

Ad-blocking: A Study on Performance, Privacy and Counter-measures

Kiran Garimella

Aalto University

Helsinki, Finland

kiran.garimella@aalto.fi

Orestis Kostakis

Aalto University

Helsinki, Finland

orestis.kostakis@aalto.fi

Michael Mathioudakis

Aalto University

Helsinki, Finland

michael.mathioudakis@aalto.fi

ABSTRACT

Many internet ventures rely on advertising for their revenue. However, users feel discontent by the presence of ads on the websites they visit, as the data-size of ads is often comparable to that of the actual content. This has an impact not only on the loading time of webpages, but also on the internet bill of the user in some cases. In absence of a mutually-agreed procedure for opting out of advertisements, many users resort to ad-blocking browser-extensions.

In this work, we study the performance of popular ad-blockers on a large set of news websites. Moreover, we investigate the benefits of ad-blockers on user privacy as well as the mechanisms used by websites to counter them. Finally, we explore the traffic overhead due to the ad-blockers themselves.

ACM Reference format:

Kiran Garimella, Orestis Kostakis, and Michael Mathioudakis. 2017. Ad-blocking: A Study on Performance, Privacy and Counter-measures. In *Proceedings of WebSci '17, Troy, NY, USA, June 25-28, 2017*, 4 pages.

DOI: <http://dx.doi.org/10.1145/3091478.3091514>

INTRODUCTION

Ad-blocking is now a widespread practice among web users. According to current estimates, it is employed by around 200 million desktop and 300 million mobile users, a user base that still grows at 40% annually [7]. The same estimates indicate that ad-blocking translates into significant losses for publishers, reaching 41 billion US dollars for 2016. It is therefore not surprising that there is an arms race between publishers and advertisers, on one hand, and ad-blocking tools (or ‘ad-blockers’), on the other¹. In this context, it is clear that understanding the mechanisms implemented by ad-blockers and their effects is of wide and continuous interest.

In this paper, we study a number of popular ad-blockers and analyze their performance in desktop and mobile settings for a large number of webpages. Specifically, we employ the ad-blockers on two sets of webpages: one consisting of the front pages of popular news websites; and another consisting of around 30,000 web articles.

¹<https://techcrunch.com/2016/08/11/facebook-blocker-blocking-ads-blocked-by-blockers/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebSci '17, Troy, NY, USA

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM.

978-1-4503-4896-6/17/06...\$15.00

DOI: <http://dx.doi.org/10.1145/3091478.3091514>

For each webpage, we collect a number of measures that describe the browser workload with and without the use of an ad-blocker – including, for instance, the amount of data loaded, the load-time of webpages, or the number of data requests. Our results indicate that ad-blockers decrease the data consumption significantly, though the benefits in terms of load times are limited. Moreover, by taking a deeper look into the set of requests incurred by websites and ad-blockers, we identify cases where the websites attempt to counter the ad-blockers, but also requests that the ad-blockers incur on their own. Finally, we investigate the extent to which ad-blockers prevent the transfer of user-tracking information, resulting in privacy benefits for users. When comparing a range of different ad-blockers, we find that uBlock performs the best, in terms of both data savings and user-tracking.

RELATED WORK

Online Advertising Advertising has emerged as one of the main sources of revenue on the web [2]. On one hand, advertisers claim that ads on the web help keep most parts of the web free for consumers [1]. On the other hand, consumers find advertising annoying and obstructive. In one study, the New York Times analyzed the top 50 news websites² and found that more than half of all data came from ads.

Moreover, many users perceive web ads as a threat to privacy [3] and tracking is cited as one of the most common reasons for users to install ad-blockers [6]. This is corroborated by many studies that show large scale tracking behavior on the web [3, 9]. For instance, Yu et al. [12] find that 95% of the pages visited contain 3rd party requests to potential trackers and 78% attempt to transfer unsafe data. To counter tracking, ad-blockers also remove tracking buttons (such as Facebook’s ‘like’ button) and protect their users from known malware domains.

Ad blocking In [4], the authors propose methods to detect the users who have installed an ad-blocker and characterize ad-block usage for a large set (2 millions) of users. In addition, they give details on demographics and geographic penetration of ad-blockers on the web. In [8], the authors analyze the use of two ad-blockers, AdblockPlus and Ghostery, on web traffic data from a European ISP. They show that there is a drop in requests to third party services when using ad-block and estimate that around 18% of the traffic is due to ads. There are a few differences between our study and theirs: (i) their study handles only two ad-blockers, (ii) with a constant battle between advertisers and ad-blockers, the ad-blocking and counter ad-blocking scene has been changing over the last few years, and so have some of the related performance measurements.

