What Ad Blockers Are (and Are Not) Doing

Craig E. Wills Doruk C. Uzunoglu

Computer Science Department Worcester Polytechnic Institute Worcester, MA 01609

WPI-CS-TR-16-02

June 2016

Abstract

The Web has many types of third-party domains and has a variety of available ad blockers. This work evaluates ad blocking tools for their effectiveness in blocking the retrieval of different categories of third-party content during download of popular websites. The results of this work demonstrate that there is much variation in the effectiveness of current ad blocking tools to prevent requests to different types of third-party domains. Non-configurable tools such as Blur and Disconnect provide only modest blockage of third-party domains in most categories. The tool uBlock generally demonstrates the best default configuration performance. By default, Ghostery provides no protection while Adblock Plus and Adguard provide minimal protection. They must be manually configured to obtain effective protection.

The behavior of Adblock Plus is particularly notable as usage data indicates it has an 85% share of the ad blocking tool market. Recent results based on network traces suggest that approximately 80% of these Adblock Plus users employ its default configuration. Construction of a "composite" ad blocker reflecting current usage of ad blockers and their configurations shows this composite ad blocker provides only a modest range reduction of 13-34% in the set of third-party domains retrieved in each category relative to not employing any ad blocker.

1 Introduction

The Web has evolved so that requests for most popular "first-party" pages by a browser result in a number of requests to "third-party" server domains. We found over 300 third-party domains (e.g. doubleclick.net) with a presence on at least 1% of popular Web pages and almost 100 third-party domains with at least a 5% presence on these pages. These third parties provide a variety of services such as tracking users, serving ads, performing site analytics, connecting with social media and providing interfaces to other services.

At the same time, a 2015 industry report indicates that ad blocking had grown by 41% globally and 48% in the U.S. in the past year [20]. Given the prevalence of different types of third parties and ad blockers, this work seeks to understand what ad blockers are (or are not) doing with these different types. The work makes a number of contributions:

- 1. measures the longitudinal use of different types of third parties by popular websites,
- 2. examines ad blockers and their filter lists for what third-party servers are blocked and allowed, and
- 3. evaluates ad blocking tools for their effectiveness in blocking the retrieval of different types of third-party content during download of popular websites.

Our work is builds on prior work in this area. Mayer and Mitchell [17] did a study on the effectiveness of available ad blocker tools in 2011. However, the set of available tools and their configurations has changed. In addition, the effectiveness results did not distinguish between different types of third-party trackers.

More recently, Pujol, *et al.* [21] examined the use of two ad blocking tools, Adblock Plus [3] and Ghostery [12]. As part of their work, they estimated 22% of the most active users employed Adblock Plus. Their results show a drop in HTTP requests to third-party servers using these tools, but it is not clear whether all, or only some, requests to specific third-party servers are blocked.

Walls *et al.* [27] performed an extensive examination of the Acceptable Ads program of Adblock Plus. The work looked at the growth of this program as well as measured filter behavior and surveyed user perception of acceptable ads.

Work in [18] characterizes the extent that users are exposed to third parties. It provides a table of ad blocking tools, but does not measure the effectiveness on different types of third parties. Earlier work by these researchers indicated that 20% of users had installed ad blockers in 2014 [19].

The remainder of this paper describes how the third-party domains examined in this work were identified along with their classification into categories. We go on to track the growth of these categories over time. The ad blocker tools and their filter lists used in this study are then described followed by evaluating the effectiveness of these tools for the different types of third parties. The paper concludes with a summary of the results as well as directions for future work.

2 Third-Party Domains

The initial part of our work involved understanding the range of third-party servers employed on popular websites. As a means to gather information on commonly used third-party servers

we employed a similar methodology as originally developed in [15]. That work used the Alexa website rankings [7] for the 100 most popular sites across different categories. Duplicate sites were removed resulting in over a 1000 popular websites, which were then retrieved while recording the set of URLs retrieved for each page.

In 2015, we repeated this methodology by first obtaining the 100 most popular websites across each of 15 categories as ranked by Alexa. The result is 1275 unique Web pages. In December 2015, we used the Selenium IDE [14] to drive a Mozilla Firefox browser to load each Web page while using the HTTP LiveHeaders [22] add-on to record all HTTP requests resulting in a response. The procedure was repeated twice with the two sets of results combined for reporting.

