Measuring the Impact and Perception of Acceptable Advertisements

Robert J. Walls Eric D. Kilmer Nathaniel Lageman Patrick D. McDaniel
Department of Computer Science and Engineering
Pennsylvania State University, University Park, PA, USA
{rjwalls, ekilmer, njl5114, mcdaniel}@cse.psu.edu

ABSTRACT

In 2011, Adblock Plus—the most widely-used ad blocking software—began to permit some advertisements as part of their *Acceptable Ads* program. Under this program, some ad networks and content providers pay to have their advertisements shown to users. Such practices have been controversial among both users and publishers. In a step towards informing the discussion about these practices, we present the first comprehensive study of the Acceptable Ads program. Specifically, we characterize which advertisements are allowed and how the whitelisting has changed since its introduction in 2011. We show that the list of filters used to whitelist acceptable advertisements has been updated on average every 1.5 days and grew from 9 filters in 2011 to over 5,900 in the Spring of 2015. More broadly, the current whitelist triggers filters on 59% of the top 5,000 websites. Our measurements also show that the program allows advertisements on 2.6 million parked domains. Lastly, we take the lessons learned from our analysis and suggest ways to improve the transparency of the whitelisting process.

Categories and Subject Descriptors

H.3.5 [**On-line Information Services**]: Web-based services; K.4.4 [**Computers and Society**]: Electronic Commerce

General Terms

Acceptable Ads; Adblock Plus; Ad Avoidance

1. INTRODUCTION

Over 144 million users employ ad blocking software [27]. Users are motivated by a desire to hide intrusive ads, increase their privacy, or protect themselves from malicious adverts [34]. Yet, some claim ad blocking threatens the Web's business model. Indeed, Google lost an estimated $887 million in revenue to blocking in Q2 2013 [26,31].

In 2011 Eyeo GmbH—the maker of the most popular ad blocker, Adblock Plus—introduced their *Acceptable Ads* program. Through this program, Adblock Plus allows some "non-intrusive" ads that satisfy a set of community-driven guidelines, such as "ads should never obscure page content." According to Eyeo, their goal is to strike a balance between the needs of users and publishers, and they emphasize *transparency* as key to the program's success [6, 22]. However, Eyeo drew strong criticism when they confirmed some companies—including Google, Microsoft, and Amazon—paid undisclosed amounts to be included in the whitelist [9,32]. Some view this arrangement as a conflict of interest; the organization that provides blocking software is in a position to indirectly profit from ads being shown.

The Acceptable Ads program impacts millions of users and billions of dollars,[1] but little is known about the whitelisting process or how it impacts users. In this paper, we provide the first comprehensive study of the Acceptable Ads program. We identify how the users experience the Web under this program by exploring the use of ad policies (called filter lists, or just whitelists). We develop tools and techniques to explore and correlate information from Internet measurements, a complete history of the program's whitelist, instrumented browser behavior, and user surveys. In this, we have focused on the following questions:

1. *What is in the whitelist and how has it changed over time?* We find that at the current revision, Rev. 988, the whitelist contains 5,936 filters and is updated every 1.5 days to add or modify 11.4 filters on average.
2. *Who benefits from the whitelist?* We find that the whitelist identifies 3,545 unique explicitly listed publisher domains (including 15 of the top 100), and that five general-purpose filter types are responsible for allowing content on 2,676,165 parked domains.
3. *How do we measure the impact of the whitelist?* We survey whitelist use in the Alexa top 5,000 most popular websites as well as 5,000 sites from the 5k to 1 million most popular. The current whitelist triggers filters on 59% of the top 5,000 websites but explicitly whitelist only a few percent of less popular sites.
4. *How do users perceive acceptable advertisements?* A survey of over 300 users showed wide dissension on many advertisements that were judged as being invasive. One area of agreement was clear: advertisements interspersed with and largely indistinguishable from web content were deemed as undesirable.

Our study is motivated by other large-scale Web and security measurement studies, including those characterizing SPAM [14, 17], affiliate programs [20], domain abuse [4,7], and malicious advertising [19,34]. We begin by detailing the operation of Adblock

[1]The Internet Advertising Bureau reported a record high of $12.4 billion in U.S. advertising revenue for Q3 2014, breaking the previous record of $12.1 billion in Q4 of 2013 [13].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
IMC'15, October 28–30, 2015, Tokyo, Japan.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3848-6/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2815675.2815703 .

```
<iframe id="ad_main" frameborder="0" scrolling="no" name="ad_main" src="http://static.adzerk.
    net/reddit/ads.html?sr=-reddit.com,loggedout&amp;bust2#http://www.reddit.com"></iframe>
```

Figure 1: Sample ad code from Reddit. This code displays an `iframe` for an Adzerk advertisement on the right side of the page. Similar code constructions are common across different sites using the same ad network. This allows Adblock Plus to use a single filter to block ads on multiple domains.

Plus and the Acceptable Ads program. We then characterize how the program works in practice. Finally, we offer suggestions for improving the transparency of the whitelisting process.

2. ADBLOCK PLUS

Adblock Plus is the most widely used browser extension with over 50 million users across all major browsers.[2] In 2014, the extension's Firefox version was downloaded 68 million times and boasted 19.2 million users daily.[3] Adblock Plus is open source and available free of charge.

Adblock Plus was created by Michael McDonald's as a fork of Henrik Aasted Sørensen's Adblock project. In January 2006, Wladimir Palant rewrote the code and released it as a separate project for Firefox. Since then, Adblock Plus has been ported to run on all major browsers: Chrome (Dec. 2010 [28], formerly AdThwart [3]), Opera (Nov. 2012 [10]), Internet Explorer (Aug. 2013 [29]), and Safari (Jan. 2014 [25]). Eyeo offers an Android version, but it is not available in the Google Play store [30].

Adblock Plus uses textually encoded *filters* to determine the content shown on a page. *Blocking filters* restrict page content, while *exception filters* override any matching blocking filters to allow the content. Filter definitions generally consist of: *(i)* a *matching expression* that specifies what content to block (or allow), e.g., the URL of an advertising network; and *(ii)* a set of filter options, e.g., the `image` option applies the filter to image requests. A detailed description of the filter syntax is included in Appendix A.

Adblock Plus users rarely write their own filters. Instead, they subscribe to regularly published text-based *filter lists*. By default, Adblock Plus subscribes users to two filter lists: the first, *EasyList*, contains tens of thousands of filters to block advertisements and covers most common ad networks. Other blocking extensions also use EasyList, including the second most popular blocker, AdBlock. The second default filterlist—which we refer to as the *Acceptable Ads whitelist*—is used to implement the Acceptable Ads program. In short, this list overrides the user's other filter lists allowing certain publishers to show advertisments. We characterize the scope and impact of the whitelist in later sections.

Users can subscribe to additional filter lists that provide functionality beyond blocking advertisements including: disabling tracking, allowing the user to "browse the web truly anonymously", blocking known malicious domains, and removing social media buttons, such as the Facebook Like button.[4] We defer analysis of these lists to future work.

2.1 Filter Matching

Broadly, individual filters match one of two types of content: *Web requests* or *page elements*. We briefly describe each below.

2.1.1 Matching Web Requests

Publishers often rely on third-party ad networks, such as Adzerk or Google Adsense, to display advertisements on their site. These networks make it possible for publishers to show ads by simply including a small snippet of code provided by the ad network. This straightforward interface also simplifies the blocking process by allowing a single filter to block ads on multiple sites.

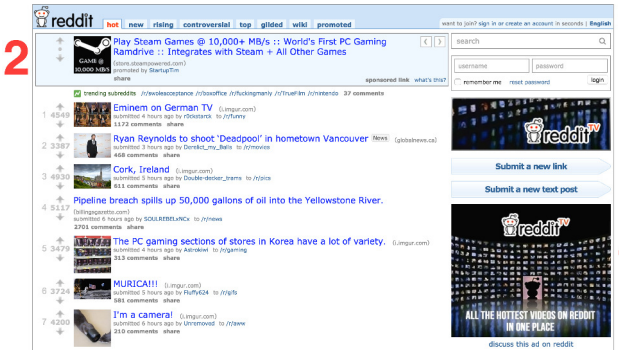


Figure 2: Acceptable ads on `Reddit.com`. Reddit is a member of the Acceptable Ads program. Consequently, Adblock Plus allows both of the ads on this page. A third-party network, Adzerk, serves the ad on the right side (labeled **1**). The sponsored link (labeled **2**) is embedded directly into the page.

For example, `reddit.com` uses the code in Figure 1 to show Adzerk advertisements. When an Adblock Plus user visits the page, their browser will make a *third-party* web request to fetch the advertisement from Adzerk. Adblock Plus will preempt this request to check if the request URL matches any filters. If the match is for a blocking filter, such as the following, Adblock Plus will cancel the request, stopping the browser from fetching the ad.

```
1 ||adzerk.net^$third-party
```

In short, the above filter will block all third-party requests to `adzerk.net` or any of its subdomains. For a more complete explanation of filter syntax see Appendix A.

If the request matches an exception filter,[5] then Adblock Plus allows it, *regardless of any blocking filter matches*. The advertisement is then shown as an image on the right of the page—denoted by the bold **1** in Figure 2.

2.1.2 Matching Page Elements

Adblock Plus uses different filter syntax for matching advertising elements embedded directly into the page. Similar to how request filters match URLs, element filters use CSS Selectors[6] to identify elements based on attributes such as the element's `class` or `id`.

The following filter blocks the "sponsored link" at the top of Reddit's front page (bold **2** in Figure 2) by matching elements with an `id` attribute of `siteTable_organic`.

```
1 ||reddit.com###siteTable_organic
```

However, Adblock Plus does not block either of the advertisements in our example. It allows these ads because Reddit is a part of the Acceptable Ads program.

