Quantifying Web Adblocker Privacy

Arthur Gervais*, Alexandros Filios*, Vincent Lenders†, Srdjan Čapkun*
*ETH Zurich, Switzerland
{1, 2, 4}@inf.ethz.ch
†Armasuisse, Switzerland
{3}@armasuisse.ch

***Abstract*—Web advertisements, an integral part of today's web browsing experience, financially support countless websites. Meaningful advertisements, however, require behavioral targeting, user tracking and profile fingerprinting that raise serious privacy concerns. To counter privacy issues and enhance usability, adblockers emerged as a popular way to filter web requests that do not serve the website's main content. Despite their popularity, little work has focused on quantifying the privacy provisions of adblockers.**

In this paper, we develop a quantitative approach to objectively compare the privacy of adblockers. We propose a model based on a set of privacy metrics that captures not only the technical web architecture, but also the underlying corporate institutions of the problem across time and geography.

We investigate experimentally the effect of various combinations of ad-blocking software and browser settings on 1000 Web sites. Our results highlight a significant difference among adblockers in terms of filtering performance, in particular affected by the applied configurations. Besides the ability to judge the filtering capabilities of existing adblockers and their particular configurations, our work provides a general framework to evaluate new adblocker proposals.

1. Introduction

Online advertising provides a viable way to support online businesses that offer content free of charge to their users, such as news, blogs and social networks. To achieve targeted and hence more effective advertising however, advertisers and tracking companies record user browsing behavior, e.g. pages viewed, searches conducted, products purchased [9], [15], [22], [27], [34]. Such techniques are known as *online profiling* and have raised significant privacy concerns because online user profiles can be used to infer private sensitive information and user interests [11], [12], [25], [28].

Adblockers aim to improve the user experience and privacy by eliminating undesired advertising content, as well as preventing the leakage of sensitive user information towards third-party servers. The most well-known adblocker solutions are browser extensions such as *Ghostery* or *Adblock Plus* which suppress unnecessary requests to third-party advertisements and tracking servers, thereby limiting the risk of data leakage towards these servers. Recently, users' privacy concerns and awareness about online profiling and tracking practices have strongly increased, leading to a proliferation of adblocker browser extensions in the wild. According to Mozilla and Google usage statistics [2], [4], already more than thirty million surfers are actively using a browser with the Adblock Plus extension enabled. In a recent measurement study [30], researchers show that 22% of the most active users are using the Adblock Plus adblocker while surfing the Web.

Despite the popularity of adblocking tools, surprisingly little research has been performed to understand how well adblocking actually improves the privacy of its users. While the methods employed in advertisement and tracking and their privacy implications have been well researched in the literature [14], [21], [23], [29], the protection that adblockers offer, has not been investigated that much in the literature. Works such as [10], [20], [30], [33] analyze ablockers's performance, however the impact of user privacy is not in the main scope of these studies, as they focus on the effectiveness of the adblocker's implementations and the usage in the wild. Understanding how adblockers affect user privacy is fundamental to their use, because it not only provides feedback to the users, but also helps at correctly using and configuring those systems. Adblockers rely on complex filter configurations in the form of blacklisted URLs and regular expressions, and as we show in this paper, existing adblockers are not necessarily configured by default to provide the best privacy protection to their users.

Our goal in this work is to quantify the privacy that web adblockers provide. We address this problem by developing a quantitative model to compare adblocker filtering performance across various privacy dimensions. Our model includes simple count metrics to third-parties, but also considers more advanced metrics on the level of organizations (legal entities) and countries as well as their relationships. In order to also understand the temporal dynamics of the system, we further incorporate temporal metrics to track the filtering performance over time.

We have developed a testbed system which allows us to repetitively browse the same Web sites in a systematic way and classify the number of HTTP requests that go to first and third parties without any classification errors. We evaluate 12 different browser profile configurations in our testbed,

capturing different adblocker instances and combinations of desktop/mobile user client agents. During three weeks, we repetitively surfed Alexa's top 500 global sites and 500 randomly selected sites and analyzed how different configurations influence these privacy metrics.

Our results show that the usage of adblockers provides a significant improvement in terms of user privacy. However, the degree of protection is highly depending on the configuration. For example, by default Ghostery does not block any third-party requests and Adblock Plus still allows a significant amount of requests to third parties. These results are consistent for the desktop and the mobile user agents. When increasing the level of protection in Ghostery and Adblock Plus however, these tools manage to effectively suppress requests to third-parties and thus improve the privacy. Except for Google Inc. which still receives around 50 % of third-party requests because it hosts relevant content not related to advertisement and tracking, the amount of third-party requests towards the other top ten companies in our experiments is only 2.6 % of the total amount that would result when surfing without an adblocker.

Our contributions in this paper can be summarized as follows:

- We provide a quantitative methodology to objectively compare the filtering performance of web adblockers.
- We capture the temporal evolution of adblocker filtering performances and study the differences between mobile and desktop devices, as well as the impact of the *do not track* header. Our methodology further allows to measure the influence of other parameters (e.g. third-party cookies) on adblocker filtering performance.
- Beyond the domain of the third parties, our model takes into account the underlying legal entities, their corresponding geographical locations as well as their relationships.
- Using our model, we quantify the privacy of 12 different adblocker browser profile configurations over 1000 different Web sites for repetitive daily measurements over the duration of three weeks and discuss the implications in terms of user protection.

The remainder of the paper is organized as follows. In Section 2 we illustrate the objective and functionality of adblockers, while in Section 3 we outline our privacy metrics. Section 5 discusses the experimental setup and the results. Section 7 presents the related work and Section 8 summarizes our work.

2. Web Tracking and Adblockers Background

This section provides relevant background on third-party tracking in the web and how adblocker browser extensions aim at improving user experience and privacy.

2.1. Third-party Tracking

When visiting an HTTP-based website on a domain (commonly referred to as first party), the web browser sends an HTTP request to the first-party server that hosts the website and loads the content of the first-party domain. The HTML code of the first party is then able to trigger (without the awareness of the user) further HTTP requests to remote servers (commonly referred to as third parties) in order to load further resources that they host. External resources vary in their format and are applied with different objectives, such as the inclusion of external libraries —e.g. jQuery— that are indispensable for the functionality of the website itself. Further reasons include the promotion of advertising content that can be externally loaded and placed at a pre-allocated space on the website.

This third-party content loading mechanism clearly facilitates the development and deployment of dynamic websites because it allows to use different content providers to load resources that do not need to be served from the first party. However, as shown in previous works [9], [22], [27], [34], HTTP requests to third-parties lead to severe privacy implications because third parties can follow the activity of the users and reveal the pages they are looking at while surfing the web. For example, it has been shown in [15] that dominant players in the market such as Google Inc. are embedded as third-parties in so many web sites that they can follow 80 % percent of all web activities. Since the web page content and thus user interests can be inferred by the uploaded requests to the third parties, personal profiles of users can easily be derived and potentially used to discriminate people or spy on their interests and habits without getting noticed by the users.

2.2. Adblocker browser extensions

To address the aforementioned implications and challenges, numerous software and hardware-based solutions — commonly referred to as *adblockers* — have been proposed in order to remove or alter the advertising and third party content in a web page. Although there exist multiple adblocking methods (e.g. DNS sinkholing, proxies run by internet providers (externally) or by an application on the same client machine, special hardware) we focus in this work on one of the most popular solutions: browser extensions, such as *Ghostery* and *AdblockPlus*.

Adblocker browser extensions use one or more lists that describe the content that is to be allowed (whitelists) or blocked (blacklists) and update those on a regular basis. There are two principal methods how adblockers apply these lists to remove ads/third parties from a web page: One is filtering the resource according to the result of an URL-pattern matching, before this resource is loaded by the web browser. The second consists in hiding loaded content with the use of CSS rules (*element hiding*) within the HTML content. In terms of privacy, filtering the resources before they are requested by the browser is the only effective method because these requests are the ones revealing the activity of the users.