We processed the resulting data to determine the list of web servers contacted for each page. For purposes of analysis, servers are mapped to their Internet domain with multiple servers having the same domain (e.g. ads.doubleclick.net, stats.g.doubleclick.net) merged together. Web servers not in the same Internet domain as the originating page are considered to be third-party servers.

As done in [16], we also determine the domain of the Authoritative Domain Name Server (ADNS) for each server in our dataset. In some cases the ADNS of the server indicates is being provided by another entity. For example, the ADNS for metrics.cnn.com is nsl.omtrdc.net, which is part of Omniture owned by Adobe. On lookup, the ADNS returns the DNS CNAME alias cnn.122.207.net. Thus metrics.cnn.com is a "hidden" third-party server because syntactically it is part of the cnn.com, but its "root" domain is not. We use the root domain of a server in our analysis when a hidden third-party server is detected.

We found 316 third-party domains with a presence on at least 1% of popular websites. Furthermore, 97 third-party domains were found to be retrieved on at least 5% of these popular sites.

3 Third-Party Domain Classification

In the modern Web, third-party servers provide a variety of services. We determined a set of categories and classified these services using information on third-party services available from Ghostery [12], Abine [2] and TrendMicro [24]. The Ghostery categorization was used for most third parties with the other sites employed as needed. More details on the categories and classifications are available in [26].

We classified third-party domains into six categories, which are given below along with a short description and a sample of domains with a large presence on popular first-party sites.

- 1. **AdTrackers** (e.g. doubleclick.net, adnxs.com, quantserve.com): deliver advertisements, serve ads and track user activity across first-party sites.
- 2. **Analytics** (e.g. google-analytics.com, omtrdc.net, imrworldwide.com): provide data aggregation for first-party sites.
- 3. **Beacons** (e.g. googlesyndication.com, scorecardresearch.com): do not directly serve ads, but track and aggregate user activity across sites.
- 4. **Social** (e.g. facebook.com, twitter.com): are icons/links to connect user activity with social media sites.

- 5. Widgets (e.g. google.com, googleapis.com): provide page functionality.
- 6. **Other** (e.g. cloudfront.net, amazonaws.com, truste.com): provide other third-party services.

Using this classification, Figure 1 shows the proportion of third parties with at least a 5% and 1% presence of popular first-party sites in each of the six categories. Results for each set of third parties is similar with roughly 40% of third-party domains classified as AdTrackers. Another 15-20% of third parties are classified in each of the Analytics and Beacons categories. Widgets and Social third parties each make up less than 10% of the total with Other third parties constituting the remainder.

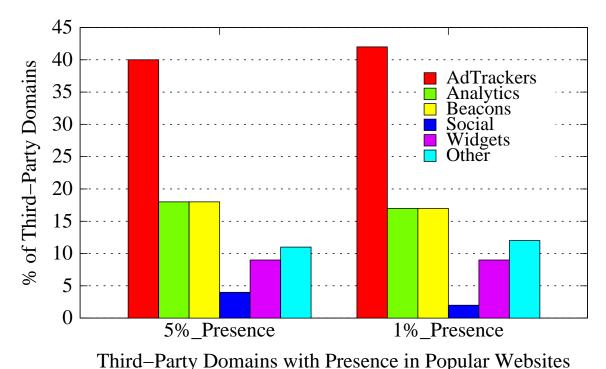


Figure 1: Proportion of Third-Party Domains in Each Category

4 Longitudinal Tracking of Third-Party Categories

While the breakdown of third-party domains into their respective categories is interesting, data that was originally gathered for [15] and [16] as well as data that have been gathered since allows us to track the presence of domains in these third-party categories on first-party sites over a ten-year period. Data on the presence of third parties were gathered at four epochs in time for first-party sites that were popular in one of Alexa's categories at that time. The results of this longitudinal tracking are shown in Figure 2.