²<http://www.nytimes.com/interactive/2015/10/01/business/cost-of-mobile-ads.html>

(iii) our study handles in-html ads, which are used by advertisers nowadays to bypass ad-blocking.

In [11], the authors classify third party trackers into various categories including ad trackers, analytics, social, etc and compare the performance of various ad-blocking tools with respect to blocking these third party requests. They find that there is a lot of variance in the type of requests that are blocked by each ad-blocker. They then look deeper into the individual default and possible configurations of these ad-blockers and study the changes in blocking capabilities with the different settings.

There have been privacy concerns with ad-blockers too. Many popular ad-blockers (including AdBlock and AdblockPlus) participate in the *Acceptable Ads* program³, allowing ‘non-intrusive’ ads to go through. For an extensive study on the issue see [10].

To counter the effects of increasing ad-blocking, advertisers are relying on counter-ad-blocking tools. In [5], the authors study their use on the most popular five thousand websites and find that at least 6.7% of these sites use them.

DATA

We analyze two datasets, chosen to capture potential differences between content-heavy homepages of large websites and individual webpages.

Top150 This is a list of the top 150 news websites, as ranked by Alexa⁴. For each of them, we have a URL that points to the website homepage, for both its desktop and mobile version (unless the latter does not exist).

GDELT A list of 30,000 URLs pointing to different news articles published on a single day (November 8, 2016). The list was obtained from GDELT⁵, a project that collects news stories from all around the world over the years.

We load the webpage of each URL with a clean instance of Chrome browser (no cached content or extensions), using the Selenium Python library⁶ on a Macbook pro with 8 cores, with no other major processes running. On each load, we capture the HAR file for the load. The HAR (HTTP Archive File) is a JSON-formatted file that captures the interactions between the browser and the website, including network requests, types and size of objects, and load times.

We load the same URL in six browser modes, all simultaneously: a vanilla mode (no ad-blocker), and one mode for each of five ad-blockers. The ad-blockers are AdBlock⁷, AdblockPlus⁸, Ghostery⁹, uBlock¹⁰ and Privacy Badger¹¹ – chosen from the most popular ad-blockers on the Chrome Store. For the Top150 dataset, we also load the mobile version of the page (as loaded on Google Nexus 5) using Selenium’s mobile emulation tool¹². Note that, though we

³<https://acceptableleads.com/>

⁴<http://www.alexa.com/topsites/category/News/Newspapers>. We processed the list manually and removed a few which were not standard news sites (e.g. reddit.com)

⁵<http://gdeltproject.org/>

⁶<http://selenium-python.readthedocs.io/>

⁷<https://getadblock.com/>

⁸<https://adblockplus.org>

⁹<https://www.ghostery.com/>

¹⁰<https://www.ublock.org>

¹¹<https://www.eff.org/privacybadger>

¹²<https://sites.google.com/a/chromium.org/chromedriver/mobile-emulation>

report the data about mobile websites with and without ad-block, the ad-blockers used were for the desktop version. There are no ad-blockers for the mobile version of Chrome and none for any of the Android browsers. Our datasets and code can be accessed here¹³.

FINDINGS

This section details our findings on the efficiency of ad-blockers, the ad-blocker benefits on user privacy, the counter-measures used by advertisers, and finally the traffic load that ad-blockers incur on their own.

Ad-blocker Efficiency

For each URL and setting (desktop or mobile, with or without ad-blocker), we collect measures that describe the average performance of the browser in loading the webpages of a dataset. The measures include: the number of distinct domains for all HTTP requests performed during loading; the maximum number of threads that run concurrently at any point; the total number of HTTP requests; the total size of downloaded and uploaded data; the cumulative time for loading the webpage, obtained by summing up the durations of all requests; and, finally, the wall-clock time. We acquire these statistics by parsing the HAR files.

The number of distinct domains indicate the number of different parties that acquire information of the user, such as IP address and the user-agent string (contains device type, name and version of browser and OS, etc). The number of HTTP requests and maximum number of concurrent threads give an indication of the load on the user’s machine. The amount of transferred data, and the cumulative and wall-clock time affect the user experience.

Figures 1 and 2 show the results for the desktop and mobile versions of the Top150 webpages; and Figure 3 for the GDELT webpages. Numbers represent the ratio of the measures over the vanilla mode – for example, a value of 0.6 for the number of domains means that using an ad-blocker loads 60% the number of domains compared to using no ad-blocker.

