Online Advertising under Internet Censorship

Hira Javaid
LUMS, Pakistan

Hafiz Kamran Khalil
LUMS, Pakistan

Zartash Afzal Uzmi
LUMS, Pakistan

Ihsan Ayyub Qazi
LUMS, Pakistan

ABSTRACT

Online advertising plays a critical role in enabling the free Web by allowing publishers to monetize their services. However, the rise in internet censorship events globally poses an economic threat to the advertising ecosystem. This paper studies this interplay and presents ADVENTION, a system that provides censorship circumvention while serving *relevant* ads. ADVENTION leverages the observation that ad systems are usually hosted on domains that are different from the publisher domains and are almost always uncensored. Taking cue from this, ADVENTION fetches ads via the direct, uncensored, channel between users and the ad system. Preliminary results show that ADVENTION not only offers high ad relevance compared to other popular relay-based circumvention tools, it also offers smaller page load times.

1 INTRODUCTION

Online advertising revenues are projected to reach $83 billion in 2017, an increase of 40% since 2015, and now accounting for the largest share of the advertising market, surpassing both print and TV advertising [1]. Online advertising plays a critical role in fueling the "free" Web and its growth can be attributed both to increase in internet users and the ability to individually customize advertisements to users. While targeted advertising—driven by collecting information about a user's digital habits, locality, and demographics [8, 12]—has raised concerns about user privacy and surveillance, it also offers a number of consumer benefits (e.g., seeing more interesting or *relevant* ads).

With internet censorship events on the rise and over 70 countries filtering Web content [13, 15], the online advertising ecosystem faces a serious challenge. The growing use of censorship circumvention tools, such as Tor [10], Lantern [22], Hotspot Shield, and uProxy [34], implies that advertisers do not infer the *correct* user location information to generate ads as these tools typically rely on proxy relays located outside the censorship region, thus masking the true user location. As a result, users see irrelevant ads (e.g., ads in a language they do not understand or ads for products unavailable in their region). This frustrates the end-user[1], reduces the click through rates, and disrupts the ad campaign, causing a loss of revenue to publishers and advertisers.

This work attempts to explore the adverse impact of censorship circumvention on the relevance of advertisements and how to alleviate such impact. In particular, we ask, "*How can we design censorship circumvention systems that allow advertisers to serve geographically relevant ads to users while retaining circumvention effectiveness?*" To this end, (a) we conduct an exploratory measurement study to understand how ads are impacted by incorrect user location information and quantify the decrease in ad relevance, and (b) present the initial design of ADVENTION, a system that exposes correct user location information to advertisers while still allowing relay-based circumvention of internet censorship.

ADVENTION leverages the observation that the ads served by advertising systems are hosted on domains that are different from the publisher domains and are almost always uncensored to avoid potential collateral damage[2]. Based on this observation, ADVENTION fetches ads through the (uncensored) direct path to the advertisement system—to mimic normal ad requests that are sent without routing through a relay. We show that such an approach also leads to smaller page load times (PLTs) and hence improved user experience as ad requests tend to get served faster over the direct channel. While ADVENTION allows users to circumvent censorship, the publisher—accessed over the relayed path—still sees incorrect user location. The publisher may then generate irrelevant content[3], thus passing incorrect *publisher context* to the ad service, leading to irrelevant ads. To address this challenge, ADVENTION allows the possibility of improving publisher context by intelligent relay selection (IRS) via which requests to publishers are routed.

We evaluate ADVENTION for a range of popular websites including both censored and uncensored websites in the region of our experimental study. Our results show a substantial loss in ad relevance when user traffic is routed via relays, as is the current practice with most common circumvention

HotNets-XVI, November 30–December 1, 2017, Palo Alto, CA, USA

ACM ISBN 978-1-4503-5569-8/17/11…$15.00
https://doi.org/10.1145/3152434.3152455

[1] This may also incent more users to employ ad blockers [25].