[2] `https://adblockplus.org/en/about`
[3] `https://addons.mozilla.org/en-US/firefox/addon/adblock-plus/statistics/`
[4] `https://adblockplus.org/en/features`
[5] Request exception filters are denoted by the `@@` prefix.
[6] `http://www.w3.org/TR/CSS21/selector.html`

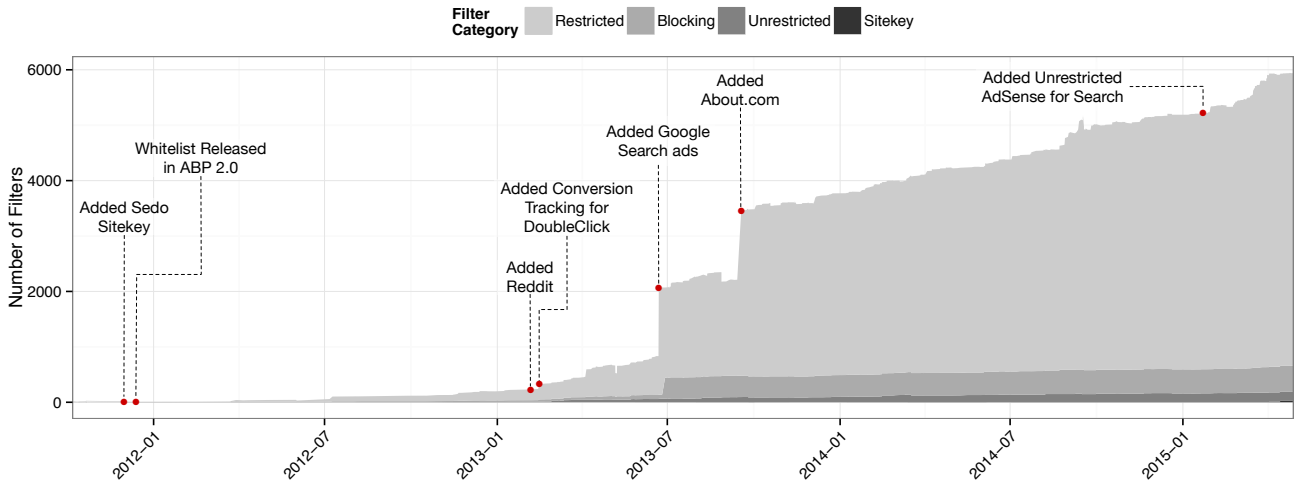


Figure 3: The growth of the Acceptable Ads whitelist. The whitelist has grown steadily since its introduction in 2011. On average, this filter list is updated every 1.5 days, adding or modifying 11.4 exception filters. The most recent version (Rev. 988 on April 28, 2015) includes 5,936 filters. Among these are exceptions for domain parking services, conversions tracking, and third-party ad networks.

3. ACCEPTABLE ADS

In May 2011, Adblock Plus's creator, Wladimir Palant, outlined the criteria for Acceptable Ads with the goal of "encourag[ing] websites to use advertising that users don't perceive as annoying."[7] These guidelines have been refined multiple times based on feedback from the community. Whitelisted sites are required to adhere to these guidelines. To paraphrase the most recent criteria [1], sites must ensure that:

1. Advertisements cannot contain animations, sounds, or "attention-grabbing" images.
2. Advertisements cannot obscure page content or obstruct reading flow, i.e., the ad cannot be placed in the middle of a block of text.
3. Advertisements must be clearly distinguished from the page content and must be labeled using the word "advertisement" or equivalent terms.
4. Banner advertisements should not force the user to scroll down to view page content.

After publishing the initial Acceptable Ads requirements, Palant and his partner, Till Faida, created Eyeo GmbH in August 2011. Their goal was to make the Adblock Plus project "more sustainable" [8]. The following month, the company began surveying users about their willingness to allow advertising in some form.[8] By the year's end released Acceptable Ads as an opt-out feature in Adblock Plus version 2.0.[9]

The program has been controversial. Eyeo drew criticism from both users and publishers when they confirmed that some companies—including Google, Microsoft, and Amazon—paid undisclosed amounts to be included in the whitelist [9,32]. Eyeo has stated that received funds are used to sustain the program. Further, they state that whitelisting is free for smaller sites and *all* participants of the program must abide by the acceptable ads criteria.

[7] https://adblockplus.org/forum/viewtopic.php?f=4&t=7551

[8] https://adblockplus.org/releases/adblock-plus-1310-released

[9] https://adblockplus.org/releases/adblock-plus-20-released

Eyeo's practices have incensed some publishers to such a degree that they have accused the company of "extortion" and "shaking down" websites [23]. Recently, this criticism has escalated to the courtroom, and Eyeo is facing lawsuits in France and Germany [11, 12,16]. In Germany, publishers filed suit against Eyeo claiming that their product is anti-competitive and threatens their ability to generate revenue. The regional court in Hamburg ruled in favor of Eyeo after a four month trial [15,33].

3.1 Filter List Maintenance

Eyeo regularly updates the whitelist. On average, the company adds or modifies 11.4 filters every 1.5 days, and has a documented process for requesting new whitelist filters to be added. The process of adding new sites to the Acceptable Ads program comprises four steps: contact, application, agreement, and inclusion.

The first step, *contact*, consists of communication between Eyeo and a perspective publisher. Either party may initiate this dialog [24]. Next, Eyeo works with the publisher to ensure their site follows the Acceptable Ads guidelines (the *application* step). Once Eyeo confirms the site's advertisements adhere to their policy, they establish a private *agreement* with the publisher. These agreements may involve a fee, but Eyeo does not currently disclose the fee structure, monetary value, or list of paying publishers. However, anecdotal sources claim the company has requested up to 30% of recovered revenue [5], and Eyeo acknowledges they have experimented with both flat and performance-based fees [24]. After reaching an agreement, Adblock Plus adds the filter to the list and solicits community feedback on the application via an online forum.[10] Forum posts made after Nov. 2014 include links to the whitelist revision and optional sample screenshots.

4. WHITELIST ANALYSIS

In the following sections, we examine the complete history of Eyeo's Acceptable Ads program. Our analysis combines whitelist changes with public disclosures and empirical observations of browser behavior. We focus on answering the following broad questions.

[10] https://adblockplus.org/forum/viewforum.php?f=12

Year	Revisions	Filters Added	Filters Removed	Domains Added	Domains Removed
2011	26	25	17	5	1
2012	47	225	30	59	4
2013	311	5152	1555	2248	73
2014	386	2179	775	859	125
2015	219	1227	495	371	207
Total	989	8808	2872	3542	410

Table 1: Yearly activity for the Acceptable Ads whitelist. This table shows the year, number of revisions, number of first-party domains, and number of changes to exception filters—modifications are counted as new filters. The data covers changes up to Apr. 28, 2015 (Rev. 988).

1. How has the whitelist changed over time? We analyze all revisions of the whitelist to quantify the number of domains and filters. Section 4.1.
2. Who benefits explicitly from the whitelist? We first characterize the whitelist filters based on their scope and target content. Then, we parse the whitelist filters to extract explicitly listed publisher domains and group these domains based on popularity and category. Section 4.2.
3. How do we measure the impact of the whitelist? We run an automated survey across the top 5,000 most popular websites and 5,000 additional sites from the 5K to 1M top sites. Section 5.
4. How do users perceive acceptable advertisements? We survey 305 users on Amazon's Mechanical Turk platform and ask them to rate 15 "acceptable" advertisements. Section 6

4.1 Whitelist History

Eyeo tracks all whitelist changes in a public Mercurial repository.[11] Using this repository, we extracted 988 versions of the whitelist dating from the start of the whitelist in Oct. 2011 to Apr. 2015. The most recent version (Rev. 988) comprises 5,936 distinct filters. The majority of these exceptions allow advertisements and other advertising functionality such as conversion tracking.

Figure 3 shows the growth of the Acceptable Ads whitelist over time. There are two large jumps visible in the figure. The first corresponds to Google's official addition to the whitelist on June 21, 2013 (Rev. 200).[12] In total, this revision added 1,262 filters for Google search advertisements on `google.com` and other variations of this URL, e.g., `google.co.uk`. The second jump was caused by filters for `ask.com`, `about.com`, and related subdomains, e.g., `cars.about.com`. These additions are discussed in Section 7.

Table 1 summarizes the yearly changes. The second full year of the whitelist (2013) saw the list grow by an order of magnitude over the previous year: Eyeo made 4,633 filter changes, bringing the total to 2,319 publisher domains compared to just 60 domains in 2012. However, as we discuss below, these numbers belie the actual scope of the whitelist.

4.2 Whitelist Scope

In order to understand who benefits from the Acceptable Ads program, we need to first understand the *scope* of a whitelist filter, i.e., the set of domains that can activate the filter. For some filters, this list of applicable domains is explicitly enumerated in the filter's definition. We call these *restricted* filters. For others—namely *unrestricted* and *sitekey* filters—the filter can apply to *any* domain. The implication here is that it is impossible to determine the actual impact of these filters using filter definitions alone. Instead, we can only empirically estimate this value through site surveys (Section 5).

[11] `https://hg.adblockplus.org/exceptionrules`

[12] `https://hg.adblockplus.org/exceptionrules/rev/8bdf815a5291`

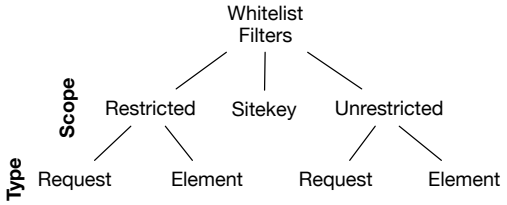


Figure 4: Hierarchy of Filter Types in the Whitelist.

Figure 4 illustrates a hierarchy of whitelist filter types and their scope. We discuss each class in detail below.

