On the Impact of Internal Webpage Selection when Evaluating Ad Blocker Performance

Philip Gunnarsson
Linköping University, Sweden

Adam Jakobsson
Linköping University, Sweden

Niklas Carlsson
Linköping University, Sweden

***Abstract*—Not all ad blockers achieve the same blocking success and, depending on their implementation, they can either improve or hurt the web performance experienced by users. Borgolte and Feamster (2020) recently provided the first extensive evaluation of how privacy-focused browser extensions affect a user's web performance. However, while their work provides a nice comparison of the performance impact that different extensions may have, their evaluation only considered landing pages of a single set of websites. In this paper, we focus specifically on performance comparisons when considering different sets of webpages. For example, we study the impact of whether a page is a landing page or an internal page, whether a page is popular or less popular, as well as the impact of in which country/region the company registering the website is operating (used as a proxy for the primary target market). For our evaluations, we use pairs of webpages carefully selected from the recently proposed Hispar list (Aqeel et al. 2020) and compare the performance of the most popular blocking extensions (Adblock Plus, uBlock Origin, Ghostery, and Private Badger) considered by Borgolte and Feamster with a baseline case in which we do not use any extension. While we observe clear differences in the distribution statistics of the metrics considered, several observations were consistent across all dimensions, including whether we consider landing pages or internal pages. The paper highlights some of these invariants and discusses their implications. In addition, our measurements (and the differences observed by different ad-blockers) also reveal new insights into how internal vs. landing pages of different webpage categories (e.g., based on popularity or region) differ in their composition and resource usage.**

***Index Terms*—Ad blockers, Internal pages, Performance**

I. Introduction

In 2021, the online advertising market exceeded 450 billion US dollars [1]. It is therefore no surprise that users today are highly tracked and constantly are presented personalized advertisements. However, many people find ads interruptive/annoying and do not agree that personalized advertising is an acceptable price to pay for receiving free content/service [2]. Instead, they think that the advertisements slow down their web browsing and/or have privacy concerns with current advertising practices [3]. For these and other reasons [3], many users have started using ad blockers and other privacy enhancing browser extensions [4]–[6].

Such extensions typically attempt to block third-party trackers, ads, or both. However, not all ad blockers are designed the same, achieve the same blocking rates, or impact the web performance experienced by users the same. The increased use of ad blockers has prompted much research on the effectiveness of ad blockers [7] and the battle between them and ad providers wanting to evade detection [8], [9]. Much less work has focused on their performance tradeoffs.

Borgolte and Feamster [10] recently provided the first extensive evaluation of how privacy-focused browser extensions affect users' web performance. In their work, they developed a framework for evaluating the web performance when visiting different websites with or without different extensions enabled, and then used the framework to compare the performance impact that different extensions have when visiting a set of landing pages selected from the Tranco list [11].

While their work provides a very nice comparison of the performance impact that the evaluated extensions may have on the users' web performance, their evaluation only considered landing pages and did not consider the impact that website selection bias has on the evaluation (e.g., whether a website is popular vs. less popular, or in which region of the world it is hosted). As argued and shown in recent work by Aqeel et al. [12], best paper winners at IMC 2020, it is important to also consider web performance when visiting internal pages. Motivated by current ranking lists typically used in web performance studies only including landing pages, and the substantial differences that can be observed when evaluating different systems on visited internal pages compared to the landing pages, they argue that many web performance studies should be repeated also for internal pages.

In this paper, we complement the performance study by Borgolte and Feamster to address some of the concerns raised by Aqeel et al. First, we use the evaluation framework by Borgolte and Feamster to compare the performance differences observed when using landing pages vs. internal pages. Second, we compare the impact of using the most popular websites on such a ranking list compared to using the least popular websites. Finally, we compare the performance that the clients observe when interacting with websites hosted by organizations based in distinct regions the world. For the evaluations, we use different webpage sets carefully selected from the extended Hispar list (created+shared by Aqeel et al. [12]) and compare the performance of the most popular blocking extensions (Adblock Plus, uBlock Origin, Ghostery, Private Badger) used by Borgolte and Feamster with a baseline without extension. Here, we are interested both in the impact that the choice of evaluation pages has on the relative performance observed by the blockers and if there are differences in the page composition and resource usage of the different webpage subsets that impact the adblockers differently.

The definitive version was published in *Proc. IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*, Nice, France, Oct. 2022. https://doi.org/tbd.

While we observe clear differences in the distribution statistics of the eight performance metrics of interest (five browser performance metrics and three system performance metrics), several observations are consistent across the website classes. These seemingly invariant observations include the worst-case performer with regards to each metric. For example, Adblock Plus consistently achieved by far the worst page-load times and smallest reduction in the number of transferred bytes, cookies, and resource objects of the four extensions. In most cases, Adblock Plus (the most popular of the extensions) substantially inflated the page-load times compared to the baseline case. In contrast, uBlock Origin typically performed the best with regards to these metrics for all cases except for transfer sizes (where Ghostery in a few cases performs slightly better). Furthermore, while ad blockers typically reduced the amount of downloaded content (including cookies) somewhat, all extensions increased the on content load times.

Overall, our results suggest that the relative ad-blocker comparisons when using only landing pages (e.g., as collected by Borgolte and Feamster) for the most part holds also when using internal pages, as well as when breaking up the analysis based on popularity or hosting country. The main differences are instead typically visible in the website composition and their resource requirements associated with each page visit.

We next present our methodology (§2), performance comparisons (§3), related work (§4), and conclusions (§5).

