sent through our data processor described in the next section.

4   Anti Ad-blockers Detection

In this paper, we are trying to replicate and reproduce the work done by Haris et al. [1] in their paper "Detecting Anti Ad-blockers in the wild", and by Rishab et al. [4] in their paper "Adblocking and Counter-Blocking: A Slice of the Arms Race", but in a different context. Haris et al. used websites from the Alexa top-100K which show a slight bias towards the American Ad-blocking industry. This is evident from the country of origin of Alexa top-100 websites listed on Alexa's website [24] where Google Germany and Amazon Germany are the only German sites in that list.

```
1  <aws:CategoryListingsResponse xmlns:aws="http://alexa.amazonaws.com/doc/2005-10-05/">
2  <aws:Response xmlns:aws="http://awis.amazonaws.com/doc/2005-07-11">
3  <aws:OperationRequest>
4  <aws:RequestId>3ffe5d1a-8939-5b7a-9002-bec5a1243i9a</aws:RequestId>
5  </aws:OperationRequest>
6  <aws:CategoryListingsResult>
7  <aws:Alexa>
8    <aws:CategoryListings>
9      <aws:RecursiveCount>21</aws:RecursiveCount>
10     <aws:Count>2</aws:Count>
11     <aws:Listings>
12       <aws:Listing>
13         <aws:DataUrl type="navigable">http://www.reddit.com</aws:DataUrl>
14         <aws:Title>Reddit.com</aws:Title>
15         <aws:PopularityRank>2</aws:PopularityRank>
16       </aws:Listing>
17       <aws:Listing>
18         <aws:DataUrl type="navigable">http://www.cnn.com/</aws:DataUrl>
19         <aws:Title>CNN</aws:Title>
20         <aws:PopularityRank>3</aws:PopularityRank>
21       </aws:Listing>
22     </aws:Listings>
23    </aws:CategoryListings>
24  </aws:Alexa>
25  </aws:CategoryListingsResult>
26  <aws:ResponseStatus xmlns:aws="http://alexa.amazonaws.com/doc/2005-10-05/">
27  <aws:StatusCode>Success</aws:StatusCode>
28  </aws:ResponseStatus>
29  </aws:Response>
30  </aws:CategoryListingsResponse>
```

Listing 2: An HTTP response in XML for top news sites.

Our focus in this paper are the popular sites in Germany, DACH region (Germany, Austria, Switzerland) and we also take a look at the sub-categories of News websites. We plan to compare our results with the results obtained by Haris et al. and Rishab et al. thereby producing an analogy to their findings. Our assumption is that our results will be similar to theirs. The previously mentioned papers also do not provide any source code to their technique but explain their methodology in the paper which has been adopted by us. In this section, we

try to explain how we have adopted and modified the methodology proposed by them.

4.1 Overview

As mentioned earlier in this paper, HTML and DOM elements in a browser page contains source code of websites which use Anti Ad-blockers and which do not. Websites rely mostly on adding and/or modifying HTML content for detecting Ad-blocker usage.

Haris et al. in their paper categorize these changes into the following categories:

Nodal Changes: These changes are concerned with the addition of extra HTML DOM elements.

Style Changes: These changes are concerned with the modification of HTML DOM elements relating to their style and appearance attributes.

Textual Changes: These changes are concerned with the modification in textual elements in any HTML page.

Structural Changes: These changes are concerned with the modification in elements such as the innerHTML DOM property and whether any changes lead to redirection to another URL.

For websites that show such changes as mentioned above with Ad-blockers enabled, they are said to be using Anti Ad-blockers.

4.2 Methodology

The methodology we use is to capture and record these changes made by Websites and create a model to predict the usage of Anti Ad-blocker scripts. To not over complicate the technique and simplify the process, we consider only Nodal Changes, Textual Changes and Structural Changes as described in the previous section and altogether ignore Style changes and the innerHTML component of Structural changes from our experiments. Similar to Haris et al. we perform A/B testing using Selenium Chrome WebDriver [25] using an extension of Ad-blocker in one instance and not using one in another instance.