4.2.1 Restricted Exception Filters

A *restricted* exception filter explicitly defines the first-party domains for which it activates, i.e., the filter will only match when the user is browsing a page on one of these domains. The syntax for defining the domain list depends on the filter's type. For example, the following are two of the whitelist's restricted exceptions for `reddit.com`.

```
1 || reddit.com#@##ad_main
2 || @@||adzerk.net/reddit/$subdocument,
  ||     document,domain=reddit.com
```

The first filter—an element exception—instructs Adblock Plus to allow any `reddit.com` elements with the `ad_main` identifier. For element filters, restricted domains are prepended to the start. The second filter allows requests to `adzerk.net`, the ad provider for `reddit.com`. In this case, the filter is restricted via the `domain` option.

Restricted exception filters make up 89% of the whitelist and cover 3,545 fully qualified domains, including search engines (Microsoft, Yahoo, Google), commercial sales (Walmart, Amazon), content publishers (Reddit, About.com, Cracked), and ISPs (Comcast, Time Warner), amongst others. Many of the fully qualified domains appearing in the whitelist map to the same publisher. For instance, the whitelist includes over 1,044 subdomains for `about.com`, including `cars.about.com` and `food.about.com`. Additionally, there are 919 country-based domains for Google properties included in the whitelist, e.g., `google.co.uk` and `google.de`. Table 2 shows the respective count and Alexa rankings for fully qualified domains.

4.2.2 Unrestricted Exception Filters

An *unrestricted* exception filter applies to all first-party domains, i.e., these exceptions can match on any site. The whitelist uses unrestricted exceptions primarily for two purposes. First, many of these filters enable conversion tracking. Broadly, conversion tracking is used to measure if an advertisement resulted in some user action, e.g., user purchased the product after clicking on the ad. Second, unrestricted filters are used to whitelist specific ad networks.

Alexa Partition	Fully Qualified Domains
All	1,990
Top 1,000,000	1,286 (0.12%)
Top 5,000	316 (6.32%)
Top 1,000	167(16.70%)
Top 500	112(22.40%)
Top 100	33(33.00%)

Table 2: Domains explicitly included in the whitelist. The whitelist contains restricted exception filters for 3,544 fully qualified publisher domains. This corresponds to 1,990 effective second-level domains, e.g., `google.com` is the effective second-level domain of `maps.google.com`. Percentages are on the total number of domains within the particular Alexa partition.

This practice allows the ad network to show advertisements across multiple domains without having to explicitly include these domain in the whitelist. PageFair is an example of the latter.

PageFair. While PageFair is oft-quoted for their reports on the prevalence of ad blocking and its monetary costs—in fact, we cite one in the introduction—the company is also an ad network involved in the Acceptable Ads program.[13]

```
1 || @@||pagefair.net^$third-party
2 || @@||tracking.admarketplace.net^$third-
  ||     party
3 || @@||imp.admarketplace.net^$third-party
```

These unrestricted exception filters allow PageFair to show advertisements on any partnered website without needing a separate exception for each domain. The company takes a "minority share of the additional advertising revenue" that they provide publishers. PageFair also pays Eyeo to participate in the Acceptable Ads program.

Another advertising network, Influads, has similar unrestricted exceptions.

```
1 || @@||influads.com^$script,image
2 || #@##influads_block
```

The second of these filters—an element exception—is not actually limited to Influads advertisements; instead, this filter will match any element on any site as long as the element's id is `influads_block`. This is the only example of an unrestricted element filter in the whitelist, and possibly an oversight by the whitelist's authors.

In Section 5, we further explore the impact of the 156 unrestricted exception filters using a survey of popular domains.

4.2.3 Sitekey Exception Filters

A *sitekey* exception filter includes a DER-encoded, base-64 representation of an RSA public key.

```
1 || ! Text ads on Sedo parking domains
2 || @@$sitekey=MFwwDQYJK...wEAAQ,document
```

Adblock Plus allows advertisements on *any* domain that presents a valid signature signed with a current sitekey. Effectively, sitekeys delegate the task of whitelisting to the publisher.

Adblock Plus calculates the signature by signing a string containing the URI, hostname, and user-agent string of the HTTP request. Adblock Plus then compares the result with the signature returned by the server in: *(i)* the `X-Adblock-key` header of the HTTP response, and *(ii)* the `data-adblockkey` attribute of the returned page.

The first sitekey—added to the whitelist before its release—belongs to Sedo, a domain parking and hosting company based in Cologne, Germany (the same city as Eyeo). The company's founder, Tim Schumacher, is also Eyeo's chief investor and participated in the initial development of the Acceptable Ads program [24].

[13] `https://pagefair.com/about/`

Company	Whitelisted	Domains (.com)
Sedo	2011-11-30	1,060,129
ParkingCrew	2013-05-27	368,703
RookMedia	2013-07-31	949
Uniregistry	2013-09-25	1,246,359
Digimedia	2014-07-02	25
		2,676,165

Table 3: Domain statistics for the five parking services whitelisted by Adblock Plus. RookMedia was removed from the whitelist on Sept. 16, 2014 (rev. 656); all others remain active.

Currently, there are 4 sitekeys and 25 sitekey filters in the whitelist, all belonging to domain parking services. They are, in order of introduction, Sedo, ParkingCrew, Uniregistry, and Digimedia. A fifth sitekey (for Rook Media) was removed from the whitelist in Sept. 2014.

Parked domains typically exist to show advertisements (and sell domains), usually in the form of links to pages related to the domain name. Misspellings of popular sites are also frequently parked. For example, `reddit.cm` is a parked domain that advertises dating services and photos of celebrities; this domain is whitelisted under the Acceptable Ads program using a sitekey. For a more complete treatment of domain parking, see the recent work by Alrwais et al. [4].

Using the top-level domain zone file for `.com` domains, we identified approximately 3 million parked domains managed by one of the parking services listed in Table 3. Specifically, we focused on those domains whose name servers belong to one of the sitekey parking services. For example, Sedo domains use the `ns1.sedoparking.com` and `ns2.sedoparking.com` nameservers. The list of parking name servers, in part, was derived from the example sites given in Adblock Plus online forums. We used automated tools to visit each suspected domain and only recorded those that presented a sitekey signature.

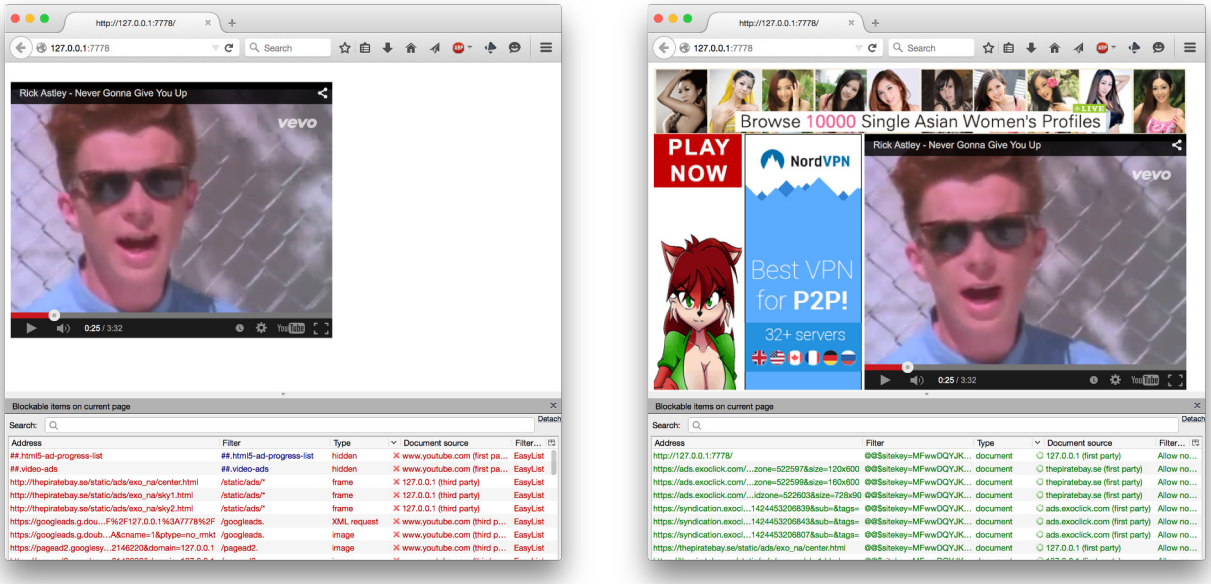
Table 3 provides a lower bound on the number of domains for each parking service. In total, we find the four active sitekeys account for at least 2,676,165 distinct whitelisted domains.

Finally, some of the above sites required special accommodations to scrape. For example, ParkingCrew domains employ countermeasures to prevent scraping, returning a `403` response if the user-agent string matches that of a tool like `curl`. Further, some domains, e.g., Uniregistry, behave differently given the presence or absence of specific cookie values. For instance, when a user visits a Uniregistry domain for the first time, the site will return a page that first generates a cookie and then redirects the user to another page with advertisements (and the sitekey signature).

Factoring Sitekeys. All current sitekeys use 512 bit RSA keys (RSA-155). Such small key sizes are well within the factoring capabilities of an individual or publisher with modest hardware resources. To demonstrate this, we constructed a cluster comprising 8 desktop computers running Ubuntu 14.04, each with an Intel Xeon E5-2630 clocked at 2.30GHz and 32 GB of memory. We used the CADO-NFS[14] implementation of the Number Field Sieve algorithm. This setup took approximately one week on average to factor each sitekey.

In Figure 5, we demonstrate how an adversarial publisher could use a factored sitekey to show intrusive or malicious advertising.

[14] `http://cado-nfs.gforge.inria.fr/`



(a) Without sitekey. **(b)** With sitekey.

Figure 5: Exploiting sitekeys. Using a cluster of desktop computers, we were able to factor one of the whitelist sitekeys in five days. The derived key allowed our test site to bypass Adblock Plus's blocking entirely.