II. Methodology

For our experiments, we used the framework created by Borgolte and Feamster [10]. The framework allows us to visit a set of URLs using different extensions. Given an ordered list of webpages, the framework evaluates one webpage at a time. For each webpage, the framework (1) visits the webpage to make sure that all tests have the same prerequisites regarding access, DNS cache, etc., and then it (2) runs the same experiment with each of the extensions as well as a baseline experiment without any extension. In the second step, the order that each experiment is done is selected at random each time that we process a URL. For fair comparison, we always consider one pair of webpages at a time, where a pair consists of an internal page and the corresponding landing page. While this approach allows for fair comparisons of extensions, it is time consuming since each extension needs to be reinstalled for every test.

A. Webpage selection

During the website selection process, we selected contrasting sets of webpages that allow direct head-to-head comparisons. These sets were selected to allow comparisons along two primary dimensions. First, and most importantly, we wanted to compare the impact of going to a landing page (as done by Borgholte and Feamster) vs. going to internal pages of the same domain (as argued important by Aqeel et al. [12]). Second, we wanted to compare the impact of webpage popularity and the region/country that different websites are listed.

To achieve the above objectives, we used the extended Hispar list shared by Aqeel et al. [12]. Using the list from March 29, 2021, we first extracted the URLs for both the 10,000 top-ranked internal webpages on the list and the 10,000 bottom-ranked internal webpages on the list. (The Hispar list contains 100,000 webpages.) Second, we made sure that we added the corresponding landing page for each internal URL. (Since Hispar typically contains several internal webpages for each domain, our complete list of URLs contains many landing pages several times.) This step ensured that we always collect matching data points for the landing page associated with every tested internal page. In our experiments, described later, the tests of each pair of landing + internal webpage were performed back-to-back (close in time). Combined with random ordering of the pairs tested as well as which of the two members of each pair is performed first, this helps ensure fair head-to-head comparison between the relative performance seen when visiting landing vs. internal webpages.

B. Metrics

For each test, we extracted and saved the same metrics as those used by Borgolte and Feamster [10]. However, for our evaluation, we excluded the CPU-clock parameter. This parameter is faulty and always reports the same value as the *task clock* (which we report) when using the Linux `perf` data tool [13] (also used in the original paper [10]).

Browser metrics: HTTP Archive format (HAR) files [14] were used to capture browser performance. Here, we extracted and investigated five metrics capturing how the browser downloads and renders a webpage. First, we extracted two timing-based metrics measuring the load time of a webpage: *on content load* and *on load*. The *on content load* metric measures the time until the initial HTML has loaded but not all resources have been included, whereas the *on load* metric measures the time until the onLoad event that is triggered when the browser finish loading and rendering the page. The other three browser metrics are the number of requested *resources*, the overall page-load size (referred to as the *body size* in the figures, to match the figures in [10]), and the number of *cookies*.

System metrics: We used three system metrics. (1) The *task clock* measures the total time spent computing across all processor cores, excluding the time that the processor is sleeping. (2) The number of *CPU migrations* measures the how many memory page faults occur due to a process attempting to access memory currently not mapped or loaded into the virtual address space. (3) The number of *context switches* measures how many times that the operating system switch between processes (which require it to store away state for a process). As noted by Borgolte and Feamster [10], these metrics are all correlated to the energy usage that can be associated with an individual process. Like them, we measure the system metrics using Linux's `perf_events` [13].

C. Extension selection

We selected to use the four most popular blockers studied by Borgolte and Feamster [10]: Adblock Plus [15] (20M users), uBlock Origin [4] (15M users), Ghostery [16] (4M users), and Privacy Badger extension [5] (2M users). While there exists

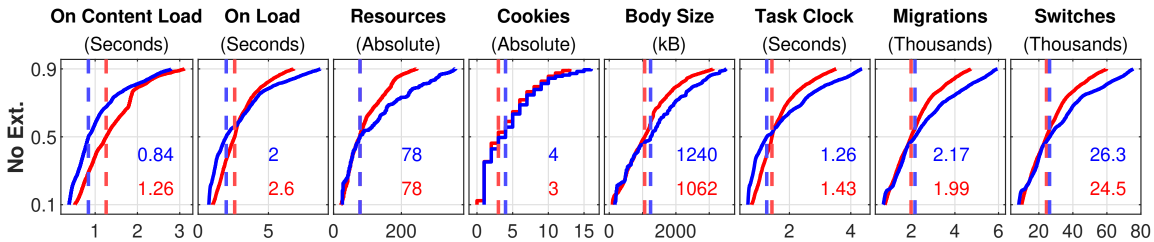

Fig. 1. CDF-based comparison between landing pages (blue) and internal pages (red) when not using any extension. Results are shown for eight metrics. Vertical lines (and values) show the medians. The top- and bottom 10% of each curve are removed for easier visual comparison.

other popular adblockers, this selection allows for comparisons with their results and provides another reference point for comparison between internal and landing pages.

Clearly, Adblock Plus and uBlock Origin are by far the most popular extensions. While both these extensions use blocklists, they take somewhat different approaches to adblocking. For example, AdBlock Plus does not block all ads by default, but allows some ads through via the acceptable ads program [17]. In contrast, uBlock Origin attempts to block all ads. Ghostery is similar to these two ad blockers in that it too uses a blocklist. However, in contrast to them (who rely on public lists such as EasyPrivacy [18] and/or EasyList [19], curated by the user community), Ghostery uses its own private list.

Blocklists take time to update. This results in a delay before new ads/trackers are possible to block. To address this shortcoming, Privacy Badger [5] implements heuristics that try to identity third-party content that match specific properties and then modify or block the corresponding requests [20]. To reduce the false positive rate, Privacy Badger often modify rather than block requests (e.g., by removing cookies to protect the user's privacy). In contrast to the others, Privacy Badger primary aim to address tracking, not blocking ads.