The results show that third-party servers classified as AdTrackers have grown from a presence of 34% on popular sites in 2005 to 86% at the end of 2015. Analytics third parties have grown

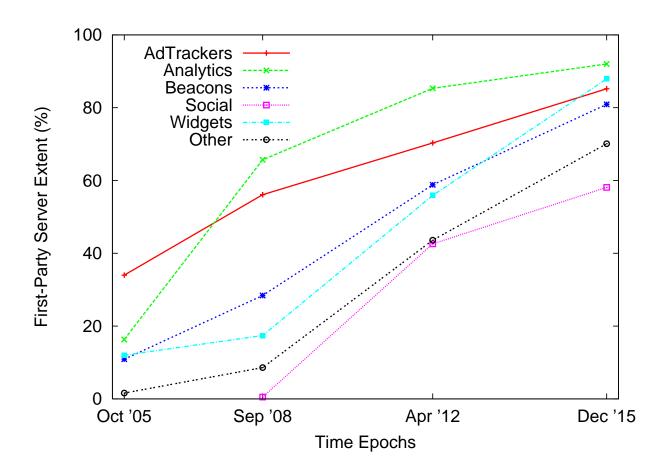


Figure 2: Longitudinal Presence of Third-Party Categories on Currently Popular Websites

even more dramatically from a presence of 16% in 2005 to 93% in late 2015. Servers classified as Beacons and Widgets are now present in over 80% of popular sites. Third-party servers classified in the Social category did not exist in 2005, but were present on 58% of popular sites by the end of 2015.

5 Ad Blockers and Filter Lists

Clearly third parties have grown to have a significant presence on popular websites. The content served by, and the functionality of, these third parties has become enough of a concern, whether because of privacy or annoyance considerations, that users are employing tools to prevent third-party content from being requested while browsing. A common type of tool for such prevention is an ad blocker extension for a Web browser.

5.1 Filter Lists

Each ad blocker is driven by a filter list, which is a set of syntactic matching rules to block (or allow) the retrieval of a URL by a browser. Ad blockers either embed a filter list as part of their implementation (e.g. Disconnect [10]) or allow users to load one or more public filter lists into the tool (e.g. Adblock Plus [3]).

Figure 3 shows the percentage of third-party domains with at least a 5% presence that are either blocked, allowed or have no action taken by a sampling of available filter lists. See [26] for actions taken by additional filter lists. The figure shows that the hpHosts [13] list blocks 91% of these domains with the Ghostery list blocking 78%. The EasyList list, which is the default for the Adblock Plus tool, blocks 58% of domains, but explicitly allows 11% of them. The Acceptable Ads list, also an Adblock Plus default, explicitly allows 26% of domains.

5.2 Ad Blockers for Study and Their Default Configurations

Figure 3 provides limited insight into how extensive the block or allow actions are on the most widely employed third parties. However, it does not provide specific actions on the types of third parties nor does it show the effectiveness of ad blockers deploying these filter lists to prevent different types of third parties from being contacted while downloading popular websites.

For measuring the effectiveness of current ad blocking tools we identify six ad block browser extensions for a detailed performance study. Each of these extensions is available for a variety of browsers. The six ad blocking tools and their default configurations are:

- 1. **Adblock Plus** [3] employs the EasyList and Acceptable Ads filter lists by default.
- 2. **Adguard** [5] by default employs an English filter list based on the EasyList filter and a Useful Ads list.
- 3. **Blur** [1] provides its own tracker blocking amongst other privacy features. The free version of the tool was tested, but the paid premium version does not appear to provide any additional tracking protection features.

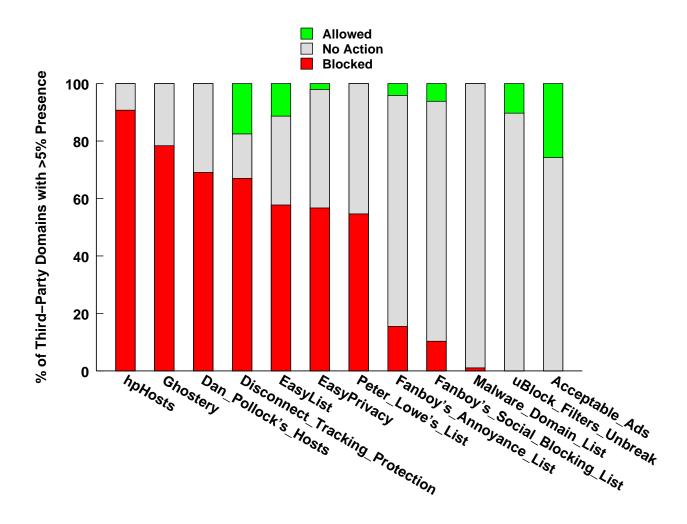


Figure 3: Ad Blocker Filter List Actions on Third-Party Domains with >5% Presence

- 4. **Disconnect** [10] employs its own Tracking Protection List, which cannot be changed by a user.
- 5. **Ghostery** [12] does not block any requests by default. It has its own filter list, which can be deployed in its entirety or blocking rules for different types of third parties can be user selected.
- 6. **uBlock** [25] by default employs its own blocking lists as well as EasyList, Peter Lowe's List, EasyPrivacy and a Malware Domain List. It also provides other lists that can be selected. It evolved from uBlock Origin, which yielded similar results in testing. We report only results from uBlock.