Browser instance automation To automate this process, we make use of Selenium Chrome Webdriver [31] to launch two separate instances of Google Chrome web browser one with an Ad-block extension installed and one without one. We open the sites under test with each instance.

Browser profiles We create two new browser profiles in Google Chrome. In one of the profiles we install Ad-blocker extension and in the other we do not. This provides a clean system under test not subject to any previous website caching for that profile. We use Adblock Plus [6] due to its popularity. It has options to configure various filter lists such as [14]. We use the EasyList filter [12] after removing all the Anti Ad-block rules as well as disallow any possible Acceptable Ads. The Anti Ad-block rules are marked by the comments ! Anti Ad-block in the filter list.

Python Script for Data Collection A script in python was used for the data collection purpose. It reads a list of websites and then checks if the website is live. Then it opens the same website in two modes: One with Ad-block installed and one without and takes a screen grab (which can be used later to verify the results) and saves the HTML content for both cases.

BeautifulSoup library [26] is used for parsing the HTML file and recording the difference of various features and the presence of certain keywords in the text such as adblocker, adblock, ad block, ad-block, whitelist, block-adblock, pagefair, etc. This is then written to a csv file.

Nodal Features: We consider the following nodal features for our processing: anchor, div, h1, h2, h3, img, table, p, iframe and the text nodes.

Textual features: We consider the number of lines, words and the text characters.The following key words are also used in method A of the A/B testing: adblocker, adblock, ad block, ad-block, whitelist, block-adblock, pagefair, fuck-adblock.

Structural features: We note any URL redirections by using a simple yes or a no in generation of the csv file.

Creating and Training the model For the creation and training of the model, we used the Weka toolkit [27] that provides out of the box machine learning algorithms used for classifications. It also has a Java interface and a GUI which makes it easier to work with the data.

We used the following machine learning classifiers such as J48 Decision Tree, Random Forest and Naive Bayes which can then be used to identify websites employing anti Ad-blockers.

For the learning phase, we required labeled data. No such public database is available so it need to be done manually. For this purpose, we need some negative samples (websites that do not employ Anti Ad-blockers) and positive samples (websites that employ Anti Ad-blockers).

To collect positive samples we look into the issue list of anti-adblock-killer [32], it is used for blocking anti-adblockers so naturally websites on its issue list use Anti ad-blockers. We scraped the websites list from its issue list and manually checked them to confirm this. This formed our positive samples.

For negative samples, we collected a list of global websites using a collection of sources such as Amazon Alexa Web Ranking, Similarweb [28] and Quantcast [29] and manually checked them. We then ran these websites through our Python data collection script and manually labelled the websites in the csv file.

This was used as our training set and we ran this through various Weka classifiers to create our model.

Analysis of the features Our next step is to perform feature analysis on the features that we had generated in the previous section from the training data. Different attributes have different weights for discriminating between the classes to be learned. We use information gain ratio [30] for the same.

The rankings shown in Table 1 show that textual features have higher information gain which is similar to the results of Haris et al. although the information gain from the attributes maybe different.

Table 1. Features ranked based on Information Gain

Information Gain	Feature
25.65%	lines
20.66%	p
20.60%	a
20.44%	div
19.2%	words
18.05%	tags
16.12%	img
11.76%	h1
11.05%	keyword
11.05%	h3
9.15%	iframe
8.67%	h2
4.87%	table
0.78%	url change

4.3   Evaluating Classifier Models

Building on the previous work we perform different machine learning classification methods using the Weka toolkit. We use the most common ROC metrics such as precision, recall, and the area under ROC curve (AUC) to compare the different classifier algorithms used in our experiments.

The following tables from 2 to 7 summarizes the classification accuracy and the effectiveness of the classifiers we use for our training model.