D. Experimental setup and limitations

All experiments were done from a single (anonymized) location (in Europe) with good network connectivity. We used a laptop with Intel Core m3-7Y30 processor, a Mesa Intel HD Graphics 615 (KBL GT2) graphic card, 8 GB RAM, that ran Ubuntu 20.04 LTS. The tests presented here ran from June 23 to July 13, 2021. The Tranco list and Hispar lists used were obtained on March 29, 2021. For comparison purposes, we used the same browser and extension versions as used in prior work: Firefox (68.0.2), Adblock Plus (3.6.3), Ghostery (8.4.2), Privacy Badger (2019.7.1.1), uBlock (1.22.2).

Despite good connectivity, some tests failed. For fair comparison, the analysis therefore only includes sets of measurements that include ten successfully completed tests for a given landing-internal pairing. In particular, we required the tests and corresponding HAR data to have been successfully obtained for all five configurations (i.e., baseline without extension + four extensions) for both the internal- and the landing page.

Limitations: First, we ran our experiments from a single location. Naturally, the network connectivity, location, and time of day each website is evaluated may impact the absolute performance experienced by a client. To limit impact of such biases and ensure fair comparisons, our analysis focuses on the relative performance differences between the extensions and if an internal- or landing page is accessed. Furthermore, we made sure that the tests for each set of 10 tests always were done close in time (within a few minutes due to the time of each test), we only included measurements for which all 10 tests were successful in the analysis, and the order each extension is tested is selected at random for each sample domain. We also evaluated significantly more webpages than Borgolte and Feamster [10], focus on the relative performance of the extensions, and use hypothesis tests (described later) to support the significance of our findings.

Second, for easier comparison, we restricted our analysis to the same metrics as used by Borgolte and Feamster [10] and did not study additional parameters such as the impact of the amount of compression that each website employs, for example. While the amount of compression (and other measures) may impact the performance of individual webpages, we note that all performance comparisons (between extensions) are done on a per-website basis. The compression seen by each extension for that webpage is therefore likely to be similar.

Finally, we do not consider every blocker. Instead, we focus on the most popular blockers studied in prior work, and provide dates and versions for the lists and extensions used. For reproducibility, we share our datasets (https://www.ida.liu.se/~nikca89/papers/mascots22.html).

III. Performance comparisons

A. Baseline tests without any extension

Let us first compare the performance when accessing the internal pages compared to their corresponding landing pages.

As baseline comparison, we ran experiments without any extension. Figure 1 shows the empirical Cumulative Distribution Functions (CDFs) of the eight metrics of interest. The red curves show the CDF of the internal pages, with the vertical red line (and red value) showing the median value. The corresponding curves and median values for landing pages are shown in blue. To ease visual comparison, we removed the top- and bottom 10% from each curve.

Landing pages more optimized: We found that most landing pages were better optimized with regards to the load times than the internal pages. For example, despite having larger *body size* (total compressed data sent) and requiring more *resources* to be downloaded than internal pages, the landing pages typically had significantly faster *on content load* times. While also the median *on load* times were somewhat faster for landing pages, the variations were larger for them.

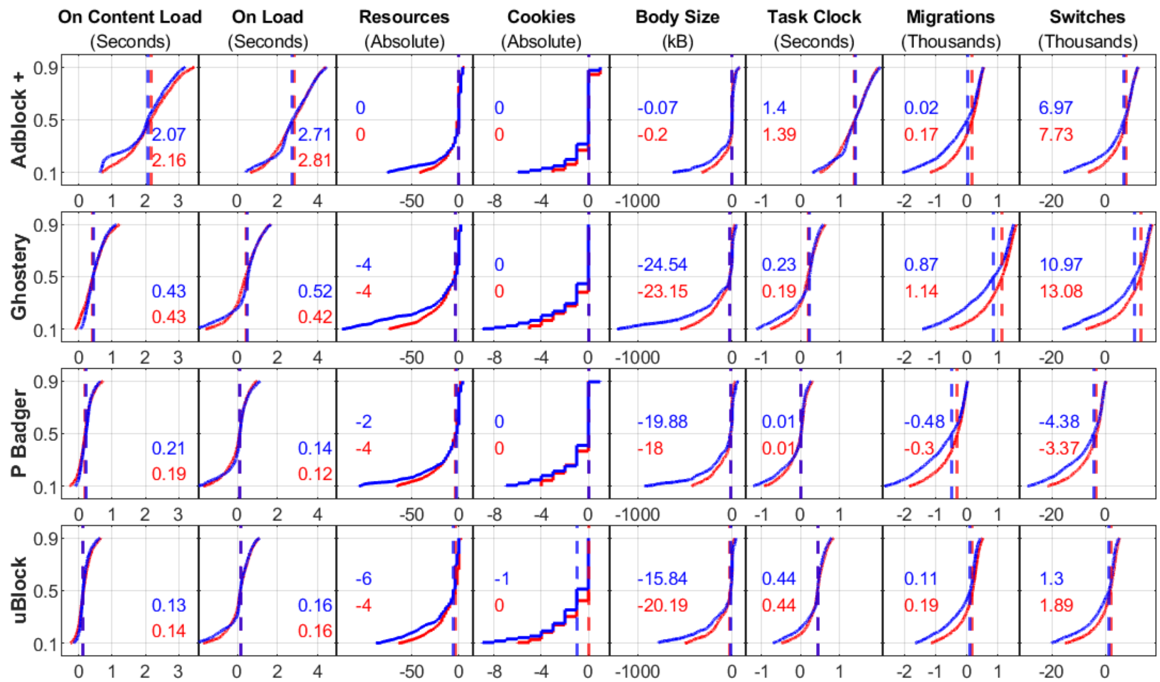

Fig. 2. CDFs of the relative increase (positive values) of decreases (negative values) when using each extension compared to the baseline(without extension). For each extension (rows), we show the relative increase/decrease of each metric (column) compared to the baseline.