[2] Authorities rarely, if ever, block ad system domains as it would cause blocking of ads displayed on scores of popular publishers. For example, Google's DoubleClick is used by 1,843,854 publishers and PubMatic is used by 215,046 publishers [11, 29].

[3] Many popular publishers offer region-specific content.

tools [10, 22, 26, 32, 34]. ADVENTION provides significant improvements in fetching relevant ads, with further benefits when it employs IRS. From our experimental study, we also demonstrate that compared to commonly-used tools for circumvention (e.g., Tor and static proxies), ADVENTION shows improvement in PLTs across a variety of popular websites, thus providing a deployment incentive for end-users.

ADVENTION keeps the users interested (lower PLTs and relevant ads) and enables effective ad campaigns (relevant ads with high click-through rates). The mechanism behind ADVENTION is powerful yet simple to implement as a browser extension, for example by making use of proxy automatic configuration (PAC) files [27] or other similar configurations.

ADVENTION's use of the direct path for routing ad requests raises three challenges. First, it can make easier for censors to block traffic by inspecting the HTTP referer field of ad requests. Through measurements of Alexa top 500 websites, we find that over 82% ad requests are sent over HTTPS, which encrypts the referer field. This prevents the censor from blocking most traffic based on this field[4]. Second, if censors collude with ad systems, they can potentially block user traffic. However, this may be less likely in case of popular ad servers (e.g., DoubleClick and PubMatic) that are typically hosted in uncensored regions and serve majority of publisher websites. Finally, exposing user location information to ad servers raises potential user privacy and anonymity concerns. We argue that for users circumventing censorship, this may be less of a concern. Furthermore, ADVENTION focuses on circumvention rather than provisioning anonymous communication, in line with many popular circumvention tools such as Lantern, Hotspot Shield, uProxy, and static proxies.

We summarize our contributions in this work as follows:

- We quantify via an extensive measurement study the decrease in ad relevance when users employ relay-based censorship circumvention.
- We present the design and implementation of ADVENTION, a system for serving relevant ads while retaining circumvention effectiveness, with a provision for IRS.
- A preliminary evaluation of ADVENTION performance, compared to other circumvention tools, which captures: (i) increase in ad relevance, (ii) additional increase in ad relevance when ADVENTION uses IRS, and (iii) decrease in PLTs.

Our goal in this work is three-fold: (a) to highlight the adverse impact of circumvention tools on online advertising, (b) to highlight opportunities for improving ad relevance and PLTs in relay-based circumvention, and (c) to come up with a feasible road-map for developing circumvention tools that realize these opportunities. We hope this work ignites a broader discussion in the community about the interplay between censorship, privacy, and online advertising and how it may impact different stakeholders in the internet ecosystem.

[4]The remaining 18% of ad requests can use IRS.

2 BACKGROUND

Ad Classification and Targeting. Targeted advertising is often driven by: (i) publisher page context (e.g., language and keywords) to serve *contextual* ads, (ii) user profile and browsing history for *behavioral* ads, or (iii) user location to deliver *geo-targeted* ads. We only consider these 'dynamic' ads, served on-the-fly via ad systems[5] and real-time bidding [7]. *Profile ads* such as those served over online social networks largely remain relevant even with the use of a relay, and are not interesting from the point of view of this work.

Information Flow in Serving Ads. On receiving a page request from a client, the publisher sends back the webpage along with an embedded *ad tag*, which contains the page context. The client browser, behind the scenes, forwards the ad tag to the ad server which may also infer user profile and browsing history (via cookies) and location (via geo-mapping). The ad server then uses the available information (e.g., user location, user profile, or publisher context) to serve *geo-targeted*, *behavioral*, or *contextual* ads. This information flow usually allows *relevant* ads to be served.

Information Flow with Relays. When a user routes a connection via circumvention relays, incorrect inference of user location can cause the ad server to serve irrelevant ads. This has two important consequences: (1) it degrades user experience as users prefer relevant, targeted ads over random, untargeted ads [9, 35] and (2) it disrupts the ad campaign, causing a loss of revenue to advertisers, ad servers, and publishers. With a global rise in internet censorship [15] and the use of circumvention relays [26], the advertisement ecosystem may face a large economic impact.