Table 2. Random Forest - Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.024	0.993	1.000	0.996	0.984	0.999	1.000	FALSE
	0.976	0.000	1.000	0.976	0.988	0.984	0.999	0.998	TRUE
Weighted Avg.	0.995	0.019	0.995	0.995	0.994	0.984	0.999	0.999	

Table 3. Random Forest - Confusion Matrix

a	b	<− classified as
422	0	a = FALSE
3	121	b = TRUE

Table 4. Naive Bayes - Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.976	0.798	0.806	0.976	0.883	0.304	0.817	0.931	FALSE
	0.202	0.024	0.714	0.202	0.314	0.304	0.816	0.545	TRUE
Weighted Avg.	0.800	0.622	0.785	0.800	0.754	0.304	0.817	0.844	

Table 5. Naive Bayes - Confusion Matrix

a	b	<− classified as
412	10	a = FALSE
99	25	b = TRUE

Table 6. J48 Decision Tree - Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.988	0.250	0.931	0.988	0.959	0.806	0.896	0.943	FALSE
	0.750	0.012	0.949	0.750	0.838	0.806	0.896	0.837	TRUE
Weighted Avg.	0.934	0.196	0.935	0.934	0.931	0.806	0.896	0.919	

Table 7. J48 Decision Tree - Confusion Matrix

a	b	<− classified as
417	5	a = FALSE
31	93	b = TRUE

5   Results Analysis

5.1   Analysis of Test Set

Once we have trained our models, we need to run our test set which is the website list that we collected earlier through the Weka toolkit to generate the classification metrics. For each region and category, we generate the True Positives (TP), False Positives (FP) and the Precision involved for each of the three classifiers being used. We do not consider the True Negatives (TN) and the False Negatives (FN) due to the manual work and limitations for the time frame for this paper.

Germany Region. We collected 500 top websites belonging to Germany region from Alexa Website Rankings *https://www.alexa.com/topsites* for our data analysis. Running it through our methodology we were able to successfully identify and create a test set of 418 websites. The results are listed in table 8.

Table 8. TP, FP, Precision of Classifiers in Germany region

	Naive Bayes Classifier	J48 Classifier	Random Forest Classifier
Predicted	4	24	26
TP	3	8	8
FP	1	16	18
Precision	0.75	0.333	0.308

We observe that 8 websites from the 418 websites deploy an Anti Ad-blocker of some sorts. This accounts to 1.9% of the websites we identified for Germany region.

DACH Region. We collected 500 top websites belonging to Germany, Austria and Switzerland (DACH region) from Alexa Wesbite Rankings

https://www.alexa.com/topsites for our data analysis. From this set of 1500 websites, we took a unique set of websites, removing any duplicate websites that exist for each country. This accumulated to a total of 809 unique websites. The results for this set are listed in table 9.

Table 9. TP, FP, Precision of Classifiers in DACH region

	Naive Bayes Classifier	J48 Classifier	Random Forest Classifier
Predicted	10	40	41
TP	3	11	10
FP	7	29	31
Precision	0.3	0.275	0.244

We observe that 11 out of the 809 websites deploy an Anti Ad-blocker of some sorts. This accounts to 1.4% of the websites we identified for DACH region.

News Category.

We collected 500 top websites belonging to News Category from Alexa Website Rankings *https://www.alexa.com/topsites* for our data analysis. Running it through our methodology we were able to successfully identify and create a test set of 357 websites. The results are listed in table 10.

Table 10. TP, FP, Precision of Classifiers in News Category

	Naive Bayes Classifier	J48 Classifier	Random Forest Classifier
Predicted	7	18	16
TP	6	10	11
FP	1	8	5
Precision	0.857	0.556	0.688

We observe that 11 out of the 357 websites deploy an Anti Ad-blocker of some sorts. This accounts to 3.1% of the websites we identified for News category.

Comparison. We observed during verification that some websites have certainly designed different methods to evade any kind of Ad-blocker detection. This includes, but is not limited to ads that are allowed under the Acceptable Ads programme [18]. However we do not consider such a behaviour in generating the analysis metrics.

Haris et al. [1] observed in their paper that out of the Alexa top 100K websites they evaluated, they were able to identify 686 websites that used Anti Ad-blockers using their methodology.

5.2   Types of Ad-block Detection Responses

We manually collected and verified the positive samples for Ad-block detector responses. We use this to categorize how they behave when encountering an Ad-blocker.