We find that: (i) All ad-blockers except Ghostery give around 25-34% savings in the amount of data transferred (on average). This is a bit higher than the 18% saving reported by [8] and 13-34% reported by [11]. Ghostery is the exception because it is not an ad-blocking tool per se, but gives a choice for a user to track who is tracking them and block those of their choice. So by default Ghostery does not block any content. We still include it in all our measurements, as it still makes sense to study it in other analysis cases that follow in this section. (ii) On average, there is an *increase* in the wall clock time when using an ad-blocker, even as the cumulative time decreases for some of the ad-blockers. This means that ad-blockers incur overhead, but block ad-related threads that were meant to run in parallel while loading content; this overhead is not necessarily experienced by the user, since various scripts can often run in the background, after the important content has loaded. (iii) uBlock gives the best performance, both in terms of data and time saved.

¹³<https://users.ics.aalto.fi/kiran/adblock/>

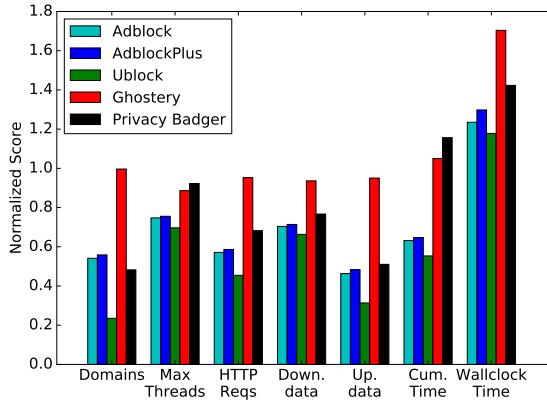


Figure 1: Performance benchmark when using an ad-blocker on the desktop version of Top150, normalized over the vanilla mode.

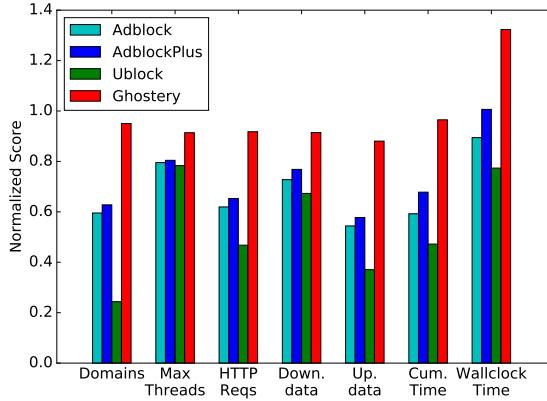


Figure 2: Performance benchmark when using an ad-blocker on the mobile version of Top150, normalized over the vanilla mode. Privacy Badger was discarded because of inconsistencies in the data collection.

Benefits for User Privacy

By examining the HTTP requests performed by the browser when loading a webpage, we notice that a significant amount of these are intended for tracking the users' behaviour. We attempt to quantify the potential privacy hazard for the user when accessing a webpage by analyzing the parameters passed through these requests. Specifically, to measure the impact of potentially privacy invading requests when using no ad-blocker, we manually examined the tracking parameters from the request headers and found nine frequently used tracking parameters that accompanied a large set of requests. The parameters we are: 'pixel', 'track', 'google_gid', 'user_id', 'partnerid', 'partnerUID', 'partner_device_id', 'uid', 'user_cookie'. The presence of these parameters offers a high-precision method to identify user-tracking, although we currently do not quantify its recall.

Figure 4 shows the fraction of requests containing one, two and three tracking parameters in a request when using an ad-blocker, again normalized over the vanilla mode. We can see from the figure

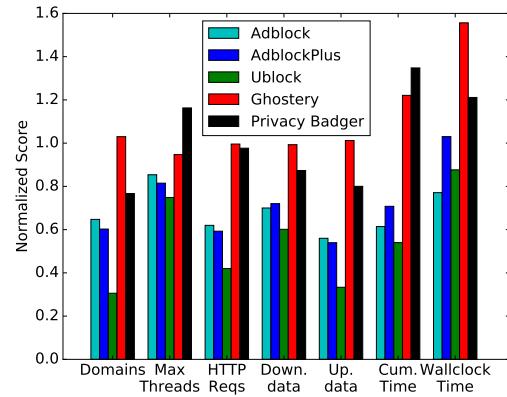


Figure 3: Performance measures when using an ad-blocker on the GDELT dataset, normalized over the vanilla mode.