Table 1 shows the relative popularity of these ad block extensions by the combined users of the Chrome and Mozilla Firefox browsers. These results, compiled from usage numbers for each extension reported by Chrome [9] and Firefox [11], show that Adblock Plus is the predominant ad blocking tool currently in use with roughly 85% of users. Prior work [21] done in Germany based on network traces from a large European ISP found that 22% of the most active users employ the Adblock Plus browser extension. Combining these respective results suggests approximately 25% of users are employing an ad blocking browser extension. This number is consistent with a 2015 report by PageFair [20] on ad blocking, which reported ad blocking software usage at 24.7% in Germany. It indicated 16% of the U.S. online population blocked ads.

Table 1: Combined Users (Millions) of Chrome and Mozilla Firefox Ad Blocker Tool Extensions

Extension	# Users (M)	Share
Adblock Plus	75.4	85.2%
Adguard	2.0	2.2%
Blur	0.5	0.6%
Disconnect	1.2	1.4%
Ghostery	3.9	4.4%
uBlock & uBlock Origin	5.5	6.2%

6 Effectiveness of Ad Blockers on Popular Websites

Given that ad blockers are being used in significant numbers, the question is what do these ad blockers do and are they effective at doing it. Our measure of effectiveness is to determine if a request is sent to at least one third-party server of different types when popular first-party websites are retrieved by a browser, which is the metric used in results shown in Figure 2.

6.1 Methodology

Our approach for studying this effectiveness was to first re-run our tests in May 2016 using the same set of popular websites as determined in 2015. These tests were initially run with the Firefox

browser configured with no ad blocking software. The results provide a baseline for the impact of ad blockers on the presence of third parties during the download of popular first-party Web pages.

As expected, the May 2016 presence results for each category of third-party domains is similar to what is shown for December 2015 in Figure 2. We also used these results to update the category membership to include additional third-party domains with at least a 1% presence amongst this set of popular pages using the same procedure as done for our original classification.

6.2 Default Ad Blocker Configurations

The six browser extensions described in Section 5.2 were then tested using default configurations for each browser. In each case, the current version of the extension was installed and employed with current versions of the filter lists. Figure 4 shows the first-party presence results when retrievals of popular Web pages are made. The first result is a baseline when no ad blocker is installed with subsequent results shown for each ad blocker with its default configuration employed. Ghostery does not block any requests by default so its behavior mimics the no-ad-blocking results.

The results in Figure 4 show that all ad blockers have a relatively minor effect on domains in the Widgets category. The range of percentage reductions for the four remaining categories, relative to the baseline for the category, is shown as part of each label as an aid in comparing results.

A striking result in the figure is that Adblock Plus, with deployed default filter lists of EasyList and Acceptable Ads, yields a relatively minor reduction (3-21%) in the presence of third parties during the retrieval of popular Web pages across the set of categories. Adguard, with its default based on EasyList and allowing Useful Ads, provides similar results. At the other extreme, uBlock significantly reduces the presence of different categories of third-parties. For example, it results in a 7% presence (91% reduction) of AdTrackers with the presence of domains in the Beacons category dropping to around 2% (98% reduction) of popular first-party Web servers. Between these two extremes are the results for Blur and Disconnect, which modestly reduce the presence of third-party domains in these four categories from 22-62% and 54-78%.

While not shown in the figure, the results for Adguard are interesting in that during initial testing, its effectiveness results were comparable to those for uBlock, but it generated a request to its own server sb.adtidy.com for 95% of the first-party pages, presumably to collect statistics. However, on subsequent testing the tool's default performance changed to have effectiveness comparable to Adblock Plus with requests to its own server no longer present. We do not know why this change in default behavior occurred, but Figure 4 reflects these latter testing results.