We also measure the extent to which websites go to inform or request users from disabling Ad-blockers. We define a new terminology for this measure called "CIA". It involves the following three types:

Cost Model (C): We define this type as a cost or monetization model. This includes websites that ask for paid subscriptions (Figure 6), content for a limited number of days (Figure 7) or even donations.

Invisibility (I): We define this type as not allowing users to view content making their content invisible until Ad-blockers are disabled. Two stand out examples of this measure are: `www.bild.de` or `www.sport1.de` (Figure 8).

Availability (A): We define this type as a conservative approach to dealing with Ad-blockers. It is demonstrated by *https://t3n.de/* as shown in (Figure 9). They still allow users to view content on their website but ask them politely to disable Ad-blockers, thereby enabling availability of content.
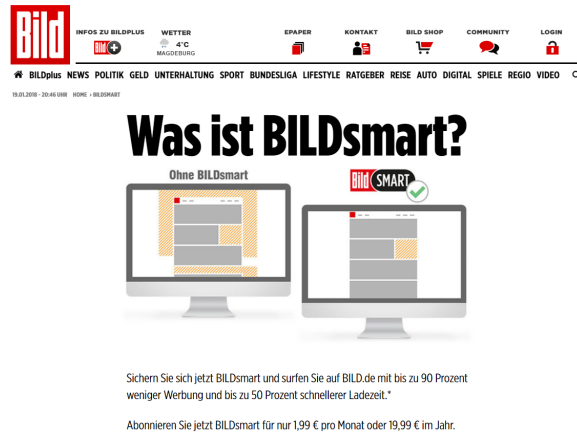
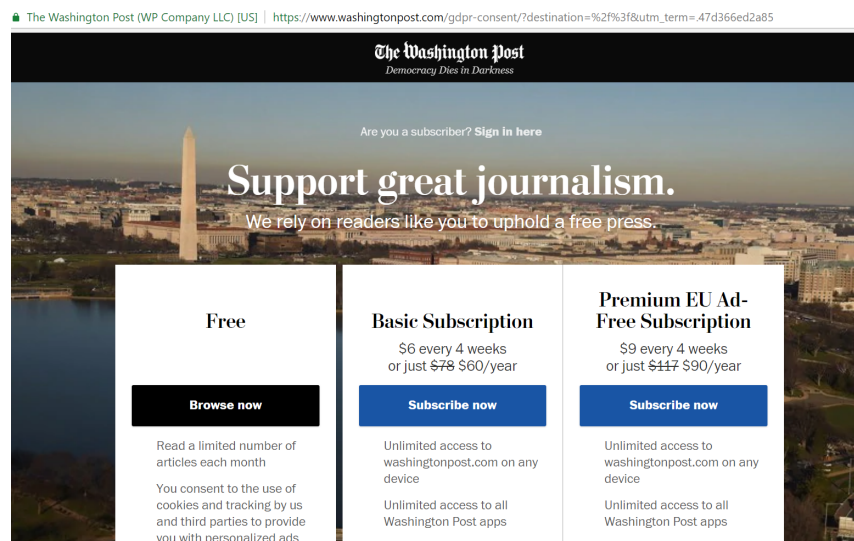**Fig. 6.** Bild.de ask users to pay to view up to 90 % less adverts.

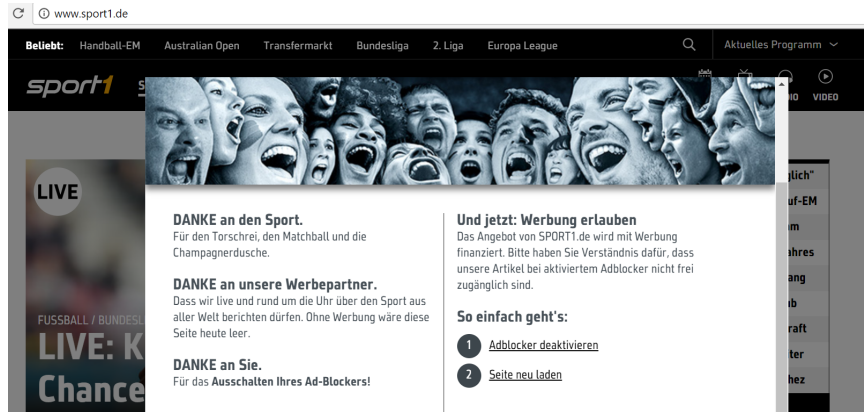**Fig. 7.** The Washington Post asks users to pay to view content based on region.