Small differences in cookie usage: Compared to the differences in *resources* and *body size*, at the aggregate level, we only observed slightly more cookies on the landing pages.

Bigger variations among landing pages: With exception of the *on content load* metric, we observed bigger variations for all metrics when looking at the landing pages. When looking at the individual metrics, we found that the landing pages typically required more *CPU migrations* and *context switches*. The differences in these hardware metrics suggest that there may be more work (in addition to more bytes to process) for landing pages than internal pages.

B. Extension comparisons

We next compare the performance with different extensions. Figure 2 summarize these results. Here, we show CDFs of the relative increase (positive values) and decreases (negative values) when using one of the four extensions compared to the baseline test (without any extension). Here, we give each extension a separate row and each metric a separate column.

Increased load times with extensions, especially AdBlock Plus: None of the extensions improved the median *on load* times or *on content load* times, and Adblock Plus consistently performs the worst with regards to both metrics. For example, while Ghostery (0.43/0.43 seconds), Privacy Badger (0.21/0.19) and uBlock Origin (0.13/0.14) all have median values of the *on load*/*on content load* within half-a-second of when not using an extension, the median values when using Adblock Plus increase by 2.07 and 2.16 second for landing pages and internal pages, respectively.

The much worse performance with AdBlock Plus, the increased *on content load* times of all extensions, as well as the relative performance ranking of the extensions are consistent with the *on content load* results by Borgolte and Feamster [10]. However, in contrast to them we also observed increased *on load* times with Ghostery, Privacy Badger and uBlock. We have found that these differences in large appears to come from the domain selection. For example, we have found that our results look more similar to theirs if only looking at the lower-ranked webpages considered here, as the blockers tend to block more content on these webpages (c.f., compare reductions in *resources* and *body size* for bottom vs. top for these three extensions as shown in Table I, discussed in the next subsection). We also note that the narrow distributions observed for these cases reduce the significance of the small differences we observe in these median values.

Interestingly, after subtracting the corresponding load times when not using an extension (shown in Figure 1), we see very limited differences in the relative performance differences between the extensions. This suggests that many of these extensions are either equally good or equally bad at handling landing pages and internal pages. For example, comparing the relative variations compared to the no-extension baseline we observe much smaller differences between the distributions of the internal and landing pages of a given extension than what we observe when comparing AdBlock Plus with Ghostery or with one of Privacy Badger or uBlock Origin (which are the two with the most similar performance).

Adblock Plus achieves the smallest reduction in *resources* downloaded and *body size*: Given most ad-block users' desire to see less ads, reductions in the number of *resources* downloaded and the total *body size* are expected. In our experiments, this is indeed the case for both landing and internal pages (and was the case for Borgolte and Feamster [10] as well) regardless of which adblocker is used.

Furthermore, given the larger variations in sizes of landing pages (e.g., Figure 1), we would expect to see bigger reductions in the number of *resources* downloaded and *body size* for landing pages (blue) than internal pages (red). For the tail of the distribution, the reductions are indeed substantial. However, for the median values the differences are more modest and in the case of Adblock Plus the reductions are very small, with almost no reductions in the median values (0 and 0, respectively, with regards to resources, and -0.07 kB and -0.2 kB, respectively, for median body size). This small reduction can partially be explained by the acceptable ads program used by Adblock Plus [17].

Performance metrics: To better understand the inflation in the two load-time metrics when using Adblock Plus, we look closer at the *task clock*. Here, more than 90% of the pages (regardless of being landing or internal pages) see increased *task clock* times (with medians of 1.40 and 1.39 seconds) when using AdBlock Plus. In contrast, the three other extensions see more balance of pages with increased/decreased *task clock* times. This shows that Adblock Plus comes with significant processing that slowed down load times substantially.

For *CPU migrations* and *context switches*, only Privacy Badger were able to improve reductions for more than half of the pages. The three other extensions saw slight increases for more than half of the pages (resulting in positive medians).

While the median values were relatively similar when comparing internal vs. landing pages, the biggest observed reductions for the three hardware metrics were observed for the landing pages. This appears to primarily be due to differences in how the websites themselves differ in their design, as the landing pages also saw the largest number of *migrations*, *switches*, and the largest *task clock* times (see Figure 1).

Cookies: While all extensions reduce the number of cookies observed on average, in most cases less than half of the pages saw any reduction (zero values), and Adblock Plus achieved the smallest reductions. The result appears consistent regardless of the use of external or internal pages.

Relative performance of extensions does not significantly depend on if we consider internal or landing pages: The relative performance of the four extensions is typically consistent regardless of whether we looked at internal or landing pages. For example, we note that whenever the median value for the landing pages is positive so is the corresponding median value also for the internal pages. Furthermore, looking at the rankings of each extension with regards to their median values, the relative rankings remain consistent for six out of eight metrics, and in the two cases (*body size* and *CPU migration*) that the rankings differed, the relative values were relatively similar. For example, for the cases of changes in median body size, Ghostery provides the biggest reduction in both cases and Adblock Plus the smallest. The differences here were in the order of the 2nd and 3rd ranks, where Privacy Badger and uBlock Origin had similar values (and changed rank when considering the internal vs. landing pages).