Adblocker browser extensions are very popular by users today and their popularity is continuously on the rise [2], [4], [30]. However, content providers and advertisers see this trend as a risk to their own business models because

they regard the application of theses tools as a way for the comsumers to evade "paying for the content". Juniper Research estimates that digital publishers are going to lose over 27 billion dollars by 2020 due to the use of ad blocking services [3]. There is therefore high pressure by these industries on the developers of adblockers to not blacklist their services. For example, Adblock Plus has introduced in 2011 the concept of "non-intrusive advertising", which basically allows third-party advertisements for ads which do not *disrupt the user's natural reading flow* [1]. However, these practices raise concern in terms of privacy because non-intrusive advertisement services may well perform intensive tracking without falling in this category. We therefore argue that it is important to quantify independently the privacy of these tools as we do in this work.

3. Privacy Model and Metrics

In this section, we introduce our privacy model and the metrics we use in order to quantify the privacy provisions of adblockers.

3.1. Threat definition

A key issue for a threat model in adblocking is to define which third-parties should be considered as a privacy threat to users. In this work, we consider all third-parties as potential threats irrespective of the type and content of the queries towards these third parties. This approach may arguably seem conservative, but it is practically impossible to exclude for sure any third-party from performing tracking and/or profiling given the multitude of possible mechanisms that are available and continuously invented for fingerprinting and tracking user behavior in the web.

In our notion, the privacy objective of the abblocker is therefore to reduce as many requests as possible towards third parties. Notice here the difference of our threat model definition to the slightly different objective that adblockers such as Adblock Plus have. By default, Adblock Plus aims at improving user satisfaction by minimizing the display of intrusive advertisements which annoy the users while third-party requests to non-disturbing advertisements and tracking services for commercial purposes are considered to be acceptable [1].

3.2. User tracking model

We model the tracking of a user U through third parties as undirected graph $G = (E, V)$, where E are edges, and V vertices. A vertex V_S represents a web domain and is connected to another vertex V_T through an edge E, if and only if at least one request has been sent from V_S to V_T. In that case, V_S is the *source* of the request and V_T the *target* of the request.

In the following, we use the term *third-party request* (TPR) to denote the requests that are sent to a target domain T that differs from the source domain S and corresponds

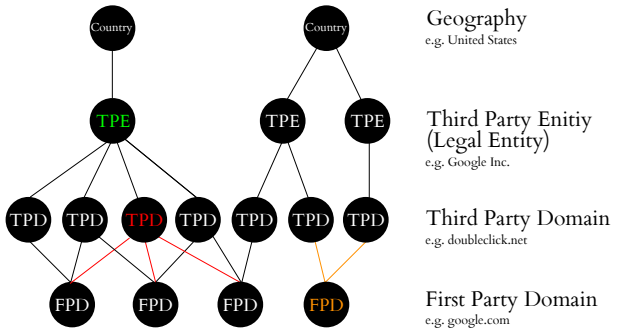


Figure 1: Graphical representation of our user tracking model. The colored third-party domain (TPD) has a node degree of 3, the colored first-party domain (FPD) has a node degree of 2. The colored third-party entity (TPE) spans all its child TPD nodes and hence has a degree of 3.

to a graph edge E between the nodes V_S and V_T. On the contrary, the requests whose source and target coincide are designated as *first-party requests* (FPR) and are not taken into consideration for the construction of G, since no information leaks to third parties and hence they do not bring about further risks for user's privacy[1]. The source and the target domain are referred to as *first-party domain* (FPD) and *third-party domain* (TPD) and correspond to FPD and TPD graph nodes, V_S and V_T, respectively.

Compared to previous works on third-party traffic characterization [10], [14], we augment G by incorporating the ownership of third party domains to their corresponding legal entities, i.e. the organizations who own the different TPDs. Two TPD, belong to the same legal entity if they are registered to the same organization (e.g., doubleclick.net and google-analytics.com both belong to Google Inc.) and are thus combined into one vertex, resulting in a hierarchical graph (cf. Figure 1). Considering the information flow of third-party requests towards legal entities is particularly important for the scope of privacy because legal entities which own multiple domains can fuse the information they collect from their different domains in order to increase their tracking and profiling coverage, thus resulting in a higher privacy threat to the users.

Finally, we further attribute each legal entity to a geographical location (the country where the headquarter of the legal entity is situated) in order to model which countries govern the regulations over which legal entities. This geographical perspective is also of special importance to privacy, because most data privacy laws are specific to local laws of the countries, thus affecting the regulations that apply to the user data that is collected by the legal entities.

1. Arguably, users also leak private information to first party domains when they visit and interact with those sites, however, since users are visiting these first parties deliberately, the privacy risks are known to the users and controllable without an adblocker.

3.3. Privacy Metrics

Given the graph representation G of our user tracking model, we evaluate the respective privacy provisions based on the following metrics.

3.3.1. Degree of First Party Domain. The degree of a FPD node of graph G refers to the number of TPDs that it has sent at least one third-party request to when loading the web page from the FPD. That is, the more edges a FPD node has — or, equivalently, the more third parties loaded by a first-party — the more third parties are able to track the web-browsing history of a user. The FPD node degree is a metric that is commonly used to evaluate the adblocker's performance [33]. However, it is alone not a sufficient metric to capture the impact on user privacy, as it does not represent the structure behind the relationships between FPD and TPD. The following metrics therefore aim at capturing these relationships.

3.3.2. Degree of Third Party Domain. The degree of a TPD node can be directly translated to the number of first-party websites that a particular third party exchanges information with and potentially tracks. Clearly, the more often a third party is accessed over the user's series of websites S_U, the less privacy the user experiences from this particular third party. To exemplify this statement, let's assume that a third party is requested by only one of the first-party websites S_U visited by U. This third party will in this case learn that the user has accessed the respective first party, but has a limited view of their browsing behavior. If the third party, however, is requested by over 80% of the user's visited websites, S_U, the third party will likely be able to recover up to 80% of the web behavior of U.

3.3.3. Degree of Legal Entity. Instead of focusing on domain degrees, the degree of a legal entity reflects the number of third-party domains that belong to a legal entity. Third-party domains such as doubleclick.net and google.com for example are both owned by the same entity Google Inc. Their collusion therefore seems more likely, and affects the privacy of a web user U more significantly, than if both were belonging to two different legal entities. By incorporating the legal relation among third party domains, we therefore capture a more realistic privacy leakage through user web surf activity.

3.3.4. Geographical location. After having mapped the TPD's to legal entities, we further assign a geographical location to the TPD. This allows our model to capture the geographical distribution of the TPDs and thus infer which geographical countries have for instance the most TPD. The geographical location of a legal entity is defined by the country in which its headquarter resides. Alternatively, we could consider the particular location of the servers as derived from the IP address, but content retrieved from web services is often hosted on distributed caches and content distribution networks and hence the server IP address does not necessarily reflect the country to which the user data is finally sent to. By choosing the headquarter's location, we thus aim at modelling the country in which the privacy laws and regulations will apply to the user data as collected by the third-party.

3.3.5. Graph Density. In addition to the degree metrics outlined above, we consider a metric based on the graph density of G. Since an edge on the graph G represents a partial tracking relationship between a third and a first party, we expect that the denser the graph G, the more information can be retrieved by third parties/can leak to third parties with respect to the browsing behavior of the user. We observe that the more dense G is, the more third parties are likely able to track the user U. The graph density therefore allows to reason about the possible privacy improvements by the respective ad-blocking software. We rely on a common definition of the graph density as:

$$D = \frac{2|E|}{|V|(|V|-1)} \tag{1}$$

Note however that we cannot achieve the maximum density of 1, because the first parties in G are not directly connected (cf. definition in Section 3.2).