Fig. 8. Sport1.de do not allow content to be viewed until Ad-blocker is disabled.

Fig. 9. t3n.de asks users to remove Ad-blockers but do not block content.

5.3   Geographical Comparison

In their paper, Haris et al. [1] identified 686 websites from the Alexa top 100K websites that deploy Anti Ad-blockers. This list mostly details websites concerning the United States of America or China. Only a few websites from this list such as `www.google.de` and `www.amazon.de` belong to the Germany region.

Rishabh et al. [4] in their paper note that 6.7% of websites in a specific category such as News deploy Anti Ad-Blocker.

In our analysis, we observe that 1.9% of the 418 identified websites in Germany region employ Anti Ad-blockers. For DACH region, this percentage stands at 1.4% for 809 websites. For News category, this percentage changes to 3.1% out of the 357 websites. Our methodology contains various limitations which is described in the next section.

5.4   Limitations

Our methodology cannot detect all websites that employ Ad-block detection due to some limitations. They are described below:

- We remove all anti Ad-block filters in EasyList. This would probably not be the default settings used by users across the goal. But as we are attempting to quantify websites which employ Anti Ad-blockers, we feel that this configuration will give us a clearer picture.
- We only check anti Ad-blockers on the home page and not on any linked or a sub page.
- We look at websites that detect Ad-blockers and make any HTML or DOM changes to the browser page as using Anti Ad-blockers.
- We use Adblock Plus for our measurements but there are other Ad-block tools which is not part of our study.
- We also don't consider style features and cosine similarity features used by Haris et al. in their paper [1].

6   Impact of Anti Ad-blockers

6.1   Economical Impact

Revenue Generation for Ads. In a 2015 report titled "Digital advertising in Europe - Statistics & Facts", The Statistics Portal [8] cited "online advertising revenues worldwide amounted to about $170B, a figure that is expected to grow to more than $330B by 2021. In a 2017 global comparison, the United Kingdom, Germany and France ranked among the largest online advertising markets in the world (Figure 10), with digital ad revenues of $11.72B, $7.37B and $5.13B, respectively. In Germany, the 2017 revenues stood at $7.37B."
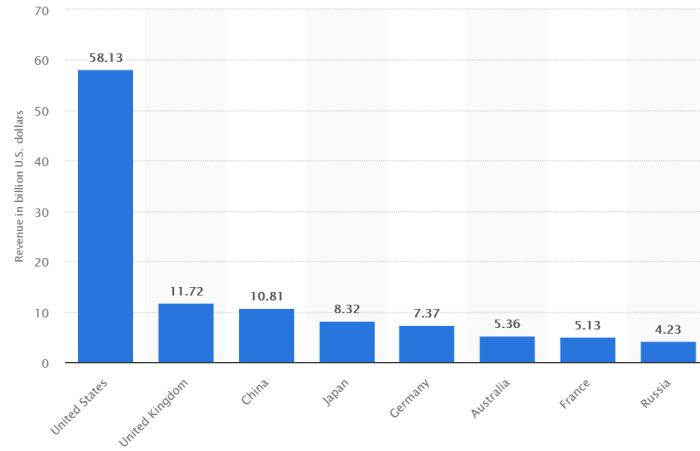
Fig. 10. Largest online advertisement markets in 2017 as published in the The Statistics Portal report titled "Online advertising revenue". [9]

6.2 Legality Aspects

European Union. Appendix A details the Article 5.3 of the European Union's ePrivacy Directive. Under this EU law and the recently introduced GDPR regulation (described in a later section), publishers must ask for permission before accessing a user's personal information, similar to how websites must ask for permission to store cookies on user devices. But as publishers are only detecting ads delivery via HTML or DOM elements, they are complying with the ePrivacy directive.