6.3 Adblock Plus Configurations

Results in Figure 4 show that the default configuration for Adblock Plus, which includes the EasyList and Acceptable Ads (AA) filter lists, have a relatively minimal impact on reducing the extent of domains in each third-party category. Given the introduction and growth of the Acceptable Ads program as documented in [27], an obvious question is the impact of this program on the effectiveness results. Figure 5 shows results from different filter list combinations employed with Adblock Plus. Baseline and default settings results are replicated from Figure 4 for comparison.

The "EasyList" results (with Acceptable Ads turned off) in the figure show little difference when compared with the default "EasyList + AA" results indicating that the Acceptable Ads filter list have virtually no impact on our measure of effectiveness when combined with EasyList. The

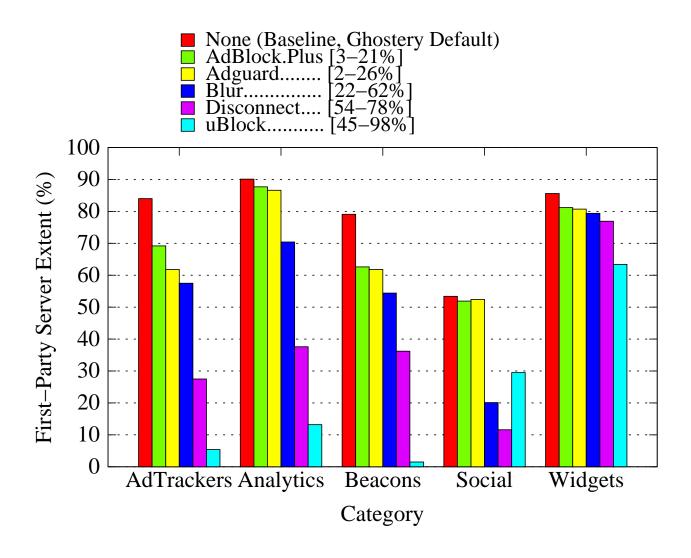


Figure 4: Third-Party Category Presence Employing Default Ad Blocker Configuration [Per-Category Reduction Relative to Baseline]

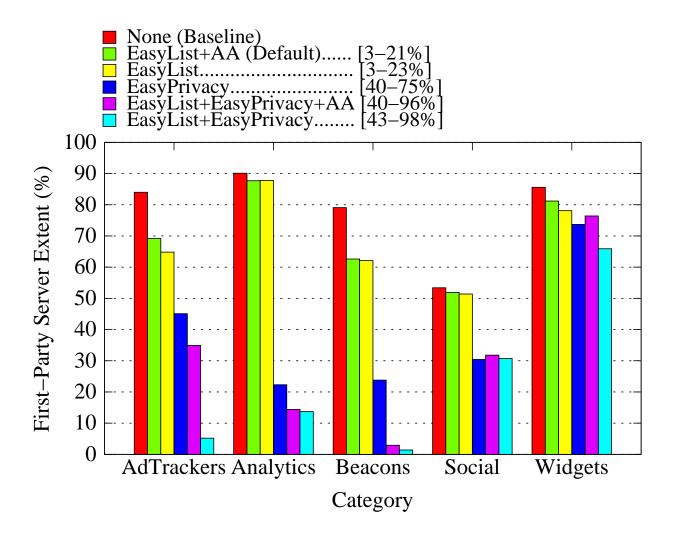


Figure 5: Third-Party Category Presence Employing Various Adblock Plus Configurations [Per-Category Reduction Relative to Baseline]

deployment of EasyList yields a relatively small reduction of third-party presence relative to the baseline whether or not the Acceptable Ads list is deployed in conjunction.

The results indicate that the default Adblock Plus filter lists may block "annoying" ads, but provide minimal protection with or without Acceptable Ads against third-party trackers. Documentation does indicate that Adblock Plus can "disable most tracking" [4] by employing the EasyPrivacy "privacy protection" filter list. Results for its deployment with no other lists are shown in Figure 5. The presence results are improved relative to default settings for all categories, but there is still a 45% presence (46% relative reduction) of AdTrackers domains.