3 A MEASUREMENT STUDY

In this section, we quantify the decrease in ad relevance due to circumvention tools by conducting a measurement study. We collect ads from a range of websites and divide them into two sets, C and U. Websites in set C were censored while those in U were not censored in the region of evaluation. The set U was taken from Alexa top 500 websites [3].

3.1 Methodology

We use Selenium [31] with Firefox to automate our experiments and employ Tor as a circumvention tool. We prevent profile-based ads by not allowing third party cookies[6].

3.1.1 Relevance: Uncensored sites (set U). We capture the set of ads served by each site in U via the direct (i.e., non-relayed) path and use it as the *ground truth* (i.e., these ads are relevant). We then use Tor to capture the set of ads from the same websites; a larger overlap of this set with ground truth means Tor exhibits greater ad relevance.

Capturing the complete ad set: To capture the complete and accurate set of ads, we visit each website multiple times,

[5]Ad networks and, in recent years, ad exchanges are part of the ad system [4, 5, 18].

[6]The Tor browser also disables third-party cookies by default.

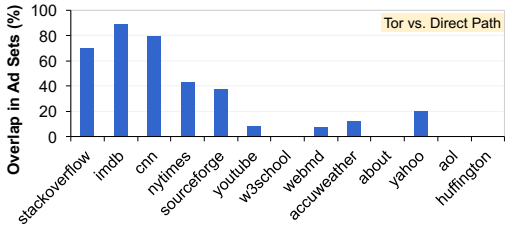

Figure 1: Overlap in ad sets obtained via Tor with the direct path (the ground truth) for various websites.

balancing completeness of ad set with ad churn. After the initial ten visits to capture the ad set as suggested in [18], we further ensure set completeness by continuing to visit the sites thrice every iteration until no new ad is discovered.

Intersection of ad sets: The intersection of ad set obtained by Tor with the ground truth indicates the Tor ad relevance. Before finding the intersection, we remove all duplicates. We use the URL of the *landing page*—the page that opens up when the ad is clicked—as the unique identifier of the ad. We remove the arguments passed in the URL but keep the complete URL path instead of just the domain name. Thus, `fb.com/advert1?foo` and `fb.com/advert1?bar` identify the same ad while `fb.com/advert2` identifies a different one.

3.1.2 Relevance for Censored sites (set C). For censored websites, ground truth cannot be established as the direct path is blocked. To define ad relevance in this case, we use two pieces of information: (i) the language of the landing page (obtained using Python's language detection API, langid [24]), and (ii) the location of the advertiser, inferred from the domain name of the landing page URL.

An ad is considered relevant only if (a) the landing page language is the same as the user language, and (b) either the top level domain (TLD) of the landing page URL is generic (e.g., `.com`, `.org`) or it matches user country.

3.2 Circumvention and Ad Relevance

To measure the impact of relay-based circumvention on ad relevance for uncensored websites (set U), we collect ad sets via Tor and the direct path. Figure 1 shows the overlap in ad sets for each site in U. We observe a 28% overlap between the two sets—averaged over all sites in U—and significant variation in percentage overlap *across* sites. We now summarize our observations from this experiment below.

Location-dependent content. For some websites, we received the same content no matter where the Tor exit relay was located. For others, we observed a change in content (e.g., text, videos, and/or language) from the publisher as we changed the region of the Tor exit relay. Some of these websites (e.g., `youtube` and `yahoo`) had country-specific domains. The set of websites from our dataset serving location-dependent content was {`youtube`, `cnn`, `yahoo`, `aol`, `huffington`}.

Location-dependent ads. For some websites, such as `imdb` and `stackoverflow`, only contextual or global ads were shown independent of user's location while other sites displayed a mix of ads—some contextual, some global, and some geo-targeted. This latter set was {`yahoo`, `nytimes`, `youtube`, `sourceforge`, `about`, `webmd`, `w3school`, `huffington`, `aol`, `accuweather`} from all the uncensored websites (set U) we considered.