4. Evaluation Methodology

In order to compare the privacy of different adblockers, as well as the influence of different browser settings on their adblocking efficiency, we create different browsing configurations without adblockers, with the Ghostery, and with the Adblock Plus browser extenstions installed in the Firefox browser.

4.1. Considered Browser Profiles

All our experiments are performed on "Linux (Release: Ubuntu 14.04.4 LTS, Version: 4.2.0-35-generic GNU/Linux)" with the version 45.0.1 of the Firefox browser. For Ghostery, we use the browser plugin version 6.1.0 and for Adblock Plus the plugin version 2.7.2. The different protection levels, *Default* or *MaxProtection*, for the two adblockers *AdblockPlus* and *Ghostery* respectively, are achieved through the use of a different combination of blacklists. AdblockPlus and Ghostery store their respective blacklists in the form of URL and CSS regular expressions. The blocking options of AdblockPlus are set through the direct inclusion of blacklists to be applied, while Ghostery's blacklist configuration consists in the selection among a multitude of tracker categories to be blocked. An overview of these configurations is presented in Table 1.

Modern web browsers such as Firefox further allow to set the *do not track* HTTP header option, to express their personal preference regarding tracking to each server they request content from, thereby allowing recipients of that preference to adjust tracking behavior, accordingly [32]. It remains the sole responsibility of the web server to

Protection Level	Lists			
	AdServers	EasyList	EasyListChina	EasyPrivacy
Default		✓		
Maximal	✓	✓	✓	✓

TABLE 1: AdblockPlus blacklist combination for default and maximal protection level. Ghostery's default and maximal protection correspond to the selection of none and all tracker categories, respectively.

respect the request of its clients. Almost 10% of the Firefox users have enabled this option on their desktop browsers in 2014 [5]. In order to evaluate to which extend the DNT header has an influence on our proposed metrics we as well include the DNT option in our evaluation.

The usage of mobile devices for web browsing has recently witnessed a steady growth [6]. As a consequence, an ever increasing number of websites has been adapting to the demands of the mobile user agents. Because of the dimensions and the reduced-bandwidth requirements of the mobile devices, the structure and content of the web pages has to be adjusted accordingly and the advertising content could not remain unaffected by these limitations. To investigate the effects of user agents from a privacy-related perspective, we consider this parameter in the design of the experimental evaluation and evaluate several mobile-device instances by setting the HTTP header *User-Agent* accordingly.

Based on above mentioned criteria, we create 12 browser profiles, U as described in Table 2. Each configuration is defined as a combination of the following parameters :

- Adblocker: No adblocker, Ghostery, or Adblock Plus
- Block policy: maximum or default protection
- User agent: mobile or desktop
- Do Not Track (DNT): header enabled or disabled

Throughout the remaining of the paper, we use the following conventions for each browser profile U (cf. Table 2):

- The *color* denotes the adblocker installed.
- The *line width* indicates the protection degree —i.e. default, maximum protection or DNT header.
- Profiles with Mobile User Agent are plotted in *dashed lines*.

4.2. Experimental Setup

The distinction between FPRs and TPRs is crucial in our attempt to precisely quantify the filtering capability for each browser profile, since they define the exact topology of the derived graph G. Passive classification of HTTP requests into first-party and third party requests is not a trivial task given the complex and dynamic structure of Web pages [30]. For this reason, we rely in this work on an active approach in which we collect our own synthetic web surfing traffic with automated web surfing agents. To create a realistic and representative dataset, the agents visits Alexa's top 500 web sites (the 500 domains with the highest incoming traffic in the web) and 500 web sites which are sampled uniformly among Alexa's top 1 million most-visited domains. The motivation for including less popular web sites is to avoid the risk of favoring an adblocker optimized to perform best for the most popular web sites, eventually biasing the experimental results. The overall sample set S of 1000 URLs is retrieved once and kept unchanged throughout the evaluation period, so as to de-correlate any variations of the results between different days.

Since nowadays most web applications are based on asynchronous calls to fetch data, it is insufficient to wait for the DOM to finish rendering to record all resource requests sent from the website to any first or third parties. To collect the complete data and better evaluate the common user browsing behavior, our agent therefore waits 20 seconds on each website of our sample set S and records any requests sent, before closing and proceeding to the next domain.

We visit the same set of web sites every day during three weeks from 28/04/2016 until 19/05/2016. To decouple the experimental conditions from the influence of any time- or location-related effects —i.e. variations of the served content, locale-based personalization— all browser profiles U execute the same crawling routine simultaneously, whilst running on the same machine, thus behind the same IP address, browser and operating system. However, some of the instances are configured to send their requests with a User-Agent HTTP header that corresponds to a mobile device (iPhone with iOS 6[2]), in order to extend our observations for mobile users.

In order to record all HTTP requests, we rely on the *Lightbeam* plugin. However in contrast to [33], we do not use Lightbeam to determine the source domain that a request is initiated from and to classify it accordingly as a FPR or a TPR because Lightbeam relies on heuristics that are too error-prone for our purpose. More precisely, the classification of Lightbeam is not always in accordance with our definitions of FPR and TPR, as introduced in Section 3. By examining the request logs after a complete crawl cycle and comparing the estimated source to the actual visited domain, two types of false-positive cases (cf. Table 3) arise in Lightbeam:

- **Unrecognized TPRs:** The request is mistakenly considered to be a FPR according to the Lightbeam heuristics, this way "hiding" a TPR edge from the graph.
- **Misclassified TPRs:** The request is correctly found to be a TPR, but not for the correct FPD node, i.e. the one corresponding to the actually crawled domain. The inaccuracy introduced to the graph results from the potential introduction of a bogus FPD node, as well as

2. User Agent: `Mozilla/5.0 (iPhone; CPU iPhone OS 6_0 like Mac OS X) AppleWebKit/536.26 (KHTML, like Gecko) Version/6.0 Mobile/10A5376e Safari/8536.25`

Browser Profile	Adblocker	Block Policy	DNT	User Agent	Legend
Ghostery_Default	Ghostery	Default	No	Desktop	
Ghostery_MaxProtection	Ghostery	Max	No	Desktop	
Adblockplus_Default	AdblockPlus	Default	No	Desktop	
Adblockplus_MaxProtection	AdblockPlus	Max	No	Desktop	
NoAdblocker	None	-	No	Desktop	
NoAdblocker_DNT	None	-	Yes	Desktop	
Ghostery_Default_MUA	Ghostery	Default	No	Mobile	
Ghostery_MaxProtection_MUA	Ghostery	Max	No	Mobile	
Adblockplus_Default_MUA	AdblockPlus	Default	No	Mobile	
Adblockplus_MaxProtection_MUA	AdblockPlus	Max	No	Mobile	
NoAdblocker_MUA	None	-	No	Mobile	
NoAdblocker_DNT_MUA	None	-	Yes	Mobile	

TABLE 2: Overview of browser profiles examined

	Visited Domain	Estimated Source	Target
Recognized	wp.pl	wp.pl	facebook.com
Misclassified	wp.pl	facebook.com	fbcdn.net
Unrecognized	wp.pl	facebook.com	facebook.com

TABLE 3: Examples of misclassified and unrecognized TPRs

the false number of TPR edges starting from the correct and the bogus FPD nodes.

As results from the experimental evaluation on the data of one full crawl cycle (1000 visited first parties) and 12 different browser profiles, the misclassified and unrecognized TPRs make up for 2.0%-12.0% and 4.0%-11.0% of the total requests, depending on the respective browser profiles that we define in the following.

We thus modify Lightbeam to account for the currently visited first-party as a priori known by the agent which triggers page visits.