For our proof of concept, we used Adblock Plus's default settings with both the EasyList blacklist and the Acceptable Ads whitelist enabled. The figure shows our test site before and after adding the sitekey. In short, our test site bypassed all blocking filters.

5. MEASURING FILTER BEHAVIOR

Our previous analysis—focusing on the content of the whitelist—provides a necessary but incomplete picture of the whitelist's behavior. For instance, the whitelist includes an exception filter for PageFair advertisements, but it does not list any of the domains that use PageFair.

Complicating matters, whitelist behavior also depends on a myriad of more subtle factors including filter ambiguity, unpredictable website behavior, and complex interactions between different filters and filter lists. In this Section, we expand on our previous analysis using empirical measurements of filter activation.

Methodology. We instrument Adblock Plus to measure filter activations on domains drawn from four sample groups: *(i)* the 5,000 most popular domains,[15] *(ii)* 1,000 domains randomly sampled from the rank 5K–50K popularity strata, *(iii)* 1,000 domains randomly sampled from the rank 50K–100K popularity strata, *(iv)* 1,000 domains randomly sampled from the rank 100K–1M popularity strata.

We instrumented Adblock Plus to record filter activations and used Selenium[16] to visit each domain. We surveyed only the landing page of each site. By limiting our visit to the first page, our survey produces a lower bound on the number of matching filters as some filters will not activate without user interaction. For instance, Google search ads only appear after a search.

Figure 6 shows the filter activations on the top 50 sites with at least one filter activation. We show the number and type of filter matches by the height and color of the bars, respectively. Domains explicitly included in the whitelist are shown in bold along the x-axis.

This figure illustrates a number of subtle issues when measuring filter behavior. First, 12 domains not explicitly included in the whitelist nevertheless activate whitelist filters, e.g., `youtube.com`. Second, whitelist filters activate needlessly. That is, the filter matches content that would not have otherwise been blocked by the EasyList blacklist. Third, sites may behave differently based on browser state and configuration. For example, `ask.com` will activate more filters if the user does not have specific cookies in the browser cache. Further, some sites will show different advertisements if the sites detects the presence of Adblock Plus, e.g., `imgur.com`.

Finally, not all whitelist filters are directly responsible for displaying advertisements, e.g., `gstatic` exceptions. Many common exceptions are for conversion tracking and do not visually impact the website.

These results suggests the need for more complex analysis techniques to fully characterize the whitelist's behavior. We leave such explorations for future work.

5.1 Active Filters on the Top 5K Domains

Of the Top 5,000 domains, 3,956 activated at least one Adblock Plus filter from either the EasyList blacklist or the Acceptable Ads whitelist. The remaining 1,044 domains were largely non-English (and thus out of the purview of Easylist) or required additional user interaction to trigger filters, e.g., logins, search queries, etc.

Figure 7 shows the distribution of the total and distinct whitelist filter matches per surveyed site. We include only domains with at least one whitelist filter activation (2,934 sites). `Toyota.com` saw the most filter activations with 83 total matches for 8 distinct filters, 5% of the surveyed sites activated at least 12 exception filters (non-distinct), and, on average, each site activated 2.6 distinct whitelist filters.

[15] Website popularity was based on Alexa rankings from Apr. 2015. `http://www.alexa.com/topsites`

[16] `http://www.seleniumhq.org/`

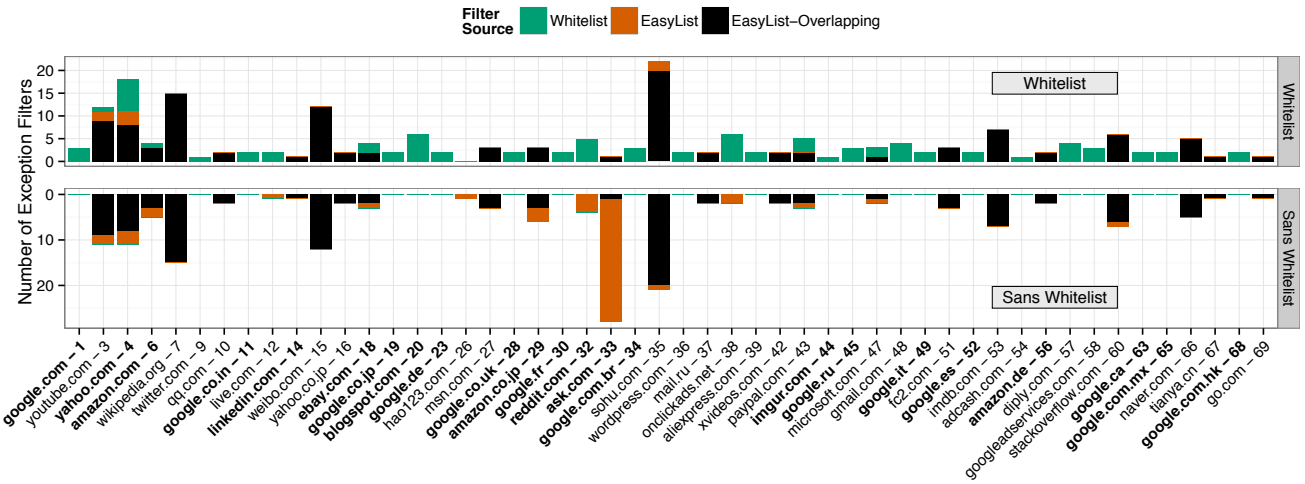


Figure 6: Filter matches with and without the whitelist enabled. The upper panel shows the filter matches when both the whitelist and EasyList are enabled, the bottom shows matches when just EasyList is enabled. We limit this figure to sites that match at least one filter from either the whitelist or EasyList (and elide `sina.com.cn` for ease of presentation). Each bar is labeled with the domain and rank. **Bold** denotes domains explicitly included in a whitelist filter definition. The height of the bar represents the number of matches and the fill specifies the filter's source. Filters that match in both configurations are shown in black.

	Filter (Truncated)	Domains	Modified	Purpose
1	`@@||stats.g.doubleclick.net^$script,image`	1,559	2013-02-21	Conversion tracking.
2	`@@||googleadservices.com^$third-party`	1,535	2013-06-21	Google search ads.
3	`@@||gstatic.com^$third-party`	1,282	2013-06-21	Google search ads.
4	`@@||googleads.g.doubleclick.net/pagead/view...`	929	2013-08-08	Experimental.
5	`@@||google.*/ads/user-lists/$image,subdoc...`	892	2013-05-31	Conversion tracking.
6	`@@||googletagmanager.com/gtm.js`	746	2013-08-08	Experimental.
7	`@@||fls.doubleclick.net^$subdocument,image`	300	2013-03-20	Conversion tracking.
8	`@@||doubleclick.net/activity*$subdocument,i...`	135	2013-08-08	Experimental.
9	`@@||google.com/adsense/search/*.js$domain=~...`	78	2015-01-22	*Google Adsense(A-filter)*
10	`@@||google.*/ads/conversion`	69	2014-11-28	Conversion tracking.
11	`@@||p.skimresources.com/px.gif?ch=1&rn=`	53	2013-11-07	Text to affiliate links.
12	`||p.skimresources.com/px.gif?ch=2&rn=`	53	2013-11-07	Blocking, text to affiliate links.
13	`@@||r.skimresources.com/api/?$script`	48	2013-08-27	Text to affiliate links.
14	`@@||s.skimresources.com/js/*.skimlinks.js^$...`	48	2013-08-14	Text to affiliate links.
15	`@@||t.skimresources.com/api/track.php?$script`	47	2013-08-27	Text to affiliate links.
16	`@@||pagefair.net^$third-party`	31	2014-01-30	PageFair ads.
17	`#@##influads_block`	30	2012-11-08	Influads ads.
18	`||viglink.com/images/pixel.gif?ch=2$third-party`	25	2014-06-02	Blocking, text to affiliate links.
19	`@@||doubleclick.net/json`	22	2013-08-08	Experimental.
20	`@@||google.com/gen_204`	20	2013-08-08	Experimental.

Table 4: Most common whitelist filters in the survey. This table displays the 20 most common exception filters from our survey of Alexa's top 5,000 websites. Number 9 on this list was added without community vetting; it allows Google's AdSense for search on nearly *all* domains.

Table 4 shows the 20 most common whitelist exception filters in the Top 5,000 group. As expected, all of these filters are unrestricted. As we discussed previously, unrestricted filters can trigger on any site.

The most activated filter, `@@||stats.g.doubleclick.net^$script,image`, triggered on 1,559 domains (31.2%). This filter is used to allows conversion tracking. The second-most popular filter, `@@||googleadservices.com^$third-party`, was observed on 1,535 domains, and allows advertisements from Google's AdSense network. The third-most popular filter, `@@||gstatic.com^$third-party`, occurred on 1,282 domains. This filter does not appear to contribute to the visibility of advertisements. Instead, the Google-owned `gstatic.com` serves fonts, scripts, images, and other resources to sites to increase browsing performance. The necessity of the `gstatic.com` filter is unclear to us, given that EasyList does not currently contain any filters that would block the observed `gstatic.com` requests.

We observed one unrestricted element exception filter, `#@##influads_block`, which activated on 30 different domains. As discussed in Section 4.2.2, this filter prevents the blocking of content contained within any element with an id of `influads_block`.