The final two sets of results in the figure show the deployment of both the EasyList and EasyPrivacy lists along with and without the Acceptable Ads list. The results show a significant reduction (96-98%) of domains in the Beacons category to only a few percent presence, but with Acceptable Ads the presence of AdTrackers domains is still 35% (58% reduction). Only when both lists are deployed and the Acceptable Ads setting turned off are the results for Adblock Plus comparable to the best non-Social results for other ad blockers. These results are particularly notable, in that [21] estimates at most 20% of users deactivate the Acceptable Ads list and more than 85% of Adblock Plus installations employ EasyList, but not EasyPrivacy. Thus at most 15% of Adblock Plus users are receiving privacy protection comparable to the best ad blocking tools. These results also suggest approximately 80% of Adblock Plus users are employing the default configuration, which is consistent with research showing that most users do not change default software settings [8, 23].

6.4 Other Ad Blocker Configurations

Blur and Disconnect employ filter lists that cannot be changed by a user, and so other lists were not tested for these tools. By default the Adguard extension also employs a Useful Ads list, which "has nothing to do with" the Adblock Plus Acceptable Ads list [6]. We tested the Adguard extension with the Useful Ads not deployed and found there is no discernible difference from the default Adguard results shown in Figure 4. We did not test other available filter lists for Adguard.

Ghostery allows third parties to be blocked based on categories it defines. The current version of the Firefox Ghostery extension (version 6.2.0) identifies eight categories of trackers, which is an increase from the five categories in previous versions such as version 5.4.11 currently available for Chrome. Figure 6 shows testing results for three Ghostery configurations along with baseline results replicated from Figure 4. The first configuration blocks the Ghostery Advertising (which appears to be a combination of the former Advertising and Beacons categories), Site Analytics and Social Media categories. The second blocks all categories except for the Ghostery Essential category, which are "technologies that are critical to the functionality of a website." The third configuration blocks all categories. This final Ghostery configuration results in a 90-99% reduction in the presence of third parties for the four non-Widgets categories.

Figure 7 repeats the baseline and default results for uBlock as well as including results for hphosts and Fanboy lists present in Figure 3. Results for the hphosts list are generally a bit worse than the default for all categories except for the Beacons category where results are much worse. Results for the Fanboy lists show that these lists prevent little presence of third parties. The final set of results in Figure 7 are for the configuration when all available filter lists are selected. The results show that this configuration is the most effective uBlock configuration at reducing the presence of third parties by 81-99% across non-Widgets categories.

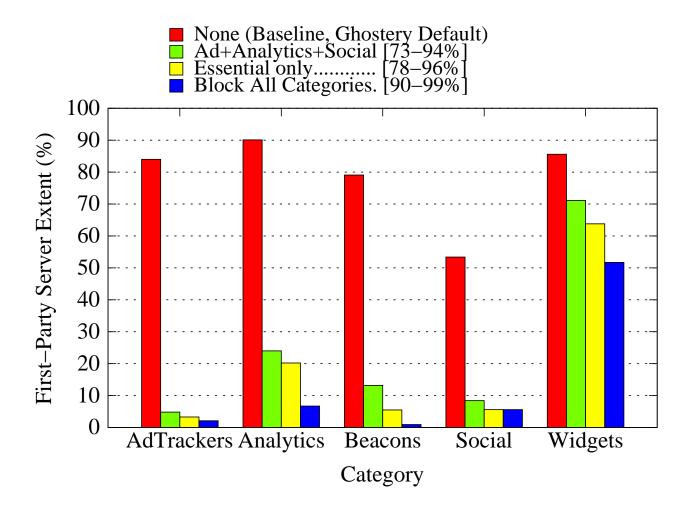


Figure 6: Third-Party Category Presence Employing Various Ghostery Configurations [Per-Category Reduction Relative to Baseline]

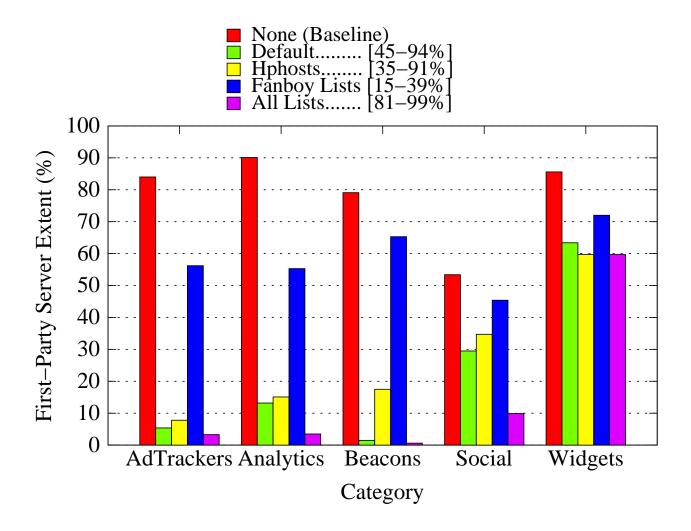


Figure 7: Third-Party Category Presence Employing Various uBlock Configurations [Per-Category Reduction Relative to Baseline]