4.3. Classification of Domains to Legal Entities and Locations

We infer the legal entities' domains and locations by inspecting the WHOIS database. The WHOIS database provides information about the holders of Web domains. For each domain, we look up the legal entity that is registered as holder and the country of the holder's address. Note that only a part of the considered domains —accounting for about 60%— could be assigned to a legal entity and followingly to a country. One reason is that WHOIS does not provide sufficient information for all of the domains loaded. Moreover, our parser that allowed for the automated extraction of the entity information depends on a relatively uniform format of the WHOIS documents and as a result, deviations from this format causes information loss.

5. Evaluation

We examine the impact of the configuration parameters on the achieved privacy level using our privacy metrics from Section 3.

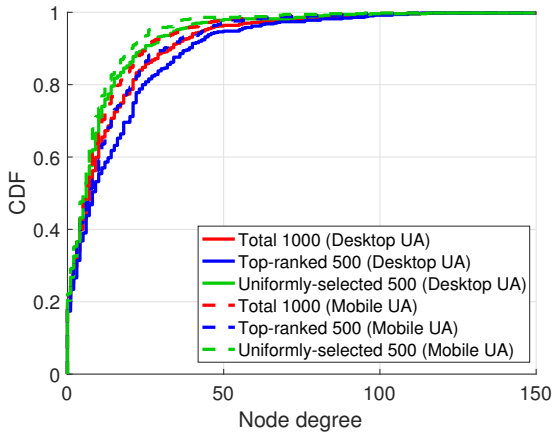


Figure 2: FPD node degree for the browser profiles *NoAdblocker* (solid line) and *NoAdblocker_MUA* (dotted line) on 28/04/2016.

5.1. Effectiveness of Adblockers at Suppressing Third-party Requests

5.1.1. Baseline without adblocking. Before investigating the effect of the different adblockers, we characterize the FPD node degree with the *NoAdblocker* and *NoAdblocker_MUA* browser profiles as a baseline. Figure 2 shows the cumulative distribution function (CDF) of the FPD node degree of both profiles on a single day (28/04/2016) for the top-ranked 500 domains and the 500 uniformly-selected ones. As can be seen, in both the top 500 and the uniformly selected domains, almost 20% of the websites did not load any third-parties at all. These domains do therefore not impose a privacy risks to the users. On the other hand, more than 80 percent of the visited domains generate requests to third parties. In general, we can say that the top 500 domains tend to generate more requests to third-parties than the uniformly selected domains, indicating that advertisement and tracking is more likely to happen on popular domains. However, even the randomly selected domains have a quite significant number of third-party requests. While the mean FPD node degree for the top 500 domains and uniformly selected domains are around 17 and 12 respectively, both

FPD node degree distributions has a quite long tail. We observe a significant number of FPD node degrees above 100 with one domain in the top 500 exhibiting a degree of 180. These sites raise serious concerns in terms of privacy since each individual third-party request could potentially leak personal information of the visiting users to these third parties.

5.1.2. Comparison of the Different Browser Profiles. To understand the effectiveness of the different adblockers and browser profiles at suppressing requests to third-parties, we plot in Figure 3 the FPD node degree distribution for all domains as a CDF. Figure 3a shows the node degree distribution averaged over the different days while Figure 3b represents the standard deviation of the node degree over the same days. Our results indicate the following findings.

The worst filtering performance is achieved with the *do not track* HTTP header options (*NoAdblocker_DNT* and *NoAdblocker_DNT_MUA*) and Ghostery in default mode (*Ghostery_Default* and *Ghostery_Default_MUA*). With these browser profile configurations, almost none of the third-party requests are blocked. AdblockPlus (*Adblockplus_Default* and *Adblockplus_Default_MUA* with its default settings has a FPD node degree that is significantly lower than the aformentioned cases, i.e., the browser profiles with the DNT header enabled and Ghostery in its default configuration. Unsurprisingly, the browser profiles that filter the most third parties are those with adblockers configured to a maximum protection level. We observe that *Ghostery_MaxProtection* decreases the mean FPD node degree by approximately 80 % compared to *NoAdblocker*. On the other hand, the FDP node degree of Adblock Plus (*AdblockPlus_MaxProtection*) is reduced by almost 75 % which is slightly behind the performance of Ghostery, but still significantly better than the default configuration option.

Interesting to note here is the large difference in blocking performance between the different configurations of the same adblockers. This result suggests that the privacy of the users is highly affected by a good configuration of the tools and that by default, these tools still permit a significant portion of the third-party requests.

The standard deviation of the FPD node degree over all domains is shown in 3b. As we can see, the profiles which have a large FPD node degree tail such as *NoAdblocker*, *NoAdblocker_MUA*, *NoAdblocker_DNT*, *NoAdblocker_DNT_MUA*, and *Ghostery_Default* also exhibit this tail in the standard deviation. However, the profiles which tend to have a small FDP node degree feature a small standard deviation as well.

5.1.3. Temporal Dynamics. To capture the temporal dynamics of third-party requests, we plot in Figure 4 the FPD over time in the considered period of 3 weeks. Figure 4a and 4b show the mean FPD for the top 500 domains and the uniformly selected domains respectively. We observe a quite stable temporal evolution over the individual days for both datasets. In particular, in none of the datasets, we can observe a change in relative order between the different browser profiles. We can therefore conclude that in general, the privacy of the users is not sensitive to web site or blacklist optimizations that happen at shorter time scales.

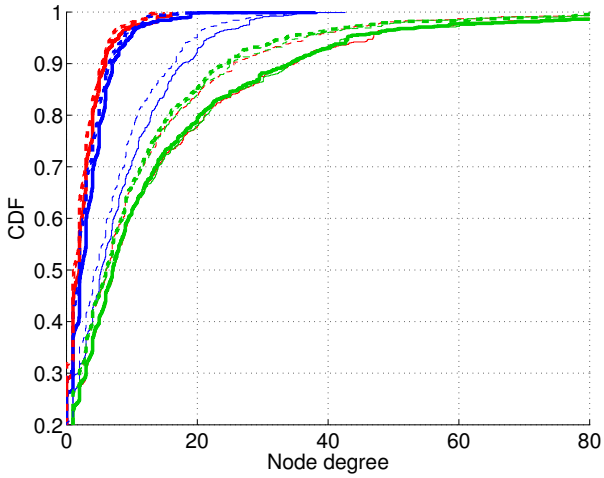
Third-Party Domain	Legal Entity	TPD Degree		
		None	Ghostery	AdblockPlus
doubleclick.net	Google Inc.	486	0	1
google-analytics.com	Google Inc.	476	4	0
google.com	Google Inc.	383	93	144
facebook.com	Facebook Inc.	318	5	164
gstatic.com	Google Inc.	308	226	235
googlesyndication.com	Google Inc.	204	0	0
google.ch	Google Inc.	189	0	0
fonts.googleapis.com	Google Inc.	185	145	141
adnxs.com	AppNexus Inc.	159	0	0
facebook.net	Facebook Inc.	157	0	140

TABLE 4: Top-loaded TPDs for browser profile *NoAdblocker* and the corresponding values for Ghostery and AdblockPlus with maximum-protection settings (browser profiles *Ghostery_MaxProtection* and *AdblockPlus_MaxProtection*) on 28/04/2016