Germany. Media groups Süddeutsche Zeitung, Pro-Sieben-Sat.1, and IP Deutschland, an RTL subsidiary recently fought a legal battle against Cologne-based Ad-blocking company Eyeo to ban Eyeo's Ad-blocking software Adblock Plus because of their revenue losses. According to a report by Wired.de [20] titled "Adblock Plus: Werbeblocker bleiben in Deutschland legal", Munich higher regional court ruled that the Ad-blocking software Adblock Plus is legal. The court also quoted that Eyeo is not dealing in "forbidden aggressive advertising" and that is why their software does not breach German laws. Although joining Eyeo's Adblock Plus whitelist is free, a few media groups are still filing cases against Eyeo because they are objecting that Eyeo is monetizing from advertisers joining these whitelists under the Acceptable Ads programme.

This court order comes is contrasting to another in Cologne. Digiday's article [21] titled "What Axel Springer's loss in ad-blocking suit means for UK publishers", describes the court judgment which states Adblock Plus must add Axel Springer to their Whitelist for free, but work within the confinements of the Acceptable Ads programme.

In a Hamburg judgement for Spiegel versus Eyeo, Eyeo cited [23] in the judgement "Urteil des Landgerichts Hamburg (Az. 315 O 293/15)", that as of August

2015: "Adblock Plus was installed on approximately 9.55 million browsers with German IP addresses, which accounted for around 5% of the computers in Germany used to access the Internet". The judges finally concluded that users do not wish to see undesired advertising and they also want safety from any malicious softwares that might take control of their data.

In a blog post, "Adblock Plus and (a little) more", detailing the court decisions [22] Adblock Plus is currently still legal in Germany and as such the judges have not favored using Ad-blocking detectors.

6.3   Ethical Aspects

The argument of Anti Ad-blocking ensues on the topic of maintenance of websites and henceforth businesses. Users of Ad-block tools are unaware as to how websites keep their business running. For most of such "free" sites, their revenue is generated through advertisements. Few websites use subscription services while others have non Internet based revenues model.

Whatever be the case, labor must be paid. It must be in accordance to what a writer, producer, musician, developer or any professional writes, creates or develops. Websites are paid according to impressions which are measured by metrics such as pay per click or cost per click on the advertisements. Small scale business are most likely to suffer because of their small workforce. Also, with increasing costs of living worldwide, publishers rely heavily on such revenue sources.

The trend among publishers is increasing with respect to the detection of Ad-blockers. They are asking users to either completely disable Ad-blockers or to pay money and subscribe to an ad-free version. This poses an ethical dilemma to users, on one hand to use an Ad-block and stopping revenue income for publishers and on the other hand for publishers, to use an Anti Ad-blocker to prevent users from blocking ads.

6.4   Impact of GDPR

The General Data Protection Regulation (GDPR) came into effect on the 25th of May, 2018 after a two year transitioning period in the European Union. The regulation carries a important changes to how user's personal data must be handled by companies. The GDPR does not allow companies to store or use user's personal data unless explicitly agreed by the user. Cookies, for example, are also considered to be part of this data. Users also have the right at any time to ask companies to delete their complete data.

Section 26 of the GDPR document is cited in Appendix B. As we have analysed in our discussion so far, detecting Ad-blockers work mostly on the concept of identifying HTML and DOM elements which are completely unrelated to the users. These HTML elements cannot be identified to any user and there is no such transmission of user's personal data to any server, since it relies on client browser elements, which would lead to a non-compliance of the GDPR.

Many websites such as `npr.org` are providing only text content or country redirecting to an ad free EU version when users do not agree to provide access to their data.

Of course, with advances in Ad-blocking detection techniques such as user targeted ads by collecting cookie information, there may be a case where a technical implementation could ideally violate the regulations of GDPR, but that is not in the scope of this paper.