5.2 Filter Activations Across Categories

Figure 8 shows the number of filters triggered by domains in each group. The top portion shows categorical filter activation frequencies while the lower portion shows filter activation frequency

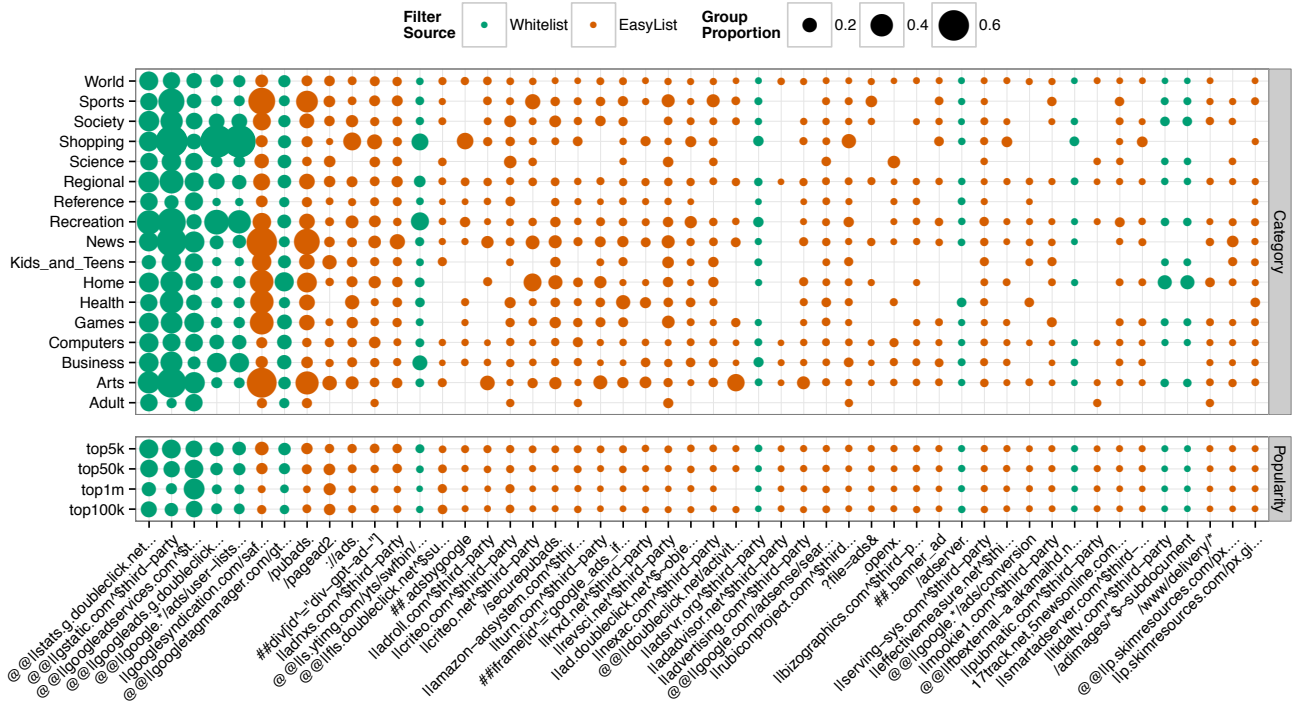


Figure 8: Filter matches per group ranking. Each row represents 1,000 domains within the respective Alexa ranking group. For each group we plot the frequency each filter is triggered by domains in that group.

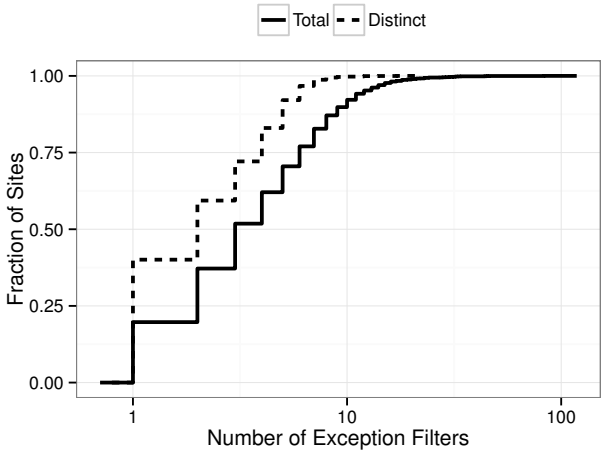


Figure 7: ECDF of filter matches per surveyed domain. We only include domains with at least one match. A single filter may match multiple elements (or requests) on a single domain; the solid line represents the total number of matches while the dashed line is the number of unique matching filters.

for the most popular sites taken from Alexa. We chose the top 50 most frequently activated filters and found that the 5 most activated filters out of both the EasyList and Whitelist were all filters from the whitelist. These filters also related to Google. This implies that these filters may be more broad than necessary.

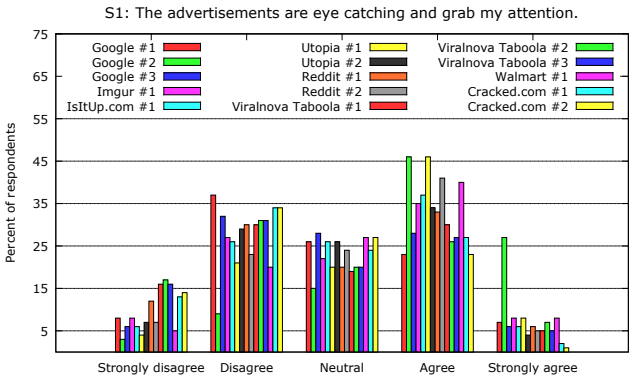
We find that the whitelist filters are skewed more towards shopping websites, which we can attribute to the filter's purpose.

We also find that 4 of these filters were triggered more often by the top 5,000 websites than by any of the other groups. There is one filter that was used most often by the group of domains consisting from Alexa rank 100,000 to 1,000,000. This particular filter is used for conversion tracking.

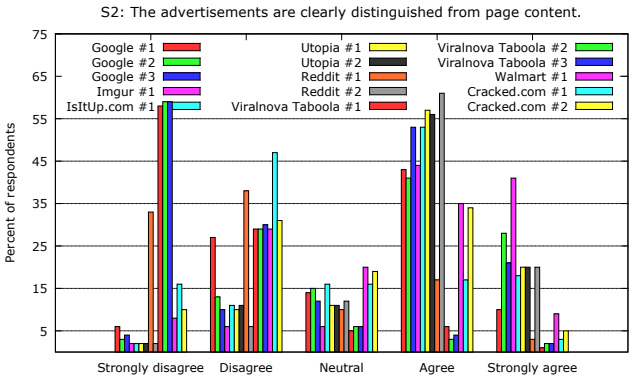
6. USER PERCEPTION

Underlying the Acceptable Ads program is the goal that both publishers and users find the whitelisted ads to be acceptable. There have been many studies of how users perceive Web advertisements (e.g., [21]), and the degree to which they intrude on the user experience (e.g., [18]). In this section, we build on these efforts to survey user perception of advertisements on popular websites based on the criteria stated in Eyeo's Acceptable Ads guidelines [1]. We used Amazon's Mechanical Turk [2] to solicit participation, limiting our pool to workers with at least 5,000 approved submissions and at least 98% approval rate. Each of the 305 respondents were paid 1$US and completed the 72 question survey in about 10 minutes. 50% of the users used some form of ad blocking software before, with 61% using the Google Chrome browser, 28% using FireFox, 9% Safari, and 1% each for Opera and Internet Explorer. The results of the survey are discussed below and shown in Figure 9.

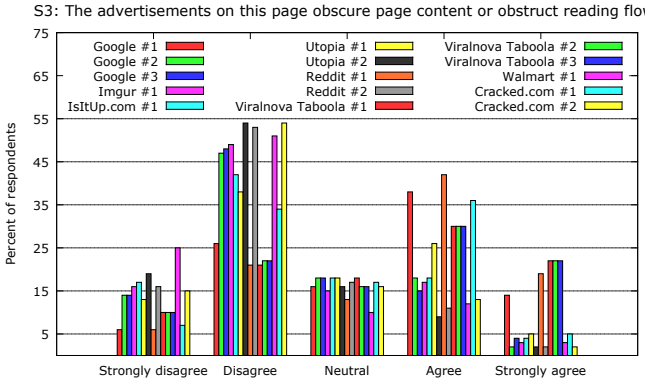
The online survey showed eight different sites, each containing one or more advertisements allowed by Adblock Plus. The eight sites were selected based on their popularity and diversity of ad placement. Specifically, we choose a search engine (Google), an image hosting service (Imgur), an online retailer (Walmart), a Web service (IsItUp.com), an online game forum (Utopia-game.com), a humor website (Cracked.com), a viral content curator (ViralNova), and a user-submitted content site (Reddit).



(a) Attention grabbing advertisements



(b) Advertisement distinguished from content



(c) Advertisement obscuring content

Search Engine Marketing Advertisements			
	Attention	Distinguished	Obscuring
μ	0.217	0.597	-0.260
$VAR(X)$	0.304	0.095	0.219
Banner Advertisements			
	Attention	Distinguished	Obscuring
μ	0.152	0.755	-0.613
$VAR(X)$	0.015	0.131	0.042
Content Advertisements			
	Attention	Distinguished	Obscuring
μ	-0.247	-0.935	0.125
$VAR(X)$	0.009	0.305	0.178

(d) Mean and variance of the survey responses. Calculated by assigning integer values [-2, 2] to the Likert scale, e.g., strongly disagree was given -2.

Figure 9: User perception survey results.

For each whitelisted advertisement, we asked the participants to rate their level of agreement with statements relating to acceptability on a Likert scale, e.g., "Strongly Disagree," "Disagree," "Neutral," "Agree," or "Strongly Agree." Each statement is a direct transcription from the Adblock Plus policy (with minor adjustments for clarity) and states characteristics that must or must not be true for an advertisement to be "acceptable". Ideally, Adblock Plus should only allow advertisements that meet these criteria.

Statement 1: **The advertisements are eye catching and grab my attention.** - This first statement measures the level of distraction that the advertisement has on the user. There were two types of advertisements that many users found to be attention grabbing or distracting. Figure 10 shows two examples of the advertisements deemed most attention getting, (10*a*, Google Ad #2, with 73% agreeing or strongly agreeing) the image-based sales advertisements displayed with search results on engines and, (10*b*, Utopia Ad #2, 45%) the advertising bar next to navigation buttons on many websites.