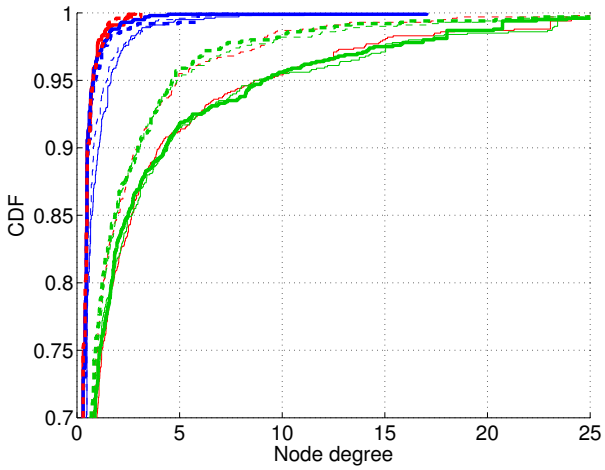
To check whether this conclusion also translates to individual domains, we take a closer look at the domains with the highest FPD in Figure 4c and 4d. Figure 4c shows the evolution of the FPD for the domain with the highest FPD in any of the dataset while Figure 4d represents the mean of the FPD over the ten domains with the highest FPD. We make two interesting observations here. First, the domains with the largest FPDs tend to exhibit a higher variation over different days. In particular, for *Ghostery_Default* and *Ghostery_Default_MUA* in Figure 4c, the filtering of third-party requests shows a larger fluctuation over time. Also, *AdblockPlus_MaxProtection* and *AdblockPlus_MaxProtection_MUA* has a significantly higher fluctuation for the top domain than on average. Second, the filtering performance of the different browser profiles is more clustered than it is was on average for all the domains. For example, on most days, the performance of *Ghostery_Default* and *Ghostery_Default_MUA* is almost identical to *NoAdblocker*, while those two profiles where significantly outperforming the *NoAdblocker* profile in Figures 4a and 4b. These two observations indicate that these domains with a high FPD score could be more active at circumventing blocking strategies by adblockers.

5.2. How do Adblockers Reduce the Tracking Range of Third-party Domains?

In order to understand the extend to which individual third-parties are able to track users while surfing across different domains, we look next at the degree of third-party domains (TPD). The TPD degree reflects how many visits to different first-party domains an individual third-party can observe. Figure 5 shows how the different browser profiles affect the TPD degree of all third-parties that we encountered in our experiments. As we can see, the TPD is highly skewed. Only 10 percent of the third-parties have a TPD of more than 10 for the *NoAdblocker* profile while



(a) CDF of the average node degree over 3 weeks.



(b) CDF of the standard deviation of the node degree over 3 weeks.

Figure 3: FPD node degree distribution for all browser profiles. Legend is provided in Table 2.

the largest TPD degree we observe is 486 (*None* column of Table 4). In general, we can therefore say that a small number of third-party domains are able to capture the vast majority of the visits to first parties.

Considering the effect of the different browser profiles, we observe a similar trend as for the FPD degree. The *Ghostery_MaxProtection* and *AdblockPlus_MaxProtection* profiles manage to effectively reduce the TPD node degree of all domains. However, in their default settings, AdblockPlus and Ghostery have only a noticeable effect on the domains with a small TPD degree, while these profiles have almost no impact on the filtering performance of domains with a large TPD node degree. Again, the browser profiles with the Do Not Track option enabled result in similar TPD node degrees as without the option.

In Table 4, we list the 10 domains with the highest TPD node degree (when no adblocker is applied) and compare how these numbers decrease with the *Ghostery_MaxProtection* and *AdblockPlus_MaxProtection* browser profiles. Ghostery achieves generally better performance, although AdblockPlus outperforms marginally Ghostery for 2 domains. Interesting to notice here is that some third-party domains from this list still exhibit a high TPD node degree with any of the adblockers enabled. These are the domains google.com, gstatic.com, and fonts.googleapis.com. These domains provide important content to render the web pages of the first parties and can therefore not be blocked. The other domains relate to advertisements, tracking, and social media and their TPD degrees are effectively reduced by Ghostery. AdblockPlus is not so effective at reducing the TPD degree of domains such as facebook.com and facebook.net.

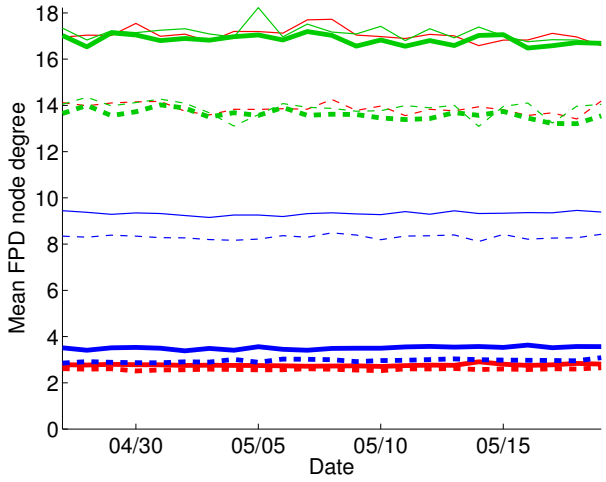
Legal Entity	Degree		
	None	Ghostery	AdblockPlus
Google Inc.	666	328	354
Facebook Inc.	328	6	211
AppNexus Inc.	159	0	0
TMRG Inc.	143	0	4
Twitter Inc.	137	9	87
Oracle Corporation	123	2	39
Adobe Systems Incorporated	107	6	32
Yahoo! Inc.	99	7	5
AOL Inc.	88	3	3
OpenX Technologies	88	0	0

TABLE 5: Legal entities with the highest TPE node degree for browser profile *NoAdblocker* and the corresponding values for Ghostery and AdblockPlus with maximum-protection settings (browser profiles *Ghostery_MaxProtection* and *AdblockPlus_MaxProtection*) on 28/04/2016.

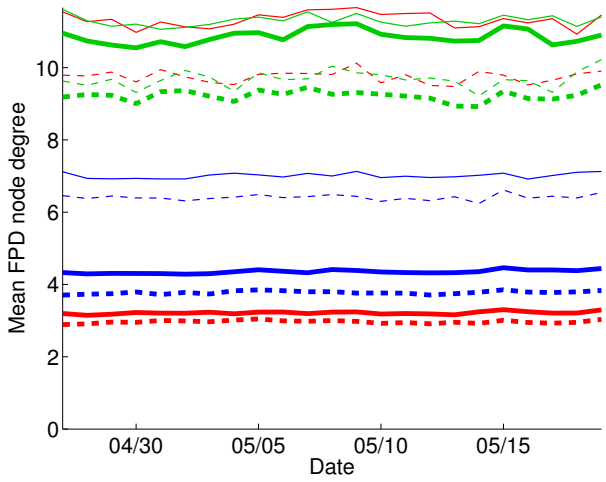
5.3. How do Adblockers Reduce the Tracking Range of Legal Entities?

As we have seen in Table 4, the TPD degree of many domains was effectively reduced with adblockers, but some domains still remain with a high TPD node degree, mostly in order to provide useful content when rendering the page of the FPD. As a next step, we aim to understand how adblockers reduce the tracking range at the level of legal entities. A legal entity may acquire multiple domains and therefore still receive a lot of third-party requests despite some of its domains being blocked by the adblockers.