C. Rank-based analysis

We next consider the relative popularity of the websites accessed. Here, we consider three aspects. First, whether the results comparing different extensions against the baseline are impacted by using the top-ranked vs. bottom-ranked webpages. Second, whether the results are impacted by the subset selection. Finally, whether the results themselves are impacted by whether we use external or internal pages.

To answer the above questions, we evaluate each extension (as well as the baseline case) on both the set of top-ranked webpages and the bottom-ranked webpages using both internal and external webpages. Table I summarizes these results. Here, all values are relative to the baseline experiments without any extension. To improve readability, we label all cases where a metric increases as red and all cases where a metric decreases as green. Note that all metrics are such that green is positive and red is negative *from the perspective of an extension*. In other words, we use green to indicate improved performance with the extension.

The results in the table confirm the generality of several observations. First, when applying majority voting, we have found that whether an extension outperforms the baseline or not is relatively independent of which of the four sample sets is used ($\{$top, bottom$\} \times \{$internal, landing$\}$). Second, the extensions tend to reduce the number of *resources* downloaded and the total *body size* but increase the *on content load* times. Third, given a metric, the relative rankings of the different extensions are typically relatively independent of which of the four sample sets is used. We next provide some statistical analysis to support these observations.

Relative to baseline: For the most part, the main results are consistent irrespective of whether the top-ranked or bottom-ranked websites are evaluated. For example, consider whether the extensions see improved (green) or worse (red) performance with regards to each of the possible cases. In total, we observe opposite colors in only 9 out of 64 cases (8 metrics $\times$ 4 extensions $\times$ $\{$landing, internal$\}$) and different colors in 17 out of 64 cases. Ignoring the probability of seeing the neutral case (white) and assuming (as a null-hypothesis) an unbiased scenario for which we would have equal probability q=0.5 to see opposite (or different) colors, these two outcomes have small p-value ($<10^{-8}$ and p=$1.13 \cdot 10^{-4}$), with z-scores of z=5.625 and -3.625, respectively, if approximating the binomial distribution with the normal distribution [21].

The results are also significant for three out of four extensions, when considering one extension at a time. For example, for the case when we ignore cases where one is neutral (which is a borderline cases) we have: Adblock Plus (1 out of 16; $p = 2.59 \cdot 10^{-4}$), Ghostery (1 out of 16; $p = 2.59 \cdot 10^{-4}$), Privacy Badger (2 out of 16; $p = 2.09 \cdot 10^{-3}$), and uBlock Origin (6 out of 16; $p = 0.227$). Clearly, the conclusion about uBlock Origin are the weakest (not significant even at 90% confidence level). The non-significant differences for uBlock can be attributed to its very small impact on the *on load* time, *CPU migration*, and *context switches* (e.g., Figure 2).

TABLE I
ANALYSIS COMPARISON OF THE TOP-RANKED VS. THE BOTTOM-RANKED WEBSITES USING BOTH INTERNAL AND LANDING PAGES.

Medians		On Content Load		On Load		Resources		Cookies		Body Size		Task Clock		CPU Migrations		Context Switches	
		Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top
AdBlock+	Internal	1.88	2.16	2.37	2.81	-2.00	0.00	0.00	0.00	-2.84	-0.20	1.26	1.39	0.04	0.17	6.70	7.73
	Landing	1.72	2.07	2.48	2.71	-2.00	0.00	0.00	0.00	-1.91	-0.07	1.31	1.40	-0.05	0.02	6.18	6.97
Ghostery	Internal	0.43	0.43	0.19	0.42	-20.00	-4.00	-1.00	0.00	-125.06	-23.15	0.10	0.19	0.77	1.14	9.73	13.08
	Landing	0.45	0.43	0.28	0.52	-22.00	-4.00	-2.00	0.00	-164.58	-24.54	0.10	0.23	0.65	0.87	8.58	10.97
P-Badger	Internal	0.20	0.19	0.00	0.12	-12.00	-4.00	-1.00	0.00	-55.53	-18.00	-0.06	0.01	-0.51	-0.30	-5.08	-3.37
	Landing	0.23	0.21	0.01	0.14	-14.00	-2.00	-1.00	0.00	-75.16	-19.88	-0.07	0.01	-0.57	-0.48	-5.80	-4.38
uBlock O.	Internal	0.15	0.14	-0.04	0.16	-18.00	-4.00	-1.00	0.00	-115.08	-20.19	0.35	0.44	-0.09	0.19	-0.88	1.89
	Landing	0.17	0.13	-0.01	0.16	-24.00	-6.00	-2.00	-1.00	-194.97	-15.84	0.33	0.44	-0.17	0.11	-1.87	1.30

Extension comparisons: Consider next the whether the choice to use top-ranked or bottom-ranked websites impacts the relative rankings of the different extensions. Here, we have 16 possible cases (8 metrics $\times$ {landing, internal}). For example, consider the *on content load* time ranking for internal pages. This is (1) uBlock Origin, (2) Privacy Badger, (3) Ghostery, and (4) Adblock Plus, irrespective if we use the top-ranked websites or bottom-ranked websites. The probability of this happening if the pages relative performance is independent is 0.042 (1/4!). Now, considering that we get the same ranking for 13 out of the 16 possible pairwise rankings (z = 14.93), we can clearly see that also here the choice of using the top-ranked or bottom-ranked websites would lead to very similar conclusions. (For simplicity, we break tied rankings in favor of the ranks being equal.) The three exceptions are the rankings for the two *on load* times and the *body size* metric for landing pages. However, also here, Adblock Plus consistently is the worst performer in 16 out of 16 cases. This result is significant. For example, using the null-hypothesis that the probability of each adblocker is the worst performer is uniform (i.e., 0.25), results in a p-value of $2.32 \cdot 10^{-10}$ (and z = 6.64).