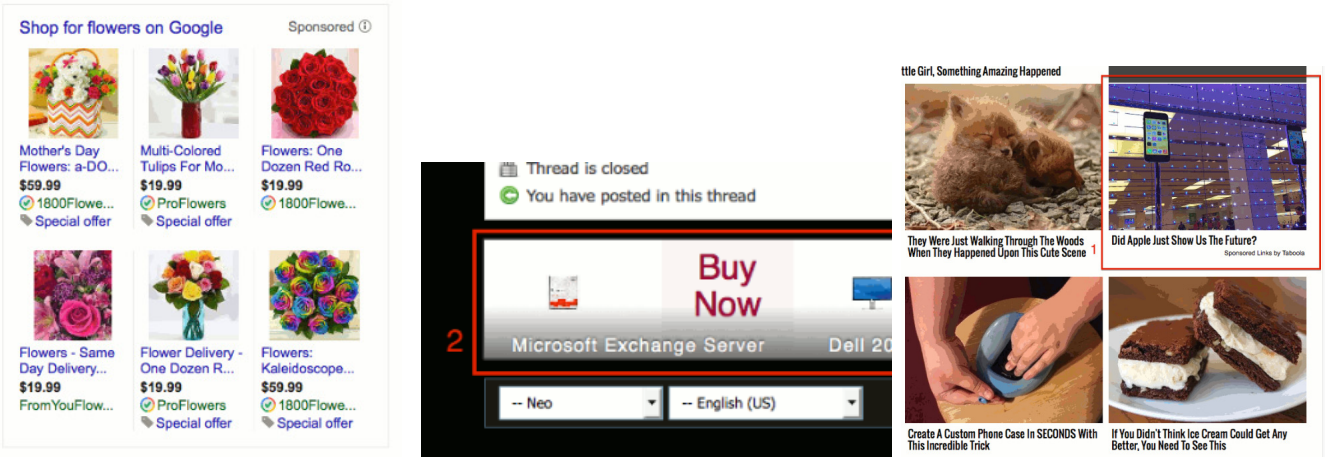
Statement 2: **The advertisements are clearly distinguished from page content.** - This statement is designed to measure how well the user can distinguish the advertisements from page content. Illustrated in Figure 10, the mixed content/advertising grid images used in curator websites such as ViralNova appears to intentionally blur the lines between advertisement and content. Almost 90% of users viewing all grid-layout ads stated that they were not distinguished from the content. Hence the allowance by Adblock Plus seems to be in conflict with the program's stated policies.

Statement 3: **The advertisements on this page obscure page content or obstruct reading flow.** - This last statement measures the degree to which users feel the ads are intrusive on content use. While the mixed content/advertising grids seem to inhibit some content use, other ad strategies were viewed as more invasive. In particular, a little more than a third of users viewed sidebar advertisements (e.g., Reddit #1, not shown), first search results (Google #1), and top bar advertisements (Cracked.com #1) as inhibiting.

Summary. While the above results are instructive, one must be careful not to over-read the meaning of one survey. A summary of the survey results in Figure 9(d) shows that there is broad dissension amongst the participants about what was acceptable, confusing, or inhibiting. However, this reenforces our experience; we have observed that each person views advertisements differently—often vastly so. Therefore, any single policy of whitelisting is unlikely to serve the needs of a large and diverse user community well. Developing deeper and larger studies of user desires is needed to develop a better understanding of user preferences and ultimately a more precise and flexible advertisement blocking policy.

7. UNDOCUMENTED FILTERS

Over the course of our analysis outlined in the previous sections, we discovered several instances where filters were added without



(a) Google Ad#2 **(b)** Utopia Ad#2 **(c)** ViralNova Ad#1

Figure 10: Survey advertisement examples.

community vetting or public disclosure. For completeness, we discuss two representative instances here.

Google's introduction. Google was officially added to the whitelist on June 21, 2013 (Rev. 200).[17] Prior to this, Eyeo allowed some publishers (but not all) to show Google AdSense for search ads on their own search pages. Of particular note among these exceptions are the two filters added for `golem.de` (Rev. 67, Dec. 2012).[18]

```
1 ||@@||google.com/ads/search/module/ads/*/
  ||    search.js$domain=suche.golem.de|www.
  ||    google.com
2 ||www.google.com#@##adBlock
```

These exception filters are structured differently than previous AdSense filters. Namely, the first exception specifies both `golem.de` and `www.google.com` in the domain list. This is unusual as the addition of Google in the first-party `domain` option is not necessary for showing search advertisements on `golem.de`. Rather, this option makes the filter active on `www.google.com`. The second filter is even more unusual as it unblocks the `adBlock` element on `www.google.com`, i.e., the filter does not make any reference to `golem.de`. None of the previous AdSense exceptions included analogous filters.

Roughly two weeks later, Eyeo modified these filters[19] to match other restricted AdSense exceptions. In particular, they removed `www.google.com` from the domain list in the first filter and deleted the second filter entirely.

```
1 ||@@||google.com/ads/search/module/ads/*/
  ||    search.js$domain=suche.golem.de
```

The forum post for the `golem.de` filters does not provide any rationale for the initial dissimilarities between these other AdSense exceptions, nor does it indicate the purpose of the later changes (or even mention such changes were made). However, during the two weeks the original filters were active, Google could conceivably have used the filters (especially the element filter) to measure the impact of whitelisting by adding an element with id `adBlock`. This element would be active via the whitelist when normal ads were not.

[17] `hg.adblockplus.org/exceptionrules/rev/8bdf815a5291`
[18] `hg.adblockplus.org/exceptionrules/rev/feb913d65a21`
[19] `hg.adblockplus.org/exceptionrules/rev/9c5f8032d88b`

```
||!A6
||@@||Ask.com^$elemhide
||@@||us.ask.com^$elemhide
||@@||uk.ask.com^$elemhide

||!A29
||@@||google.com/adsense/search/ads.
||    js$domain=search.comcast.net
||@@||google.com/ads/search/module/ads/*/
||    search.js$script,domain=search.
||    comcast.net
||@@||google.com/afs/$script,subdocument,
||    document,domain=search.comcast.net

||!A46
||@@||kayak.com.au^$elemhide
||@@||kayak.com.br^$elemhide
||@@||checkfelix.com^$elemhide

||!A50
||@@||twcc.com^$elemhide
||@@||google.com/adsense/search/ads.
||    js$domain=twcc.com
||@@||google.com/ads/search/module/ads/*/
||    search.js$script,domain=twcc.com
```

Figure 11: Subset of A-filter groups. In contrast to typical filters, A-filter groups do not contain a comment with a link to the forum. Insofar as we can determine, none of these filters were publicly disclosed by Eyeo. In total, we uncovered 59 A-filter groups.

A-List filters. There are 61 instances of Eyeo adding whitelist filters without community vetting—many of which are exceptions for large companies. We refer to these as *A-filters* because of the nondescript comments preceding each group in the whitelist, e.g. `!A1`. Figure 11 shows four example sets. Insofar as we can determine, none of these filters were publicly disclosed. In other words, none of the A-filter groups appear in Eyeo's notification forum. Nearly all A-filter additions use the same repository commit message, "Updated

whitelists."[20] In contrast, for other filters, Eyeo includes a link to the appropriate forum post in both the commit message and with a comment in the whitelist itself.

Eyeo first added A-filters in Rev. 287[21], with sets A1 and A2. Since the addition of these first filters, Eyeo has continued to add A-filters to the whitelist. The most recent one being A61 in Rev. 955.[22] Over time they have also removed 5 A-filter groups; one of these was re-added as a different A-filter, i.e. A7 as A28 in Rev. 625.[23] Among the filters added are exceptions for large companies such as Walmart and Time Warner Cable. Domains within an A-filter group appear to be closely related, possibly under the same parent organization, e.g., media company IAC owns both `about.com` and `ask.com`.[24]

In Rev. 789[25] Eyeo added A59. This filter group includes an *unrestricted* filter for Google's AdSense for search service. Specifically, this filter allows nearly *all* domains to show Google search ads.[26] As we see in Table 4, this AdSense exception was the ninth most popular filter in our survey, having been observed on 78 distinct domains.

8. INCREASING TRANSPARENCY

Acceptable Ads can benefit both users and publishers. Users are no longer subjected to annoying and intrusive advertising, while publishers—who use advertising as the primary means to support their work—can avoid erecting paywalls or reducing content. However, our analysis suggests several program areas in need of improvement. Namely, the current process lacks full transparency. Such transparency is essential for this laudable tradeoff between user desires and economic needs to become reality.

We largely agree with the Acceptable Ads criteria, but recommend that Eyeo take additional steps to reassure users that the company is acting in good faith. We recommend the following for improving the whitelisting process and policies.

Disclose financial entanglements. Other than a few isolated examples, users must guess at which filters and domains originate from a paid agreement. Eyeo claims that roughly 90% of those companies added to the list did not have to pay, however full disclosure of these statistics would strengthen public trust in the Acceptable Ads initiative.

Eyeo should clearly identify why certain sites have to pay and others do not. Eyeo states that whitelisting is free for small and medium-sized sites, but they do not say how they make this determination or if this is the only criterion. For instance, Alexa ranks `reddit.com` higher than Microsoft's Bing search; however, the former has stated they do not pay for their whitelist exceptions, while the latter does. Further, do all whitelisted ad networks pay Eyeo?

Public disclosure of whitelist agreements is especially important when one considers that a single company may own multiple domains. For instance, both `ask.com` and `about.com` are in the whitelist, and both are owned by the same parent company, IAC.[27]

Document *all* whitelist modifications. Currently, Eyeo notifies their users of whitelist changes through their online forum. However, as we discussed in previous sections, Eyeo does not give all filters the chance to be publicly vetted, e.g. `kayak.com`. It is unclear why some proposals are included in the forums and others are not.