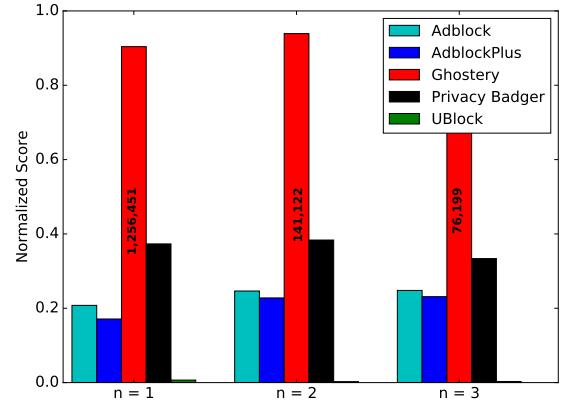


Figure 4: Fraction of distinct requests containing tracking keywords for one, two and three tracking parameters. The black text in the red bar shows the total number of requests containing these tracking parameters in the vanilla mode.

that, again, uBlock achieves near-perfect performance, whereas other ad-blockers (except Ghostery) block about 60-80% of requests with such parameters. The performance remains the same, when considering requests with different numbers of tracking parameters, though the total number of requests goes down exponentially.

Blocking the Ad-blockers

Online ventures relying on advertising try to counter to the ad-blocking tools [5]. In our analysis, we witness several techniques for countering ad-blockers, with different goals. In certain cases, the goal is simply to detect whether the browser has an active ad-blocking extension. Upon detection, a message is rendered on the webpage and notifies the user about the tool. This message may simply notify the user that the tool interferes with the content and the user-experience. Other times, the message blocks the user from accessing the actual content, until they have turned off the tool. In more extreme cases, the goal is to circumvent the ad-blocking tool altogether.

Table 1: Additional Domains loaded by Ad-blockers

Ad-blocker	Extra domains loaded
AdBlock	mixpanel.com, stripe.com
Adblock Plus	quantserve.com, crwdcntr.net
Ghostery	selectmedia.asia, streamrail.com adlooxtracking.com
Privacy Badger	cdnjs.com
uBlock	Twitter widgets

We examine again the HTTP requests performed when loading a webpage and look for specific Javascript modules designed to counter ad-blocking, such as `blockadblock.js`, `inn-anti-adblock`, and `adblock html cache buster`. When webpages from the GDELT dataset are loaded in the vanilla mode (no ad-blockers), we find more than 3,260 webpages (10.8%) with explicit attempts to determine if any ad-blocking tool is used. When AdBlock or AdblockPlus is activated, the number of webpages rises to 10,097 and 10,226 (33%), respectively, indicating that a large portion of webpages actively track the usage of these ad-blockers. When the other tools are used, the number of webpages with such requests is not as high; 15, 9, 8% of the webpages for Ghostery, Privacy Badger, and uBlock respectively. The variation in numbers between AdBlock and AdblockPlus, on one hand, and the rest of the ad-blockers, on the other, can be explained by the popularity of the former – but also the effectiveness of the ad-blockers in blocking such requests.

Requests due to Ad-blockers

In this section, we try to dissect what *additional* requests are made when using an ad-blocker. Consider the set of HTTP requests required to render a webpage in the vanilla mode, denoted by \mathcal{V} . Let \mathcal{B} be the set of requests when using an ad-blocking tool, then we define $\mathcal{A} = \mathcal{V} \setminus \mathcal{B}$ as the identified advertisements, and $\mathcal{C} = \mathcal{V} \cap \mathcal{B}$ as the “true” content of the page. Finally, $\mathcal{E} = \mathcal{B} \setminus \mathcal{V}$ is “extra” content loaded due to or by the ad-blocking tool. Unsurprisingly, the most popular domains found in \mathcal{A} are `googlesyndication.com` and `doubleclick.net`, in accordance to previous studies, and blocked (characterized as ads) by all tools.

A more interesting question is to identify the “extra” \mathcal{E} requests that incur exclusively at the presence by ad-blockers and to look for potential leakages in information¹⁴. The domains which most frequently occur only in \mathcal{E} are shown in Table 1. In many cases, we notice that \mathcal{E} contains well-known ad-serving or visitor-tracking domains. This clearly demonstrates that the tools are not perfect and are sometimes bypassed.