6.5 Why Do Blockers Fail?

Another obvious, and yet disappointing, result is that none of the ad blocker configurations provide perfect protection for any of the categories. In the best cases, results for a category are reduced to a presence of few percent, but not zero. These results lead to the obvious question of why not.

Investigation of the results yields three reasons for why even the best blockers do not block all domains within a category. The first reason is that we determine the "root" domain for a server as defined in [16] where the domain providing content of a server may not match the syntactic domain due to a DNS CNAME alias. For example, the server ls.webmd.com is not matched by a filter list rule, but it is actually an alias for oasc18033.247realmedia.com where 247realmedia.com is classified as an AdTrackers domain.

The second reason the results do not go to zero for a category is that we observe the tools are "leaky" in that despite being supposedly blocked by a rule some requests are still made by the Firefox browser. These leaked requests are less than one percent for a particular domain, but contribute to the results for a category to not go to zero. We investigated this phenomenon thinking it might have to do with our measurement platform where we both were blocking ads with an extension and recording requests and replies with a different extension. However, we ran a separate test using a proxy to record results and still observed the same small number of these leaked requests from the browser. We did observe such leakage for all extensions so this result may be due to the Firefox extension mechanism.

The third reason that not all third-party domains are blocked is simply because tool filter lists do not block the same domains. These unblocked domains may not be considered appropriate to block or are unknowingly not included by a tool in its employed list.

6.6 Composite Ad Blocker

Our final piece of our analysis is to combine the effectiveness results of various ad blockers along with the usage data in Table 1 and data from [21] on Adblock Plus configuration usage to estimate the effectiveness of the "composite" ad blocker currently in use. The performance of this composite ad blocker is shown in Figure 8 along with default tool behavior from Figure 4 for comparison. The composite ad blocker results show the heavy influence of the default Adblock Plus configuration (estimated to be used by roughly two-thirds of all ad block tool users) with only a modest range reduction from 13-34% across categories relative to not employing any ad blocker.

7 Summary and Future Work

The results of this work demonstrate that there is much variation in the effectiveness of current ad blocking tools to prevent requests to different types of third-party domains. Non-configurable tools such as Blur and Disconnect provide only modest blockage of third-party domains in most categories. The tool uBlock generally demonstrates the best default configuration performance. By default, Ghostery provides no protection while Adblock Plus and Adguard provide minimal protection. They must be manually configured to obtain effective protection. The current "composite" ad blocker, taking into account the market share of each ad blocker, provides only a modest reduction in requests to third parties across different categories.

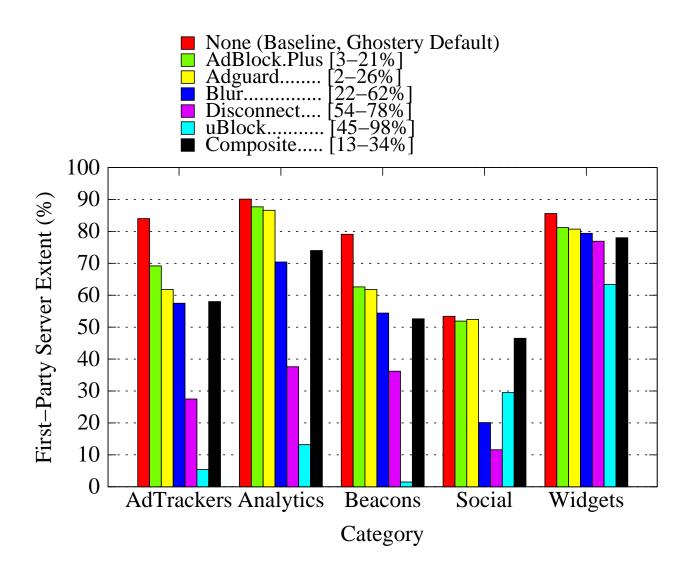


Figure 8: Composite Ad Blocker Results for Third-Party Category Presence Compared with Default Ad Blocker Configurations [Per-Category Reduction Relative to Baseline]