7   Alternatives to Anti Ad-blocking

7.1   Acceptable Ads Programme

Acceptable ads programme (by Adblock Plus) [18] lays down a technique that is focused on effecting ads on websites. They define a certain set of criteria or rules that ensure that ads placed on websites are not annoying to users, disrupt or distort any of the primary web page content and are transparent such that there are no popups or any unseen advertisements that could potentially lead to nonacceptance of an advertisement on a website.

7.2   Whitelists

Websites that comply to the Acceptable Ads programme can get their websites "whitelisted" with Adblock Plus. This is very similar to anti-viruses marking some applications as "safe" or "non-malware" and allowing them to run, while blocking other potentially harmful ones. Publishers who employ Anti Ad-blockers ask users to whitelist their websites upon detection of an Ad-blocker. Users can by themselves Whitelist most websites but they usually do not change the default configuration.

7.3   The rise of Anti Anti Ad-blockers

Anti Ad-block killers are on the rise. They deploy tricks into thinking that the user is not using an Ad-blocker. The Ad-blocker blocker lets the user keep the Ad-blocker on, making everything look normal.

One such Anti Ad-block killer is the AAK: *https://github.com/reek/anti-adblock-killer* [17]. It is a JavaScript script with a default filter list similar to that of AdBlock Plus. The AakList filter list can also be configured on various Ad-block plugins.

8   Conclusion and Future Work

Our analysis involves quantitative and qualitative analysis, which builds on the work presented by Rishabh et al. [4] and Haris et al. [1]. Our focus is mostly in the Germany and DACH region and a specific category which is News.

In our findings, we observed a range of 1.4-1.9% of websites deploy Anti Ad-Blockers when it comes to a country or a specific region such as DACH and 3.1% when it came to a specific category such as News. Rishabh et al. [4] in their paper found 6.7% of websites in a category such as News deploy Anti Ad-Blocker. We also noted that Bayes Naive classifier performs better in our analysis than the analysis presented by Haris et Al. [1].

Our methodology depends on HTML and DOM element changes, which introduces a number of limitations. We discuss different characterizations of Anti Ad-Blocker detection responses and provide an economical and ethical overview. We also discuss on the legal aspects of Anti Ad-Blocker which is still legal according to the latest rulings in the court.

Our work concludes with looking at how GDPR impacts the Anti Ad-Blocker industry and we also provide insight into some alternatives to Anti Ad-Blocking which includes Acceptable Ads Programme and Whitelists.

8.1 Future Work

Future work for this area would include conducting a more exhaustive evaluation of websites in Germany and also refining our methodology to account for the fast-changing world of Ad-blockers and Anti Ad-blockers.

8.2 Acknowledgements

The authors would like to thank Professor Jens Grossklags and fellow students taking part in the seminar and for providing valuable inputs and insights.

8.3 Source code and data release

Source code for this project can be found at the link: `https://github.com/RohitPanda/seminar-blocking-adblockers`

Appendix A ePrivacy Directive 2002/58/EC of the European Parliament and of the Council

Article 5.3 of the European Union's ePrivacy Directive [19] states: "Member States shall ensure that the use of electronic communications networks to store information or to gain access to information stored in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned is provided with clear and comprehensive information in accordance with Directive 95/46/EC, inter alia about the purposes of the processing, and is offered the right to refuse such processing by the data controller. This shall not prevent any technical storage or access for the sole purpose of carrying out or facilitating the transmission of a communication over an electronic communications network, or as strictly necessary in order to provide an information society service explicitly requested by the subscriber or user."

Appendix B   GDPR Regulation (EU) 2016/679 of the European Parliament and of the Council

Section 26 of the GDPR document [5] states, "The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes."