Impact on ad relevance. For sites showing only *location-independent* ads, the intersection of the Tor ad set with the ground truth was high as shown by the tall bars in Figure 1. In contrast, for the sites that display location-dependent ads, the intersection was low, as expected. We also found that for these sites, the set of ads significantly (or completely) changed as we picked a Tor exit relay in a different region.

Publisher-specific ad targeting types. Our measurements also enable us to infer and quantify the type of targeting being used by publishers. For instance, contextual ads are identified by taking the intersection of ad sets captured using different Tor exit relays. The remaining set of ads are the ones which changed with location and are, hence, geo-targeted.

Our study evidently shows that several popular websites in Alexa top 500 show location-dependent ads and the use of relay-based circumvention tools can significantly decrease ad relevance, which can negatively impact the advertising campaign and lead to loss in revenue for publishers and advertisers [16]. We observed a similar loss of ad relevance when static proxies (instead of Tor) were used as relays. Thus, it is reasonable to expect that any relay-based circumvention tool would cause a similar loss in ad relevance.

4 SYSTEM GOALS & DESIGN

Serving relevant ads requires availability of correct information at the ad servers i.e., correct user location is needed to serve relevant *geo-targeted* ads, correct profile information is needed to serve relevant *behavioral* ads, and correct publisher context is needed to serve relevant *contextual* ads. Based on the insights from our measurement study, we set the following design goals for a system that allows serving relevant ads while retaining circumvention effectiveness.

(1) It should allow ad server to serve relevant ads, without any performance overhead, to users who bypass censorship by using relay-based circumvention tools.
(2) It should be compatible with the existing online advertising ecosystem.
(3) It should work with a range of relay-based circumvention tools such as Tor [10] and uProxy [34].

4.1 Intelligent Relay Selection (IRS)

For serving contextually relevant ads, ad systems require correct publisher context (e.g., language, keywords, topic, and subtopic). When a user routes webpage requests through a relay, a publisher with region-specific content will return an ad tag with *context* information based on the *relay location*. This context is incorrect from a user's perspective, who will pass this information to the ad server. The contextual ads thus

Approach	Correct info @ Ad Server		Correct info @ Publisher	
	Location	Language	Location	Language
IRS	No	Yes	No	Yes
Advention (without IRS)	Yes	Yes	No	No
Advention (with IRS)	Yes	Yes	No	Yes

Table 1: Depicting what *correct* information is available to the ad server and publisher with various approaches.

served by the ad server will then be irrelevant. A simple fix is to use a relay from within the region of the end user to route connections to both the publisher and the ad server. This, however, will break circumvention effectiveness as censors typically apply censorship to an entire region [13].

An alternate approach, that we take, is to select a relay based in a region that shares the same language as the user. In this way, both the publisher and the ad server infer correct language. For example, users in an English-speaking censorship regime may intelligently select a relay in the US or the UK. This approach does not completely address the problem of incorrect context and may reduce the anonymity set to fewer relay nodes but is useful in serving linguistically relevant ads meant for global audience.

4.2 Advention Design

As an initial step towards exploring the challenges and opportunities in building such a system, we propose ADVENTION; a system that allows advertisers to serve relevant ads while retaining circumvention effectiveness. ADVENTION's design is based on the observation that ad servers are almost always uncensored to avoid possible collateral damage. Thus, with ADVENTION, users access publisher websites using the circumvention tool but obtain ads using the direct path. ADVENTION enables ad servers to correctly infer user location and language to serve geo-targeted ads. The publishers, however, are unable to correctly infer user location and language because users still obtain censored content via a proxy relay. Thus, the ads served via ADVENTION are relevant if they are geo-targeted but not if they are based on publisher context.