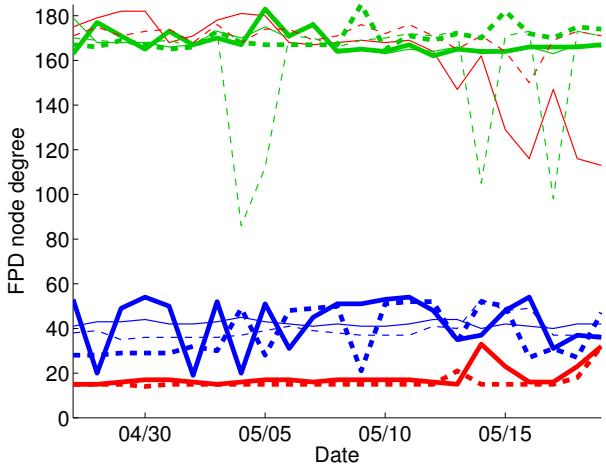
Table 5 summarizes the 10 legal entities with the highest TPD node degree, i.e. that were present on most of the visited URLs when the default Browser settings were applied (*NoAdblocker*). As the data suggests, domains owned by *Google Inc.* are loaded by 674 out of the 1000 URLs visited, thus having the most frequent presence among the



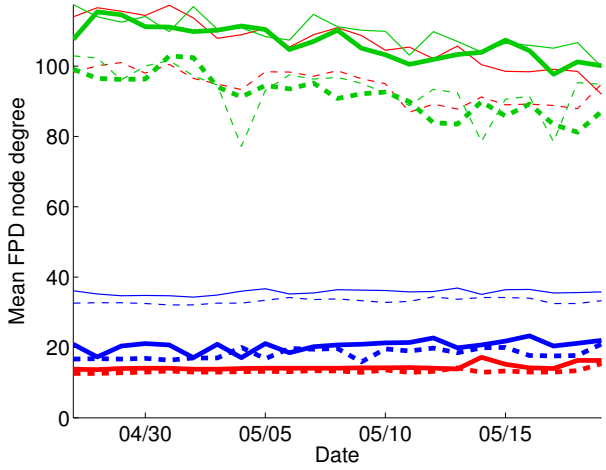
(a) Mean FPD over the top 500 domains.



(b) Mean FPD over the 500 uniformly-selected domains.



(c) Maximum FPD over all visited domains.



(d) Mean FPD over the 10 domains with the highest FPD from all visited domains.

Figure 4: Evolution over time of the first party node degree (FPD). Legend is provided in Table 2.

rest of the third-party entities. Followed by Google Inc. are Facebook Inc., AppNexus Inc., and TMRG Inc. with node degrees of 328, 159, and 143 respectively. The degree of the following domains then quickly drops below 100.

Also presented in Table 5 is the node degree of the top 10 legal entites with the *Ghostery_MaxProtection* and *AdblockPlus_MaxProtection* browser profiles enabled. Except for Google Inc., Ghostery is able to suppress the node degree of all top 10 legal entities below 10. Google Inc. however remains with a node degree of 328, meaning that despite using Ghostery, Google Inc. is able to track more than 30 percent of the page visits to the FPDs. AdblockPlus is significantly less effective than Ghostery even in the maximum protection mode. Still, it reduces significantly the TPD node degree for most TPDs.

Country	First-Party Entities
United States	35.7 %
Canada	7.4 %
Japan	4.8 %
Switzerland	4.0 %
Germany	3.8 %
India	3.5 %
Great Britain	3.0 %
Russia	2.6 %
France	2.6 %
Panama	2.0 %

TABLE 6: Countries hosting the highest percentage First-Party Entities

5.4. Geographical Considerations

Another key privacy dimension is the geographical location to which third-party requests are transferred to since

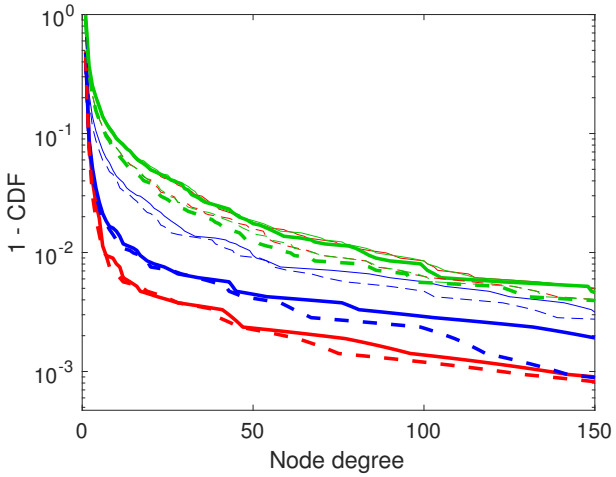


Figure 5: Inverse CDF of TPD node degree for all browser profiles.

Country	Third-Party Entities		
	None	Ghostery	AdblockPlus
United States	784 (45%)	483 (42%)	500 (45%)
Germany	106 (6%)	40 (4%)	34 (3%)
China	82 (5%)	70 (6%)	67 (6%)
Japan	80 (5%)	62 (5%)	61 (6%)
Great Britain	77 (4%)	43 (4%)	44 (4%)
France	69 (4%)	33 (3%)	31 (3%)
Canada	49 (3%)	33 (3%)	28 (3%)
India	46 (3%)	38 (3%)	38 (3%)
Panama	41 (2%)	32 (3%)	25 (2%)
Turkey	32 (2%)	27 (2%)	27 (2%)
Total	2908	1866	1812
Found	1748 (60.1%)	1140 (61.1%)	1097 (60.5%)

TABLE 7: Countries hosting the highest percentage TPEs when no adblocker is used (browser profile *NoAdblocker*), and the corresponding percentages when Ghostery and AdblockPlus are used under maximum protection settings (browser profiles *Ghostery_MaxProtection* and *Adblockplus_MaxProtection*) on 28/04/2016.

local regulations govern what legal entities may do with the personal data that they collect about users. Table 6 lists the 10 countries with the highest number of legal entities acting as first party in our traces. The country with the most first parties is the United States (35.7%) followed by Canada (7.4%) and Japan (4.8%). Figure 6a visualizes the relative number of legal entities acting as third parties in each country. The darkest regions (red) are the countries with the most TPEs loaded, while the white ones host none of the TPEs found in our graphs. As we would expect, the USA hosts most of the first and third-party domains, while regions such as Africa or Latin America contain very few TPEs.

A more detailed view of the number of TPEs hosted by the top 10 countries is presented in Table 7. For each row, the absolute numbers refer to the TPDs that were recognized and assigned to a TPE for the specific country, while the percentages refer to the ratio of these TPEs over the total number of TPEs that were recognized by our automated script. In this table, we compare the TPEs hosted by each of these countries (column *None*) to the number of TPEs loaded when the adblockers Ghostery and AdblockPlus are deployed under maximum-protection settings (columns *Ghostery* and *AdblockPlus*).

Interesting to note here is the difference in rank between countries in terms of legal entities that act as first and third parties. For example, China does not appear in the top ten list of countries for first parties, but ranks third in the ranking for legal entities that act as third-parties. This indicates that China hosts in relation to the other countries more third-party domains than first-party domains. The opposite is true for Switzerland and Russia which rank 4th and 7th in the ranking for first-party entities but don't appear in the top ten of third-party entities. Regarding the effect of the Ghostery and AdblockPlus, we can see that these adblockers do not significantly affect the overall distribution and ranking of the third-party legal entities. All countries experience a diminishing number of third-party legal entities that is in proportion relatively equal.

5.5. Graph Density

As in Figure 7, grouping the TPD nodes according to the legal entities they belong to brings a considerable reduction of the mean FDP node degree, asserting that the number of legal entities potentially collecting information about the user is indeed less than that of the actual third-party domains tracking them.

On the contrary, the mean TPD node degree, as well as the graph density do not present any significant variation, which leads us to the conclusion that the various legal entities have on average access to roughly the same first parties, although controlling multiple third-party domains.