The inclusion of forum links and a short description for groups of filters contained in the whitelist allow users to easily find relevant information. However, filter groups with nondescript titles, e.g. `A3`, are opaque.

Avoid overly general filters. A single unrestricted (or sitekey) filter in the whitelist may apply to any number of domains, making it impossible for a user to determine the filter's full scope. Overly general filters obscure whitelist understanding.

Identify whitelisted advertisements. In Google Chrome, the Adblock Plus extension logo includes a number indicating the number of blocked elements on the page. Moreover, a greyed-out logo appears when the user manually disables Adblock Plus on a page. However, there is no visible indication as to the number of whitelisted elements. Conversely, the Firefox version includes a "Blockable Items" toolbar that displays a list of page objects along with any triggered filters and the list from where the filter originates: EasyList, whitelist, etc. All Adblock Plus versions would benefit from such functionality by allowing interested users to determine which page elements were blocked and which ones were allowed (and why).

Practice good whitelist hygiene. The whitelist contains redundant, obsolete, and malformed filters. In addition to 35 duplicate filters, we observed at least 8 malformed exception filters, all of which appear to have been erroneously truncated (in Rev. 326) at a max length of 4095 characters. Similarly, AdSense for search exceptions are no longer required for individual domains. A process for retiring needed filters should be created and documented; and superfluous, malformed, and outdated filters should be purged.

9. CONCLUSIONS

This paper provides an analysis of Acceptable Ads: a program with the goal of changing Internet advertising for the better by allowing non-intrusive ads. Our study shows that the program has grown aggressively over the last four years. There is a diversity of sites being whitelisted, but a seeming (and understandable) concentration of whitelisting on popular sites. Yet, the complexity of the lists and their interaction with other web systems makes a full understanding difficult.

To us, the way forward seems clear. Public disclosures of financial relationships and more process transparency will foster trust and allow users to make informed choices about how they use blocking and ultimately what sites they visit. This community voice is essential because an open discussion is perhaps the only means of reaching a universally acceptable resolution in the blocking/whitelisting debate.

10. ACKNOWLEDGMENTS

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

[20] Rev. 304 used the message "Added new whitelists."
[21] `hg.adblockplus.org/exceptionrules/rev/b60864b2ab92`
[22] `hg.adblockplus.org/exceptionrules/rev/5a666fbe3612`
[23] `hg.adblockplus.org/exceptionrules/rev/8a9d97588927`
[24] `http://iac.com/brands`
[25] `hg.adblockplus.org/exceptionrules/rev/c182d9dc5600`
[26] The filter defines a set of 43 domains for which the AdSense filter will **not** apply.
[27] `http://iac.com/brands`

11. REFERENCES

[1] Allowing acceptable ads in adblock plus. https://adblockplus.org/en/acceptable-ads. Accessed Sept. 2014.

[2] Amazon mechanical turk. https://www.mturk.com/mturk/welcome. Accessed April, 2015.

[3] Is adblock plus for chrome (formerly adthwart) superior... https://adblockplus.org/forum/viewtopic.php?f=10&t=6509, Dec. 2010.

[4] S. Alrwais, K. Yuan, E. Alowaisheq, Z. Li, and X. Wang. Understanding the dark side of domain parking. In *Proceedings of the 23rd USENIX Security Symposium*, 2014.

[5] B. Brinkmann. Wie adblock plus mit werbung geld verdient. http://www.sueddeutsche.de/digital/browsererweiterung-in-der-kritik-wie-adblock-plus-mit-werbung-geld-verdient-1.1715051, July 2013. Accessed Oct. 2014.

[6] H. Cett. Adblock founder faida calls for transparency in ad industry. http://www.gomonews.com/adblock-founder-faida-calls-for-transparency-in-ad-instustry/, Dec. 2013. Accessed Feb. 2015.

[7] N. Chachra, D. McCoy, S. Savage, and G. Voelker. Empirically characterizing domain abuse and the revenue impact of blacklisting. In *Proceedings of the Workshop on the Economics of Information Security (WEIS)*, 2014.

[8] N. Cohen. An ad blocker opens the gate, ever so slightly. http://www.nytimes.com/2012/01/02/business/media/adblock-plus-allowing-some-online-advertisements.html?pagewanted=all&r=1&_r=0, Jan. 2012.

[9] R. Cookson. Google, microsoft and amazon pay to get around ad blocking tool. http://www.ft.com/intl/cms/s/0/80a8ce54-a61d-11e4-9bd3-00144feab7de.html#axzz3QbaQtkaZ, Feb. 2015. Accessed Feb. 2015.

[10] F. Dahlke. Adblock plus 1.3.3 for opera released. https://adblockplus.org/releases/adblock-plus-133-for-opera-released, Nov. 2012.

[11] M. Ermert. First court hearing of broadcasters and adblock plus. http://policyreview.info/articles/news/first-court-hearing-broadcasters-and-adblock-plus/347, Dec. 2014. Accessed Dec. 2014.

[12] S. Gandhe. Adblock plus sued by german marketing companies. http://www.neowin.net/news/adblock-plus-sued-by-german-marketing-companies, July 2014. Accessed Jan. 2015.

[13] IAB. Q3 2014 internet advertising revenues hit $12.4 billion, making it the highest quarter on record. http://www.iab.net/about_the_iab/recent_press_releases/press_release_archive/press_release/pr-121814, Dec. 2014.

[14] C. Kanich, N. Weaver, D. McCoy, T. Halvorson, C. Kreibich, K. Levchenko, V. Paxson, G. M. Voelker, and S. Savage. Show me the money: Characterizing spam-advertised revenue. In *USENIX Security Symposium*, pages 15–15, 2011.

[15] L. Kelion. Adblock plus defeats german publishers in court. http://www.bbc.com/news/technology-32409952, April 2015. Accessed April, 2015.

[16] L. Leong. Adblock plus responds to german lawsuit. http://news.en.softonic.com/adblock-plus-responds-to-german-lawsuit, July 2014. Accessed Jan. 2015.

[17] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Félegyházi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, et al. Click trajectories: End-to-end analysis of the spam value chain. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 431–446. IEEE, 2011.

[18] H. Li, S. M. Edwards, and J.-H. Lee. Measuring the intrusiveness of advertisements: Scale development and validation. *Journal of advertising*, 31(2):37–47, 2002.

[19] Z. Li, K. Zhang, Y. Xie, F. Yu, and X. Wang. Knowing your enemy: understanding and detecting malicious web advertising. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 674–686. ACM, 2012.

[20] D. McCoy, A. Pitsillidis, G. Jordan, N. Weaver, C. Kreibich, B. Krebs, G. M. Voelker, S. Savage, and K. Levchenko. Pharmaleaks: Understanding the business of online pharmaceutical affiliate programs. In *Proceedings of the 21st USENIX conference on Security symposium*, pages 1–1. USENIX Association, 2012.

[21] S. McCoy, A. Everard, P. Polak, and D. F. Galletta. The effects of online advertising. *Commun. ACM*, 50(3):84–88, Mar. 2007.

[22] M. McHugh. Media mafiosos: Is adblock plus shaking down websites for cash to let ads through? http://www.digitaltrends.com/web/adblock-plus-accused-of-shaking-down-websites/, Feb. 2013.

[23] M. McHugh. Media mafiosos: is adblock plus shaking down websites for cash to let ads through? http://www.digitaltrends.com/web/adblock-plus-accused-of-shaking-down-websites/, Feb. 2013. Accessed Jan. 2015.

[24] R. L. Mitchell. The business of ad blocking: A q&a with adblock plus lead investor tim schumacher. http://www.computerworld.com/article/2475721/desktop-apps/the-business-of-ad-blocking--a-q-a-with-adblock-plus-lead-investor-tim-schumacher.html, Jan. 2014.

[25] S. Noack. Adblock plus for safari beta released. https://adblockplus.org/releases/adblock-plus-for-safari-beta-released, Jan. 2014.

[26] PageFair. Acceptable ads soothe google pain. http://blog.pagefair.com/2013/acceptable-ads-soothe-google-pain/?cmp=17, Aug. 2013. Accessed Oct. 2014.

[27] PageFair and Adobe. Adblocking goes mainstream 2014 report. http://downloads.pagefair.com/reports/adblocking_goes_mainstream_2014_report.pdf, Sept. 2014. Accessed Oct. 2014.

[28] W. Palant. Adblock plus for google chrome (beta) 1.0.22 released. https://adblockplus.org/releases/adblock-plus-for-google-chrome-beta-1022-released, Dec. 2010.

[29] W. Palant. Adblock plus 1.0 for internet explorer released. https://adblockplus.org/releases/adblock-plus-10-for-internet-explorer-release, Aug. 2013.

[30] W. Palant. Adblock plus for android removed from google play store. https://adblockplus.org/blog/adblock-plus-for-android-removed-from-google-play-store, Mar. 2013.

[31] L. Stampler. Google saved an estimated $887 million by paying adblock plus to show its ads. http://www.businessinsider.com/google-saved-an-estimated-887-million-by-paying-adblock-plus-to-show-its-ads-2013-8, Oct. 2013. Accessed Oct. 2014.

[32] J. Steinschaden. Google ist geldgeber von adblock plus. http://www.horizont.at/home/detail/google-ist-geldgeber-von-adblock-plus.html, July 2013. Accessed Oct. 2014.

[33] B. Williams. Restating the obvious: adblocking declared legal. https://adblockplus.org/blog/restating-the-obvious-adblocking-declared-legal, April 2015.

[34] A. Zarras, A. Kapravelos, G. Stringhini, T. Holz, C. Kruegel, and G. Vigna. The dark alleys of madison avenue: Understanding malicious advertisements. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 373–380. ACM, 2014.

APPENDIX

A. FILTER SYNTAX

A filter is structured like a regular expression with additional modifiers and options for changing action and scope. A blocking filter prevents web content from loading or showing, and an exception filter overrides blocking filters to allow content. Options and modifiers adjust the scope of filters to affect only specific content or domains.