Combining ADVENTION with IRS. In our design, we use IRS (which increases linguistic relevance of ads) to make up for the inability of ADVENTION to serve relevant contextual ads. Thus, ADVENTION (with IRS) chooses an intelligent relay—from within a region that shares the same language as the user—to route user connections to the publishers. Users still continue to connect to the ad server without a relay. As shown in Table 1, ADVENTION (with IRS) will let the publisher infer correct user language.

Threat Model. ADVENTION makes two changes to the information flow in proxy-based circumvention tools. First, it uses different paths for obtaining content and advertisements. Second, it uses IRS to obtain correct publisher context. Both of these changes can potentially make it easier for a censor to filter user traffic. We assume an adversary that can block, modify, or reject a web connection at any time in order to filter access but is unwilling to filter all web traffic or ad servers. An adversary may attempt to block user's access to a website

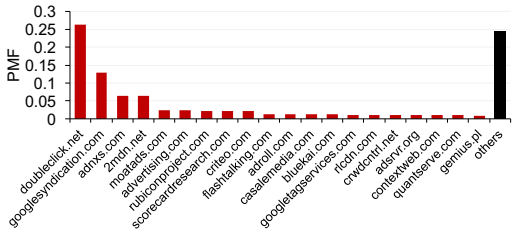

Figure 2: Distribution of ad requests across ad servers.

Protocol	Pub Requests	Ad Requests	Total Requests
HTTP	35.3 %	17.8%	14795
HTTPS	64.7 %	82.2%	31050

Table 2: Protocol distribution (HTTP vs. HTTPS) over publishers and ad requests for Alexa top 500 websites.

if it can infer from the ad request (e.g., via the HTTP referer field) the website a user is visiting by launching a website fingerprinting attack [21].

4.3 Advention's Circumvention Effectiveness

Distribution of ad servers. We conducted measurements of Alexa top 10K websites to study the distribution of ad requests across ad servers. Figure 2 shows that the top 20 ad servers served more than 75.6% of the ad requests.

What fraction of ad requests are served using HTTPS? Sending ad requests directly to the ad server, as done in ADVENTION, may provide censors with new blocking options. A censor may block an ad request by analyzing the request URL or use the HTTP referer field to spot users accessing a blocked website. However, censors are mostly interested in blocking access to the content rather than ads. This may not be a concern in most cases since popular ad servers, such as Google's DoubleClick [11] and PubMatic [29], communicate using HTTPS. Our measurements of the Alexa top 500 websites, with HTTPS Everywhere enabled [14], show that at least 82% of the ad requests use HTTPS as shown in Table 2. As the HTTP referer field is *encrypted* under HTTPS, a censor cannot use it to block *specific* ad requests or communication between the client and the publisher, whereas IP blocking (i.e., based on the destination IP address in a packet) of *all* ad requests will lead to collateral damage.

5 EVALUATION

To evaluate ADVENTION, IRS, and ADVENTION with IRS, we use the same methodology as described in §3.1.

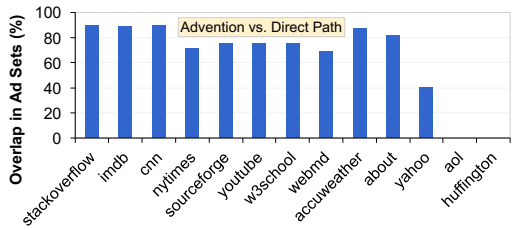

Figure 3: Ad relevance under ADVENTION.

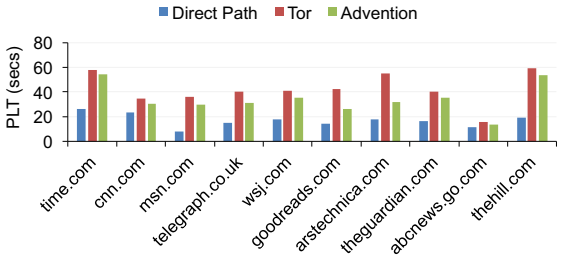

Figure 4: PLTs with ADVENTION.