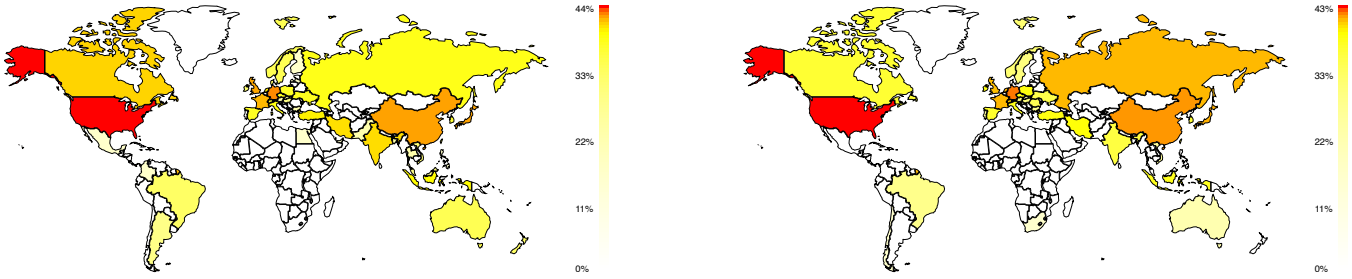
6. Discussion

This section summarizes our findings and discusses the influence of profile parameters on the user privacy:

Adblocker installed: Our results show that the use of an adblocker significantly reduces the number of third parties loaded by a factor of 40% for an adblocker with default settings. By default Ghostery does not function as adblocker, while with maximum protection, Ghostery consistently outperforms AdblockPlus.

Block policy: The block policy configured —i.e. default or maximum protection— for each adblocker results in a blocking difference of almost 80% and 50% for the mean first-party degree for Ghostery and AdblockPlus, respectively.

Do Not Track header: The activation of the DNT flag has little impact on the results, since the browser user has no control over whether the DNT flag is honored or not and hence websites and advertisers may either obey or completely ignore it.



(a) Third-party legal entity locations

(b) Third party server locations

Figure 6: World map depicting the locations of the legal entities and the servers for the third parties loaded during our experiments.

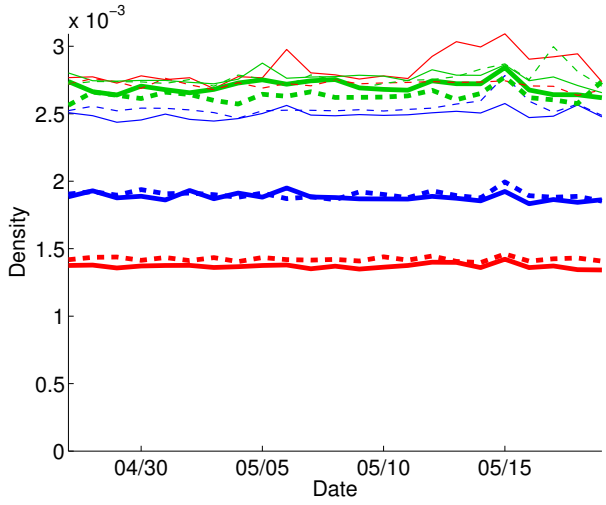


Figure 7: Time evolution of the graph density for all browser profiles.

Mobile User Agent: Websites that received requests by a mobile user agent indeed responded with a considerably lower number of third parties. A plausible explanation for this behavior is the requirement for less bandwidth that the mobile websites usually conform to. As a consequence of this limitation, less content is loaded and hence less requests are sent to third parties. This discrepancy was of course less obvious for the browser profiles with maximal protection enabled, since tracking by third parties is already reduced. To exemplify, in Figure 4a the profiles *Ghostery_Default*, *NoAdblocker* and *NoAdblocker_DNT* present a mean FPD node degree about 20% higher than the corresponding browser profiles where a Mobile User Agent is used.

Tracking: A few legal entities such as Google and Facebook are able to track users across many first-parties (67% for Google and 33% for Facebook). In their maximum protection level, adblockers such as Ghostery and Adblock Plus are able to suppress effectively third-party requests to these legal entities. However legal entities such as Google that host content (e.g., libraries) in addition to advertisements are still able to track visits to more than 30 % of the first parties.

Geographical distribution: While some countries host proportionally more third-parties than first parties such as the United States and China, adblockers tend to reduce the number of third-party requests equally well irrespective of the origin country.

Temporal dynamics: In general, the privacy of users is not very sensitive to web site or blacklist optimizations that happen at shorter time scales. However, some few domains exhibit larger temporal variations of the amount of third-party requests by more than 50 percent, suggesting that these sites may perform optimizations on their domains to reduce the impact of adblockers.

7. Related work

Privacy concerns: Many works in the literature have been dedicated to the privacy concerns as a consequence of tracking and fingerprinting by third-party domains [9], [14], [22], [25], [29], [34]. Castelluccia *et al.* [11] showed that the user's interests can be inferred by the ads they receive and their whole profile can be reconstructed. This can lead to discriminations of the users according to their profile details and configurations, as shown in [12], [28].

Countermeasures: As a result, several methods have been proposed that enable targeted advertisements without compromising user privacy [7], [17]–[19], [24], [35]. Additionally, there have been a lot of attempts for the detection of tracking behavior and ad-blocking blacklist ehnancements [16], [26], [36], while some studies have proposed further mitigation techniques [20], [31].

Comparison of mitigation-techniques: Since our work focuses on the comparison of the ad-blocking tools, it is useful to present in more detail the work done so far in this field.

Balebako *et al.* [8] propose a method to measure behavioral targeting and the effect of privacy-protection techniques —e.g. disabling of third-party cookies, Do-Not-Track header, ad-blocking tools— in the limitation of the behavioral-targeted character of the advertising content,

while Krishnamurthy *et al.* [21] compare different privacy-protection techniques against the trade-offs between privacy and page quality. Leon *et al.* [23] investigate and compare the usability of some existing tools designed to limit advertising.

Pujol *et al.* [30] aim to infer the use or no use of an adblocker by examining the HTTP(S) requests sent by a browser, using the ratio of the ad requests and the downloads of filter lists as indicators. Moreover, the filtering performance of 7 different browser profiles —adblocker-configuration combinations— is compared based upon the total number of unblocked requests per browser profile. Furthermore, the ad traffic is examined and classified through the analysis of the number of requests at different time instances throught the day, the content-type of the ad requests, as well as the effect of enabling non-intrusive ads. However, the relationship between the first-party and third-party domains is not examined and no long-term data is collected regarding the tracking behavior of the third parties.

Ruffell *et al.* [33] analyze the effectiveness of various browser add-ons in mitigating and protecting users from third-party tracking networks. In total 7 browser profiles are created, each with a different combination of multiple add-ons and browser settings, and each of the profiles visits the 500 top Alexa Rank websites, while the HTTP request data is recorded with the use of Mozilla Lightbeam. The data is collected for one crawling cycle and followingly the efficiency analysis is performed based upon various graph metrics. Nevertheless, none of these browser profiles examines the difference of the effect of the third-party tracking on devices with Mobile User Agents. Moreover, the time evolution of these metrics is not examined and no legal-entity details are taken into consideration for the graph creation. Finally, as we show, Lightbeam cannot be considered a reliable source for first and third party distinction.

Mayer and Mitchell [27] implemented the tool FourthParty —an open-source platform for measuring dynamic web content— as an extension to Mozilla Firefox. Afterwards, they created several browser profiles, so as to test the efficiency of different adblocking tools under certain settings (blacklists) and crawled the 500 top Alexa websites three times using FourthParty, in order to extract the average decrease in tracking with the use of the ad-blocking tools. The results of such an analysis may, however, be subject to biases, since the URL sample set used has been vastly examined in the bibliography so far and many adblockers may have been optimized to provide a higher efficiency when tested against it. Apart from that, they do not focus on the definition of new metrics that can be used to evaluate the tracking behavior.