The above results also reinforce that the median values presented in previous subsections provides good insights into the relative performance of the extensions. For example, Adblock Plus sees the biggest inflation in load times and appears to provide the least blocking. However, these results also shows that some care must be made when considering the extensions with the best *on load* times. For this metric, uBlock Origin is the best when considering the bottom-ranked pages but only the second best when using the top-ranked websites (beaten by Privacy Badger) to compare the extensions. Now, it should be noted that the load time differences of these extensions are comparable (including when considering the full distributions; e.g., see Figure 2) and both beat the other extensions significantly. Therefore, relative ranking changes between these two metrics are not as significant as we would have seen rank changes for other cases.

Internal vs landing pages: As noted in prior subsections, most observations are consistent when comparing internal and landing pages. To strengthen this observation, let us next use both the corresponding binomial test and ranking-based comparisons we just did for top-ranked vs. bottom-ranked websites. Here, the results are even more consistent, and we see even smaller differences. First, only 3 out of 64 cases (8 metrics $\times$ 4 extensions $\times$ {top-ranked, bottom-ranked}) see differences even when counting cases where one of the two can be neutral, and in only 1 out of 64 cases an extension see opposite majority-voting results (i.e., one green and one red) compared to the baseline for the two cases. ($z \approx$-7.125 and $z \approx$7.625) The exception here is *CPU migrations* using Adblock Plus on the bottom-ranked pages.

Similarly, when considering the rank-based comparisons, in 13 out of 16 possible cases (8 metrics $\times$ {top-ranked, bottom-ranked} subset of pages), the rankings of the extensions are the same irrespective of whether we use the landing pages or the internal pages for the performance comparison. (The exceptions are the number of *resources* downloaded when using he bottom-ranked pages and the two *body size* cases.) The result is even stronger when considering the worst-case performer. Here, the worst-case performer again is the same in 16 out of 16 cases, irrespective of which set of websites is used. These results again strengthen the generality of the median-based analysis presented in previous subsections and further confirms that these comparisons using landing pages may not be that different than those one obtains by using internal pages (which importance to study was emphasized by Aqeel et al. [12]). While this strengthens the generality of the results presented here, we have also observed some differences between internal and landing pages when trying to quantify the improvements in tail performances (Figure 2). However, much of these differences can be attributed to the bigger variations (and heavier tails) of the landing pages themselves (e.g., Figure 1) and the extensions being able to better block some of the "extra" resources associated with such webpages.

D. Country-based comparison

Finally, encouraged by the robustness observed for the median-based comparisons of prior sections, we next compare the actual performance when visiting different subsets of websites. Here, we compare the performance observed when accessing the internal pages and corresponding landing pages of all websites on our list that were listed in one out of seven countries/regions, as determined by the hosting country as classified by Netcraft [22]. Table II summarizes the results. In contrast to prior results in the paper, we now show the *absolute* values (rather than the *relative* values) and therefore include a column also for the no-extension baseline. To save space, with exception of USA and EU (who had by far the most identified websites: 2,547 and 554 landing pages, respectively, plus a significant number of internal pages per domain), we only include results for two timing metrics and the number of cookies. For USA and EU, we included all metrics (others in italics).

TABLE II
SUMMARY COMPARISONS FOR DIFFERENT COUNTRIES. EACH ROW IS COLOR CODED FROM BEST (GREEN) TO WORST (RED).

	Internal					**Landing**					
	–	**AB+**	**PB**	**uBO**	**Gh**	**–**	**AB+**	**PB**	**uBO**	**Gh**	
On Content Load	1.28	3.51	1.44	1.43	1.76	0.73	2.96	0.91	0.89	1.23	
On Load	2.47	5.25	2.53	2.70	2.82	1.60	4.85	1.81	1.81	2.23	
Cookies	2	2	2	1	1	2	2	2	1	1	
Resources	68	64	60	60	58	68	66	66	62	56	
Body Size	1,014	960	914	895	834	913	886	885	758	871	
Task Clock	1.4	2.8	1.4	1.8	1.5	1.0	2.7	1.1	1.5	1.3	
CPU Migrations	1.9	1.9	1.5	1.9	2.8	1.8	1.9	1.4	1.9	2.7	
Context Switches	23.3	29.9	18.6	23.2	34.1	22.4	30.3	17.9	23.4	33.0	**USA**
On Content Load	1.01	3.65	1.21	1.20	1.66	3.27	4.54	3.22	3.19	2.93	
On Load	2.42	5.75	2.77	2.67	3.22	7.45	6.16	4.68	6.56	4.69	
Cookies	6	6	5	5	5	8	4	4	6	4	
Resources	86	88	82	74	74	255	36	36	202	36	
Body Size	1,372	1,350	1,326	1,298	1,257	2,470	528	379	1,715	357	
Task Clock	1.4	3.0	1.5	1.8	1.7	3.7	3.1	2.2	3.4	2.3	
CPU Migrations	1.9	2.2	1.7	2.1	3.2	5.7	1.8	1.2	3.7	2.8	
Context Switches	25.1	33.2	22.3	27.1	38.6	71.0	26.2	15.8	49.7	32.1	**EU**
On Content Load	3.41	5.26	3.46	3.57	3.56	1.00	3.47	1.19	1.11	1.50	
On Load	3.91	6.66	4.15	4.23	4.34	2.055	5.22	2.21	2.37	3.00	
Cookies	4	4	4	3	3	7	5	5	5	4	**China**
On Content Load	0.67	2.275	0.92	0.71	0.84	4.435	6.18	4.58	4.765	4.64	
On Load	0.78	2.85	0.99	0.87	1.065	9.49	9.435	7.85	6.92	7.165	
Cookies	0	0	0	0	0	20	10.5	9	1	1	**Japan**
On Content Load	1.53	3.53	1.59	1.68	1.84	0.33	0.71	0.53	0.47	0.55	
On Load	4.30	7.04	3.51	3.61	3.87	0.44	0.91	0.62	0.58	0.75	
Cookies	6	6	5	4	5	1	1	1	1	1	**Canada**
On Content Load	0.62	2.77	0.67	0.68	1.01	0.82	3.26	0.92	0.91	1.45	
On Load	1.65	3.97	1.12	2.04	1.66	3.17	6.72	1.59	3.56	2.64	
Cookies	10	12	7	4	6	8	7	2	3	3	**Russia**
On Content Load	1.48	3.28	1.62	1.59	1.88	1.40	3.33	1.55	1.56	1.83	
On Load	4.15	7.04	4.3	5.94	7.23	3.57	5.87	2.63	2.40	2.82	
Cookies	7	7	4	4	4	5	4	3	3	3	**Iran**
	Internal					**Landing**					