References

1. Haris Mughees, Muhammad, Qian, Zhiyun, Shafiq, and Zubair. Detecting Anti Ad-blockers in the wild. proceedings on privacy enhancing technologies (2015)
2. Garimella, Kiran, Kostakis, Orestis, Mathioudakis, and Michael. Ad-blocking: A study on performance, privacy and counter-measures (2017)
3. Interactive Advertising Bureau. Ad block detection code access request (2017) `https://github.com/InteractiveAdvertisingBureau/AdBlockDetection`
4. Rishab Nithyanand, Sheharbano Khattak, Mobin Javed, Narseo Vallina-Rodriguez, Marjan Falahrastegar, Julia E. Powles, Emiliano De Cristofaro, Hamed Haddadi, and Steven J. Murdoch. Ad-blocking and counter blocking: A slice of the arms race (2016)
5. Full text of the General Data Protection Regulation in English (2018) `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX\%3A32016R0679`
6. Google chrome adblock plus. https://chrome.google.com/webstore/detail/adblock-plus/cfhdojbkjhnklbpkdaibdccddilifddb/support?hl=en-GB
7. Pujol, E., Hohlfeld, O., Feldmann, A.: Annoyed users: ads and ad-block usage in the wild. In: Proceedings of the 2015 ACM Conference on Internet Measurement Conference, pp. 93–106 (2015)
8. Digital advertising in Europe - Statistics and Facts (2017) `https://www.statista.com/topics/3983/digital-advertising-in-europe/`
9. Online advertising revenue in major online advertising markets in 2017 (2017) `https://www.statista.com/statistics/246570/largest-online-advertising-markets/`
10. PageFair: 2017 Adblock Report (2017) `https://pagefair.com/blog/2017/adblockreport/`
11. Desktop Ad Blocking Continues Unexpected Decline in Germany (2016) `https://www.emarketer.com/Article/Desktop-Ad-Blocking-Continues-Unexpected-Decline-Germany/1014768?mod=djemCMOToday`
12. EasyList. `https://easylist.to/easylist/easylist.txt`
13. EasyPrivacy. `https://easylist.to/easylist/easyprivacy.txt`
14. Filter Lists. `https://filterlists.com/`
15. 9% of Popular Websites Use Anti Ad-block Scripts (2018) `https://www.bleepingcomputer.com/news/technology/9-percent-of-popular-websites-use-anti-adblock-scripts/`
16. Anti AdBlock Killer List. `https://raw.githubusercontent.com/reek/anti-adblock-killer/master/anti-adblock-killer-filters.txt`

17. Anti Ad-block Killer `https://reek.github.io/anti-adblock-killer/`
18. Acceptable Ads Criteria `https://acceptableads.com/en/about/criteria`
19. Article 5.3 of the European Union's ePrivacy Directive (2002) `http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32002L0058`
20. Adblock Plus: Werbeblocker bleiben in Deutschland legal (2017) `https://www.wired.de/collection/business/gerichtsurteil-adblock-plus-verlage`
21. What Axel Springer's loss in ad-blocking suit means for UK publishers (2016) `https://digiday.com/uk/axel-springers-loss-ad-blocking-suit-means-uk-publishers/`
22. Adblock Plus and (a little) more (2016) `https://adblockplus.org/blog/it-s-still-perfectly-legal-to-block-ads-in-germany-after-latest-court-victory-over-spiegel`
23. Urteil des Landgerichts Hamburg (Az. 315 O 293/15) (2016) `http://www.landesrecht-hamburg.de/jportal/portal/page/bsharprod.psml?showdoccase=1&doc.id=JURE170021729&st=ent`
24. Most Popular Websites according to Alexa country rankings `https://www.alexa.com/topsites/countries/US`
25. Selenium WebDriver. `http://www.seleniumhq.org/projects/webdriver`
26. BeautifulSoup library. `https://www.crummy.com/software/BeautifulSoup/`
27. Weka Toolkit. `https://www.cs.waikato.ac.nz/ml/weka/`
28. Top sites ranking for all categories in Germany `https://www.similarweb.com/top-websites/germany`
29. Top Websites - Germany `https://www.quantcast.com/top-sites/DE`
30. Information Gain Ratio: T. Mitchell. Machine Learning. Mc-Graw-Hill (1997)
31. Selenium Chrome Webdriver `https://github.com/SeleniumHQ/selenium/wiki/ChromeDriver`
32. Reek Github issues `https://github.com/reek/anti-adblock-killer/issues`