There are a number of directions for future work. The defined set of categories or the classification of specific domains could be further explored. Other measures of ad blocker effectiveness could also be evaluated such as the avoidance of "annoying" ads or negative impact on page appearance and functionality. Do most Adblock Plus users employ the default "avoid annoying and allow acceptable ads" configuration because that is what they prefer or simply because that is the default? Finally, there is a clear need to extend this work to mobile platforms and understand the effectiveness of ad blocking tools in that domain.

References

- [1] Protect your privacy with blur. https://www.abine.com/.
- [2] Abine DoNotTrackPlus Third-Party Domain Information. http://www.donottrackplus.com/trackers/<domain>.php.
- [3] AdBlock Plus. http://adblockplus.org.
- [4] AdBlock Plus Features. http://adblockplus.org/en/features.
- [5] Adguard. http://adguard.com.
- [6] Adguard Useful Ads. http://adguard.com/en/whitelist.html.
- [7] Alexa: Most popular web sites. http://www.alexa.com/.
- [8] Charles Arthur. Why the default settings on your device should be right first time. *The Guardian*, December 1 2013.

```
{\tt https://www.theguardian.com/technology/2013/dec/01/default-settings-change-phones-computers.}
```

- [9] Chrome Web Store. https://chrome.google.com/webstore/category/popular.
- [10] Disconnect. http://disconnect.me.
- [11] Firefox Most Popular Extensions. https://addons.mozilla.org/en-US/firefox/extensions/?sort=users.
- [12] Ghostery. http://www.ghostery.com/.
- [13] hpHosts. http://www.hosts-file.net/.
- [14] Jason Huggins, Adam Goucher, Shinya Kasatani, Dave Hunt, and Samit Badle. Selenium IDE. Firefox Add-on. https://addons.mozilla.org/en-US/firefox/addon/selenium-ide/.
- [15] Balachander Krishnamurthy and Craig E. Wills. Generating a privacy footprint on the Internet. In *Proceedings of IMC*, October 2006.

- [16] Balachander Krishnamurthy and Craig E. Wills. Privacy diffusion on the web: A longitudinal perspective. In *Proceedings of the World Wide Web Conference*, 2009.
- [17] Jonathan R. Mayer and John C. Mitchell. Third-party web tracking: Policy and technology. In *Proceedings of IEEE Symposium on Security and Privacy*, May 2012.
- [18] Hassan Metwalley, Stefano Traverso, and Marco Mellia. Using passive measurements to demystify online trackers. *Computer*, 49(3):50–55, 2016.
- [19] Hassan Metwalley, Stefano Traverso, Marco Mellia, Stanislav Miskovic, and Mario Baldi. The online tracking horde: a view from passive measurements. In *Traffic Monitoring and Analysis*, pages 111–125. Springer, 2015.
- [20] The cost of ad blocking: PageFair and Adobe 2015 ad blocking report. https://downloads.pagefair.com/wp-content/uploads/2016/05/2015_report-the_cost_of_ad_blocking.pdf.
- [21] Enric Pujol, Oliver Hohlfeld, and Anja Feldmann. Annoyed users: Ads and ad-block usage in the wild. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, pages 93–106. ACM, 2015.
- [22] Daniel Savard. LiveHTTPHeaders. Firefox Add-on. https://addons.mozilla.org/en-US/firefox/addon/live-http-headers/.
- [23] Jared Spool. Do users change their settings? *User Interface Engineering*, September 14 2011.

 https://www.uie.com/brainsparks/2011/09/14/do-users-change-their-settings/.
- [24] Trend Micro Site Safety Center. http://global.sitesafety.trendmicro.com/.
- [25] uBlock. https://www.ublock.org/.
- [26] Doruk Uzunoglu. Understanding ad blockers. Technical Report MQP-CEW-1601, Worcester Polytechnic Institute, March 2016.
- [27] Robert J Walls, Eric D Kilmer, Nathaniel Lageman, and Patrick D McDaniel. Measuring the impact and perception of acceptable advertisements. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, pages 107–120. ACM, 2015.