Despite the smaller webpage subsets (increasing uncertainty in medians), we make several interesting observations. First, it is again reinforced that Adblock Plus (AB+) is by far the slowest ad blocker and consumes the most processing time. For example, for the timing-based metrics (*on content load*, *on load*, and *task clock*) it was the worst extension for 13 or 14 out of the 14 possible cases (7 regions $\times$ {internal, landing}). The exception is the *on load* times for the landing pages in EU, where uBlock Origin (uBO) performed slightly worse.

Second, our previous observation that also the other ad blockers frequently increase the *on content load* (compared to the no-extension case) is consistent for most of the regions (21 out of 21 cases of internal page sets and 19 of 21 cases of landing page sets). For the *on load* times, we see bigger variations but note that we also here observe increases with all three other extensions for the US (i.e., the regions with by far the most domains), China, and Canada. Finally, the extensions usually reduced the median number of cookies, regardless of region and whether internal- or landing pages were considered.

It is also interesting to see the big differences between the different regions when comparing the raw internal vs. landing page metrics. For example, the websites listed in EU seem to have bigger landing pages (compared to internal pages) with many ads that the ad blockers were able to block. In contrast, the US websites typically have larger internal websites (median case) than the landing pages. More generally, the websites listed in the US, Canada, and China appears to optimize landing pages more with regards to load times than they optimize the internal pages. This is in big contrast to pages hosted in EU and Japan, which often appears to have much heavier landing pages.

IV. RELATED WORK

Several researchers have highlighted privacy and security risks associated with using including how they can contribute to the uniqueness of a user [6] and how they can modify/observe browsing activity [10] or retrieve privacy-sensitive information [23]. Today, browsers regulate the data that extensions are allowed to collect [24].

Alrizah et al. [25] have found several concerning cases where legitimate content was incorrectly blocked or the list editors of EasyList failed to block content by advertisers. Malloy et al. [26] found limited geographic differences in both ad-block usage and fraction of blocked ads.

Several works have tried to model the arms race between ad blockers and ad providers wanting to evade detection [8], [9]. Others have studied third-party tracking in the wild [27], the personalized advertisement experienced by different persona [28], evaluated the tracking ability of a tracking service [29], or tried to improve user's awareness or control of decisions that may impact privacy leakage [30].

The most related works are performance studies of the ad blockers [7], [10], [30], [31]. Garimella et al. [7] investigated how the users' privacy is impacted by the use of ad blockers and the mechanism to counter them. Comparing the performance of several ad blockers, they found that with exception of Ghostery, the ad blockers on average reduced the total data transferred by 25-33%, blocked between 60-80% of different privacy related parameters (e.g., "track", "user-id" and "user-cookie"). Merzdovnik et al. [31] study the effectiveness of popular tracker-blocking tools but do not compare the performance for different websites. More recently, Borgolte and Feamster [10] studied the user-perceived performance when using ad blockers and other privacy-focused extensions. In this paper, we leverage this framework to investigate the impact of using different subsets of webpages. This includes comparing the performance obtained when visiting landing pages vs. internal pages, websites with different popularity rank, as well websites hosted by companies listed in different countries.

For our evaluation, we leverage the Hispar list created by Aqeel et al. [12]. In their work, they demonstrate the importance of evaluating systems on both landing- and internal pages. While there exist several popular ranking lists (e.g., see [32]), Hispar was the first list to provide such a mix.

V. Conclusions

In this paper, we extended the evaluation by Borgolte and Feamster [10] to compare the relative performance experienced with different ad blockers when accessing different sets of webpages. Of particular interest here (and motivated by Aqeel et al. [12]) is the performance comparisons when considering only landing pages (done by Borgolte and Feamster) vs. internal pages, but also to what degree it matters whether a page is more or less popular and in which country/region it is listed. The study allows us to identify properties that persist across webpage classes as well as biases within certain subsets of webpages. As part of our methodology, we carefully selected webpages from the recently proposed Hispar list [12], and for our evaluation we first collected data (June 23 to July 13) and then compared the performance of the most popular blocking extensions (Adblock Plus, uBlock Origin, Ghostery, and Private Badger) with a baseline case in which we do not use any extension.