Adblock Plus provides documentation for the syntax of filters on its website.[28] Although Adblock Plus uses regular expressions to parse filters, we provide a BNF grammar in Figure 12, which can be used to understand how filters are created.

A.1 Blocking Filters

The simplest blocking filter is ⟨*request*⟩ with just ⟨*request-match*⟩. The ⟨*request-match*⟩ is made primarily of a regular expression that defines the path to be blocked, ⟨*regex-URL-path*⟩. The filter expression

```
1 || http://example.com/ads/advert777.gif
```

will block the specified `gif` advertisement loaded from the given address. The use of wildcards can increase the effectiveness of a single filter. For convenience, each filter regular expression has an implicit wildcard operator at the beginning and end of the filter expression. For instance, the filter expression, `/ad-frame/`, will affect any domain and will block all page elements that are requested from the `ad-frame` subdirectory.

A '`|`' can be added at the beginning or end of ⟨*regex-URL-path*⟩ to override the implicit wildcard operators. Additionally, the '`||`' string appears before a domain to allow subdomains and multiple protocols, such as `http://` and `https://`. The following filter

```
1 || ||example.com/ad.jpg|
```

matches both `http://good.example.com/ad.jpg` and `https://example.com/ad.jpg`, but not `https://example.com/ad.jpg.exe`.

Moreover, the '`^`' character matches any and all separator characters within a URL, and may be placed at the beginning or end of ⟨*regex-URL-path*⟩. Separator characters appear as "anything but a letter, a digit, or one of the following: `_ - . %`."[29] For example, the filter

[28] https://adblockplus.org/en/filters

[29] https://adblockplus.org/en/filters#separators

⟨*filter*⟩ ::= ⟨*blocking*⟩
| ⟨*exception*⟩

⟨*blocking*⟩ ::= ⟨*request*⟩
| ⟨*element-hide*⟩

⟨*exception*⟩ ::= '`@@`'⟨*request*⟩
| ⟨*element-exception*⟩
| ⟨*sitekey*⟩

⟨*sitekey*⟩ ::= '`@@$sitekey=`'⟨*pub-key*⟩∗('|'⟨*pub-key*⟩)#(['∼']⟨*option*⟩)

⟨*request*⟩ ::= ⟨*request-match*⟩['`$`'1#(⟨*no-neg-opt*⟩ | ['∼']⟨*option*⟩)]

⟨*element-hide*⟩ ::= #(['∼']⟨*page-domain*⟩)'`##`'⟨*elem-selector*⟩

⟨*elemhide-except*⟩ ::= #(['∼']⟨*page-domain*⟩)'`#@#`'⟨*elem-selector*⟩

⟨*request-match*⟩ ::= ['||' | '|' | '^']⟨*regex-URL-path*⟩['|' | '^']

⟨*option*⟩ ::= '`script`'
| '`image`'
| '`stylesheet`'
| '`object`'
| '`xmlhttprequest`'
| '`object-subrequest`'
| '`subdocument`'
| '`document`'
| '`elemhide`'
| '`other`'
| '`third-party`'
| '`collapse`'

⟨*no-neg-opt*⟩ ::= '`domain=`'['∼']⟨*page-domain*⟩
∗('|'['∼']⟨*page-domain*⟩)
| '`sitekey=`'⟨*pub-key*⟩ ∗('|'⟨*pub-key*⟩)
| '`match-case`'
| '`donottrack`'

Figure 12: Adblock Plus Filter Syntax BNF Grammar.

```
1 || ^www.google.com^
```

will match `http://www.google.com/#q=foo`, where the separator characters are bolded, but this expression will not match `http://scholar.google.com` due to the replacement of `www` with `scholar`.

Request Filters and Options. While a request filter can consist of only a regular expression with modifiers, the addition of options can tune the scope of a request filter. The '`$`' character is placed after ⟨*request-match*⟩ to indicate the start of an option list. An option list consists of 1 or more ⟨*option*⟩s or ⟨*no-neg-opt*⟩s separated by a comma. A '∼' character is the negation operator, and it can be coupled with an ⟨*option*⟩ to apply the inverse. However, there are some options that cannot be negated: ⟨*no-neg-opt*⟩. Refer to Section A.4 for more information on each filter option.

Element Hiding. Elements refer to text or page areas embedded into the web page itself through the source code. An element hiding filter is required to hide and block these elements. The element hiding filter requires at least '`##`' followed by ⟨*elem-selector*⟩ to match the target page elements.

The ⟨*elem-selector*⟩ is most commonly a CSS element selector,[30] however this is not always the case.[31] Selectors can name the element explicitly, as `##.ButtonAd`, or just specify the `id` attribute to hide, `###sideads`, which is a more general match. Like ⟨*regex-URL-path*⟩, selectors may contain regular expressions.

In the current implementation of Adblock Plus, restricting element hiding filters to a certain domain requires the use of full domain names.[32] Multiple domains are separated by a comma and apply to the same selector:

```
1 || mnn.com,streamtuner.me###adv
```

A negation operator, '∼', may also be applied to domain names to create an exception for element hiding filter.

A.2 Exception Filters

An exception filter allows the user to view page content that would otherwise be blocked by one or more blocking filters. Exception filter syntax is very similar to that of blocking filters. An ⟨*exception*⟩ can be either a ⟨*request*⟩ or an ⟨*element-exception*⟩, like blocking filters. The main differences are found at the beginning of request filters and in the middle of element filters.

Request Filters. An exception ⟨*request*⟩ filter takes the form of '`@@`'⟨*request*⟩, where '`@@`' must prefix all ⟨*request*⟩s. Refer to Section A.1 and the formation of ⟨*request*⟩ for further details. If only ⟨*request*⟩ appears in the filter (no options), then the string acts like a regular expression and matches all requests containing that string.

The following example appears in the whitelist allowing DoubleClick advertisements on `references.net`:

```
1 || @@||g.doubleclick.net/pagead/$subdocument
  ||    ,domain=references.net
```

The '`domain`' option specifies the domain for which this filter applies, and the '`subdocument`' option indicates that the DoubleClick advertisement will be embedded in the '`references.net`' page.

Element Filters. Element exception filters allow page elements to be viewed. An ⟨*elemhide-except*⟩ is very similar in form to ⟨*elem-hide*⟩. However, instead of the string '`##`' for element hiding, an ⟨*elemhide-except*⟩ uses '`#@#`' to denote that it is an exception. Refer to Section A.1 for more information on element hiding.

The next example also appears in the whitelist and is an element hide exception that allows `references.net` to show the above DoubleClick ad example on the page.

```
1 || references.net#@#.adunit
```

Without this filter, a request to DoubleClick is established and the ad is loaded, but an element hiding filter hides it from view. The structure of this filter includes ⟨*page-domain*⟩ first, and the element's `class` name, '`.adunit`', second.

A.3 Sitekey Filters

Sitekey filters primarily occur as exception filters with '`document`' as ⟨*option*⟩, to allow *all* advertisements. These filters cover all sites that return a valid public key and signature. Section 4.2.3 explains how sitekeys work.

The syntax for ⟨*sitekey*⟩ is an exception ⟨*request*⟩ without a ⟨*request-match*⟩ and only specifying '`sitekey`' and '`document`' as ⟨*option*⟩s:

```
1 || @@$sitekey=MFwwDQYJK...wEAAQ,document
```

The '`sitekey`' option is given an RSA public key created by a multi-domain owner, e.g., Sedo. The filter matches all domains that have the same public sitekey and return a signature signed with the private RSA key. The '`document`' option indicates that the entire page is permitted to show ads; even allowing third-party requests.

A.4 Filter Options Explained

Filters may include a number of options to further specify their effects on page elements and scope.[33] An overview of these filters is presented in the following text.

Some options can be grouped by the specific type of element loaded from the request. For instance, '`script`' limits the filter to block (or whitelist for exception filters) the HTML script tag that loads external scripts. Other similar options include '`image`', '`stylesheet`', '`object`' (where browser plugins like Flash or Java handle page content), '`xmlhttprequest`' (for requests by the XMLHttpRequest object[34]), '`object-subrequest`' (for requests started by browser plugins), and '`subdocument`' (for pages that are embedded within the page, usually through HTML frames). The '`other`' option covers requests that are absent from the previous list.

Additional options specify a broader content selection. The '`document`' option may only be included in an exception filter and disables all blocking filters on affected domains and requests. Similarly, '`elemhide`' also only applies to exception filters and disables all element-hiding filters. In contrast, the '`third-party`' option limits a filter only to external requests from a different origin than the current web page. These and all previous options can be prefixed with '∼' to invert their original meaning.

Further, the '`domain`' option restricts the filter to certain domains, which are separated by commas. To *not* apply the filter on a certain domain, the domain can be prefixed with the negation operator. If the '`domain`' option is not specified, the request filter applies to all domains.

More options are also included in the filter syntax. The option '`match-case`' ensures that the request is matched on a case-sensitive basis. A '`collapse`' option will guarantee that the element is hidden.[35] The '`collapse`' option can also be negated. A '`donottrack`' option will send a Do-Not-Track header[36] to the web page as long as there is no matching exception rule with a '`donottrack`' option on the same page.

Finally, there are deprecated options that still exist for backwards compatibility, but their use is discouraged: '`background`', '`xbl`', '`ping`', and '`dtd`'.[37]

[30] `http://www.w3.org/TR/css3-selectors/`

[31] See `https://adblockplus.org/en/filters#elemhide` for more details.

[32] `https://adblockplus.org/en/filters#elemhide_domains`

[33] `https://adblockplus.org/en/filters#options`

[34] `http://www.w3.org/TR/XMLHttpRequest/`

[35] This option is only found once in EasyList and might be deprecated.

[36] `http://donottrack.us/`

[37] `https://adblockplus.org/en/filters#options`