For AdBlock, we notice that `mixpanel.com` appears in 5,125 (17%) of the webpages when using AdBlock, versus 170 (< 1%) of original pages (vanilla mode). This is a Google Analytics-like website that tracks visits by users of AdBlock. Similarly, `stripe.com`, an online payments platform through which users can donate to AdBlock is added to 2,616 (8%) of the webpages, compared to 40 (< 0.1%) with out ad-blocking. This is a Google Analytics-like website that tracks visits by users of AdBlock.

¹⁴e.g. <http://thehackernews.com/2016/11/web-of-trust-addon.html>

For Adblockplus, we notice an increase for `quantserve.com` (an audience-measurement website) from 14% to 21% and for `crwdcntr.net` (another audience-measurement website) from 21% to 29%. For Ghostery, we notice `selectmedia.asia` (a video-ad serving company) appears in 1,374 (4.6%) webpages’ requests, compared to 47 (< 0.1%) in the vanilla mode, `streamrail.com` (another video-ad serving company) 1,488 times (5%) versus 63, and `adlooxtracking.com` (ad tracking company) 2,190 times (7%) versus 240 (< 1%). For PrivacyBadger, `cdnjs.com` (a repository for open-source web-libraries) stands out with 2,130 webpages (7%). For uBlock we did not witness any significant increase for suspicious domains; but we did notice an increase for Twitter widgets from 11% to 20% of the webpages.

Our preliminary findings indicate that though ad-blockers block external ads and third party trackers to varying degrees, they introduce various tracking services of their own. This is a novel finding and has potential impact on users privacy. We leave a detailed analysis on this for future work.

Conclusions This paper provides a first look at the performance and privacy aspects of popular ad-blocking tools. Based on our analysis, we conclude that (i) uBlock has the best performance, in terms of ad and third party tracker filtering, and least privacy tracking, (ii) The time to load pages is not necessarily faster when using adblockers, and (iii) this is partly the case due to additional trackers and libraries introduced by the adblocking tools. As ad-blockers try to monetise their business by selling ads themselves, our datasets and findings could be used (i) by users, to get an idea on which adblocker to choose, from an increasingly competitive set of providers, and (ii) by researchers, as building blocks for further analysis on privacy and tracking behaviors by ad-blocking tools.

Acknowledgements. This work has been supported by the Academy of Finland project “Nestor” (286211) and the EC H2020 RIA project “SoBigData” (654024).

REFERENCES

- [1] Paul Barford, Igor Canadi, Darja Krushevskaja, Qiang Ma, and S Muthukrishnan. 2014. Adscape: Harvesting and analyzing online display ads. In *WWW*. 597–608.
- [2] David S Evans. 2008. The economics of the online advertising industry. *Review of network economics* 7, 3 (2008).
- [3] Marjan Falahastegar, Hamed Haddadi, Steve Uhlig, and Richard Mortier. 2014. Anatomy of the third-party web tracking ecosystem. *arXiv preprint arXiv:1409.1066* (2014).
- [4] Matthew Malloy, Mark McNamara, Aaron Cahn, and Paul Barford. 2016. Ad Blockers: Global Prevalence and Impact. In *Proceedings of the 2016 ACM on Internet Measurement Conference*. 119–125.
- [5] Rishab Nithyanand and others. 2016. Ad-Blocking and Counter Blocking: A Slice of the Arms Race. *arXiv:1605.05077* (2016).
- [6] PageFair. 2015. The 2015 Ad Blocking Report. <https://blog.pagefair.com/2015/ad-blocking-report/> (2015).
- [7] PageFair. 2016. The 2016 Mobile Ad Blocking Report. <https://pagefair.com/blog/2016/mobile-adblocking-report/> (2016).
- [8] Enric Pujol, Oliver Hohlfeld, and Anja Feldmann. 2015. Annoyed Users: Ads and Ad-Block Usage in the Wild. In *IMC*. 93–106.
- [9] Sebastian Schelter and Jérôme Kunegis. 2016. On the Ubiquity of Web Tracking: Insights from a Billion-Page Web Crawl. *arXiv preprint arXiv:1607.07403* (2016).
- [10] Robert J Walls, Eric D Kilmer, Nathaniel Lageman, and Patrick D McDaniel. 2015. Measuring the Impact and Perception of Acceptable Advertisements. In *IMC*. 107–120.
- [11] Craig E Wills and Doruk C Uzunoglu. 2016. What Ad Blockers Are (and Are Not) Doing. In *HotWeb*. 72–77.
- [12] Zhonghao Yu, Sam Macbeth, Konark Modi, and Josep M Pujol. 2016. Tracking the Trackers. In *WWW*. 121–132.