While we observed clear differences in the distribution statistics of the metrics considered, depending on whether we consider landing pages or internal pages, several observations were consistent across all dimensions. For example, with very few exceptions, Adblock Plus performed the worst and uBlock the best. We also observed that the ad blockers typically reduced the downloaded content (including cookies) somewhat but that all extensions increased the on content load times. Our results (supported using hypothesis testing) suggest that the conclusions of relative ad-blocker comparisons using only landing pages (e.g., as collected by Borgolte and Feamster) appear consistent when using internal pages as well as when considering website popularity or the registration country of each domain owner. The main differences are instead in the absolute numbers, where we observe big differences in the amount of third-party content used, resource requirements (e.g., task clock and CPU migrations, context switches), number of cookies, as well as the actual ads being shown on each page during a website visit.

References

[1] Statista, "Digital advertising spending worldwide from 2019 to 2024," https://www.statista.com/statistics/237974/online-advertising-spending-worldwide/, 2022.

[2] E. Pujol, O. Hohlfeld, and A. Feldmann, "Annoyed users: Ads and ad-block usage in the wild," in *Proc. IMC*, 2015, pp. 93–106.

[3] iabuk.com, "15% of britons online are blocking ads," https://www.iabuk.com/press-release/15-britons-online-are-blocking-ads, 2015.

[4] R. Hill and contributors, "ublock origin." [Online]. Available: https://github.com/gorhill/uBlock/

[5] Electronic Frontier Foundation, "Privacy badger." [Online]. Available: https://privacybadger.org/

[6] G. Gulyás, D. Somé, N. Bielova, and C. Castelluccia, "To extend or not to extend: On the uniqueness of browser extensions and web logins," in *Proc. WPES*, 2018.

[7] K. Garimella, O. Kostakis, and M. Mathioudakis, "Ad-blocking: A study on performance, privacy and counter-measures," in *ACM WebSci*, 2017.

[8] G. Storey, D. Reisman, J. Mayer, and A. Narayanan, "The future of ad blocking: An analytical framework and new techniques," in *arXiv:1705.08568*, 2017.

[9] S. Zhu, X. Hu, Z. Qian, Z. Shafiq, and H. Yin, "Measuring and disrupting anti-adblockers using differential execution analysis," in *NDSS*, 2018.

[10] K. Borgolte and N. Feamster, "Understanding the performance costs and benefits of privacy-focused browser extensions," in *Proc. WWW*, 2020.

[11] V. L. Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proc. NDSS*, 2019.

[12] W. Aqeel, B. Chandrasekaran, A. Feldmann, and B. M. Maggs, "On landing and internal web pages: The strange case of jekyll and hyde in web performance measurement," in *Proc. ACM IMC*, 2020.

[13] "perf: Linux profiling with performance counters." [Online]. Available: https://perf.wiki.kernel.org/index.php/Main_Page

[14] W3C Web Performance Working Group, "HTTP Archive (HAR) format," https://w3c.github.io/web-performance/specs/HAR/Overview.html.

[15] A. Plus, "Surf the web with no annoying ads." [Online]. Available: https://adblockplus.org/

[16] "Ghostery." [Online]. Available: https://www.ghostery.com/

[17] AdBlock Plus, "Acceptable ads," https://acceptableads.com/, 2021.

[18] "The Easy Subscriptions for Adblock, Adblock Plus and uBlock Origin." [Online]. Available: https://forums.lanik.us/

[19] EasyList, "Overview." [Online]. Available: https://easylist.to

[20] Electronic Frontier Foundation (EFF), "How does privacy badger work?" https://privacybadger.org/#How-does-Privacy-Badger-work.

[21] M. Schader and F. Schmid, "Two rules of thumb for the approximation of the binomial distribution by the normal distribution," *The American Statistician*, vol. 43, no. 1, pp. 23–24, 1989.

[22] Netcraft, "Netcraft." [Online]. Available: https://www.netcraft.com/

[23] Q. Chen and A. Kapravelos, "Mystique: Uncovering information leakage from browser extensions," in *Proc. ACM CCS*, 2018.

[24] A. Blondin, "Making chrome extensions more private and secure," *Chrome - safety and security*, 9 Dec, 2020.

[25] M. Alrizah, S. Zhu, X. Xing, and G. Wang, "Errors, misunderstandings, and attacks: Analyzing the crowdsourcing process of ad-blocking systems," in *Proc. IMC*, 2019.

[26] M. Malloy, M. McNamara, A. Cahn, and P. Barford, "Ad blockers: Global prevalence and impact," in *Proc. IMC*, 2016.

[27] S. Englehardt and A. Narayanan, "Online tracking: A 1-million-site measurement and analysis," in *Proc. ACM CCS*, 2016, p. 1388–1401.

[28] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan, "Adscape: Harvesting and analyzing online display ads," in *WWW*, 2014.

[29] M. Degeling and J. Nierhoff, "Tracking and tricking a profiler: Automated measuring and influencing of bluekai's interest profiling," in *Proc. WPES*, 2018.

[30] J. P. Achara, J. Parra-Arnau, and C. Castelluccia, "Mytrackingchoices: Pacifying the ad-block war by enforcing user privacy preferences," in *Proc. WEIS*, 2016.

[31] G. Merzdovnik, M. Huber, D. Buhov, N. Nikiforakis, S. Neuner, M. Schmiedecker, and E. Weippl, "Block me if you can: A large-scale study of tracker-blocking tools," in *Proc. IEEE EuroS&P*, 2017.

[32] Q. Scheitle, O. Hohlfeld, J. Gamba, J. Jelten, T. Zimmermann, S. D. Strowes, and N. Vallina-Rodriguez, "A long way to the top: Significance, structure, and stability of internet top lists," in *Proc. IMC*, 2018.