Englehardt and Narayanan [14] use OpenWPM [13], a web privacy measurement platform that can simulate users, collect data and record observations, e.g. response metadata, cookies and behavior of scripts. They introduce the "prominence" metric to rank third parties and describe its relationship with their rank, i.e. the absolute number of first parties they appear on (degree). They further test the effectiveness of Ghostery, as well as of the third-party-cookie-blocking option in terms of privacy and show that Ghostery's effectiveness drops for the less prominent third-party trackers. Moreover, they use stateful measurements (the browser's profile is not cleared between page visits) and investigate how many third parties are involved in cookie syncing. Finally, they investigate different fingerprinting techniques, build a detection criterion for each of them and perform measurements to show that the user's behavior is more likely to be fingerprinted on more popular sites. Their analysis though is not concentrated on the performance of the ad-blocking-software and does not compare different combinations of adblockers and settings. Additionally, the filtering-performance evaluation consists in the mere presentation of the most prominent third-party trackers when Ghostery is enabled, while no relevant graph analysis has been performed.

8. Conclusions

The emerging trend of web advertising as well as the earning potential that it has to offer have turned it into the driving force for the development of a broad spectrum of websites and businesses. However, this practice is in direct conflict with privacy matters of the end-user, since the protection of their personal information is at stake through fingerprinting and online-profiling techniques whose objective is to optimize the efficiency of the web advertisements. Adblockers aim to counter these risks by removing advertising content and preventing third-party tracking.

Our analysis provides a quantitative methodology to compare the filtering performance of different adblockers. After the inspection of multiple browser profiles — i.e. combinations of ad-blocking software and configurations — for desktop and mobile devices, we show that the usage of an adblocker can indeed increase the privacy level and restrain the leakage of information concerning the browsing behavior of the user towards third-party trackers. The most important factor that can determine the achieved privacy level is according to our experiments the selection of blacklists, whilst the activation of the *do not track* HTTP header only has a minor effect. Our findings indicate that the best-performing adblockers are Ghostery and then AdblockPlus, when both are set to a maximal-protection level, whilst the highest privacy risks exist when no adblocker or Ghostery with its default blacklist settings is used. Finally, our methodology allows for a quantitative evaluation and comparison of any new web ad-blocking software.

References

[1] Allowing acceptable ads in adblock plus. https://adblockplus.org/en/acceptable-ads.

[2] Google chrome adblock plus. https://chrome.google.com/webstore/detail/adblock-plus/cfhdojbkjhnklbpkdaibdccddilifddb/support?hl=en-GB.

[3] Juniper research. http://www.juniperresearch.com/press/press-releases/ad-blocking-to-cost-publishers-\$27bn-in-lost-reven.

[4] Mozilla statistics for adblock plus. https://addons.mozilla.org/en-US/firefox/addon/adblock-plus/statistics/?last=30.

[5] The state of do not track in firefox. https://dnt-dashboard.mozilla.org/.

[6] Number of mobile-only internet users now exceeds desktop-only in the u.s. http://www.comscore.com/Insights/Blog/Number-of-Mobile-Only-Internet-Users-Now-Exceeds-Desktop-Only-in-the-U.S, 04 2015.

[7] Elli Androulaki and Steven M Bellovin. A secure and privacy-preserving targeted ad-system. In *Financial Cryptography and Data Security*, pages 123–135. Springer, 2010.

[8] Rebecca Balebako, Pedro Leon, Richard Shay, Blase Ur, Yang Wang, and L Cranor. Measuring the effectiveness of privacy tools for limiting behavioral advertising. Web, 2012.

[9] Paul Barford, Igor Canadi, Darja Krushevskaja, Qiang Ma, and S Muthukrishnan. Adscape: Harvesting and analyzing online display ads. In *Proceedings of the 23rd international conference on World wide web*, pages 597–608. ACM, 2014.

[10] Michael Butkiewicz, Harsha V Madhyastha, and Vyas Sekar. Understanding website complexity: measurements, metrics, and implications. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 313–328. ACM, 2011.

[11] Claude Castelluccia, Mohamed-Ali Kaafar, and Minh-Dung Tran. Betrayed by your ads! In *Privacy Enhancing Technologies*, pages 1–17. Springer, 2012.

[12] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1):92–112, 2015.

[13] Steven Englehardt, Christian Eubank, Peter Zimmerman, Dillon Reisman, and Arvind Narayanan. Web privacy measurement: Scientific principles, engineering platform, and new results. *Manuscript posted at http://randomwalker. info/publications/WebPrivacyMeasurement. pdf*, 2014.

[14] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis draft: May 18, 2016.

[15] Phillipa Gill, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, Konstantina Papagiannaki, and Pablo Rodriguez. Best paper – follow the money: Understanding economics of online aggregation and advertising. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13, pages 141–148, New York, NY, USA, 2013. ACM.

[16] David Gugelmann, Markus Happe, Bernhard Ager, and Vincent Lenders. An automated approach for complementing ad blockers' blacklists. *Proceedings on Privacy Enhancing Technologies*, 2015(2):282–298, 2015.

[17] Saikat Guha, Bin Cheng, and Paul Francis. Privad: practical privacy in online advertising. In *USENIX conference on Networked systems design and implementation*, pages 169–182, 2011.

[18] Hamed Haddadi, Saikat Guha, and Paul Francis. Not all adware is badware: Towards privacy-aware advertising. In *Software Services for e-Business and e-Society*, pages 161–172. Springer, 2009.

[19] Ari Juels. Targeted advertising... and privacy too. In *Topics in Cryptology-CT-RSA 2001*, pages 408–424. Springer, 2001.

[20] Georgios Kontaxis and Monica Chew. Tracking protection in firefox for privacy and performance. *arXiv preprint arXiv:1506.04104*, 2015.

[21] Balachander Krishnamurthy, Delfina Malandrino, and Craig E Wills. Measuring privacy loss and the impact of privacy protection in web browsing. In *Proceedings of the 3rd symposium on Usable privacy and security*, pages 52–63. ACM, 2007.

[22] Balachander Krishnamurthy and Craig Wills. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th international conference on World wide web*, pages 541–550. ACM, 2009.

[23] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. Why johnny can't opt out: a usability evaluation of tools to limit online behavioral advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 589–598. ACM, 2012.

[24] D Levin, B Bhattacharjee, JR Douceur, JR Lorch, J Mickens, and T Moscibroda. Nurikabe: Private yet accountable targeted advertising. *Under submission. Contact johndo@ microsoft. com for copy*, 2009.

[25] Timothy Libert. Exposing the invisible web: An analysis of third-party http requests on 1 million websites. *International Journal of Communication*, 9:18, 2015.

[26] Justin Ma, Lawrence K Saul, Stefan Savage, and Geoffrey M Voelker. Beyond blacklists: learning to detect malicious web sites from suspicious urls. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1245–1254. ACM, 2009.

[27] Jonathan R Mayer and John C Mitchell. Third-party web tracking: Policy and technology. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 413–427. IEEE, 2012.

[28] Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. Detecting price and search discrimination on the internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*, pages 79–84. ACM, 2012.

[29] Nick Nikiforakis and Gunes Acar. Browse at your own risk. *Spectrum, IEEE*, 51(8):30–35, 2014.

[30] 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.

[31] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, pages 12–12. USENIX Association, 2012.

[32] David Singer Roy T. Fielding. Tracking preference expression (dnt). Technical report, 05 2015.

[33] Matthew Ruffell, Jin B Hong, and Dong Seong Kim. Analyzing the effectiveness of privacy related add-ons employed to thwart web based tracking. In *Dependable Computing (PRDC), 2015 IEEE 21st Pacific Rim International Symposium on*, pages 264–272. IEEE, 2015.

[34] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. Flash cookies and privacy. In *AAAI Spring Symposium: Intelligent Information Privacy Management*, volume 2010, pages 158–163, 2010.

[35] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas. Adnostic: Privacy preserving targeted advertising. In *Proceedings Network and Distributed System Symposium*, 2010.

[36] Minh Tran, Xinshu Dong, Zhenkai Liang, and Xuxian Jiang. Tracking the trackers: Fast and scalable dynamic analysis of web content for privacy violations. In *Applied Cryptography and Network Security*, pages 418–435. Springer, 2012.