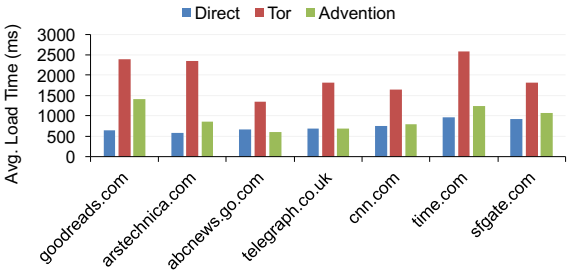

Figure 5: Average ad load times for different websites.

5.1 Advention

Ad relevance. Figure 3 shows the overlap in ad sets with ADVENTION. On average, there was 70% overlap between the two ad sets (compared to 28% with Tor). This suggests that ADVENTION can significantly improve ad relevance. Sites that do not show any location-dependent ads were already displaying relevant ads with Tor. With ADVENTION, there is no adverse impact on the relevance of ads shown on those sites. We also observe that even with ADVENTION, there are websites that exhibit only a small overlap in ad sets (e.g., `aol` and `huffington`). These websites serve location-based content and only context-based ads. Since ADVENTION (without IRS) does not attempt to improve publisher context, these sites continue to display less relevant ads.

Impact on PLTs. In our evaluation, we found that ADVENTION provide up to 47% improvement in the average PLT compared to Tor as shown in Figure 4. This is because ADVENTION avoids the longer relay path and thus speeds up the load times of ad requests (see Figure 5). The exact improvement depends on the structure of the webpage, which determines dependencies, and bottleneck resource(s) in the page load process.

5.2 Intelligent Relay Selection (IRS)

IRS exhibited more or less similar results for ad relevance as were shown by a randomly selected relay. This is because IRS enables content servers and ad servers to correctly infer just the user language but not their location. The effectiveness of IRS in serving ads that use the same language as the user (i.e., linguistically relevant ads) is shown in Figure 6. Compared to Tor, a significantly higher percentage of ads shown via IRS are in the same language as the user. From the figure, we further note that the sites displaying location dependent ads, showed fewer ads in user language when using Tor.

Figure 6: Ads in user's language with IRS compared to Tor's default relay selection mechanism.

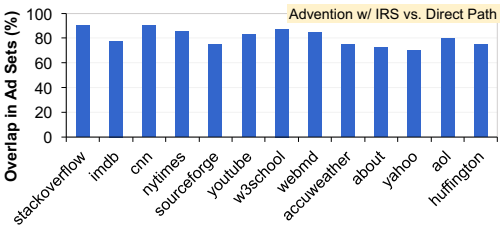

Figure 7: Ad relevance under ADVENTION w/ IRS.

5.3 Advention with IRS

ADVENTION (without IRS) works well for most websites in showing relevant ads. However, sites like `huffington` and `aol`, which serve ads based on publisher context. still display irrelevant ads. To improve the publisher context, at least to the extent of language, we use ADVENTION with IRS. Figure 7 shows the overlap in ad sets obtained via ADVENTION with IRS and a direct path. On average, this overlap is about 80%, an increase beyond just using ADVENTION (without IRS). This is because when IRS is used in conjunction with ADVENTION, ad servers and content servers both infer correct user language and the ad server infers correct user location.

Why isn't the ad relevance 100%? One might expect the overlap in ad sets with the two approaches—direct and ADVENTION w/ IRS—to be close to 100% as both approaches fetch ads via the direct path. To investigate this, we set up another experiment. For the list of websites in our dataset, we capture the ad sets twice, both times via the direct path. The intersection of these two ad sets was observed to be between 80% and 100% for all websites. Due to ad churn and randomness, one cannot guarantee that the two captured sets will have 100% overlap. Guha et al. [19] also observed that even a large number of page reloads cannot guarantee that one can capture the entire set of available ads [18]. Thus, an average overlap of 80% or more is promising as it shows that users are getting ads from the same set that would be used to select ads from, even if they were not using a proxy relay.

5.4 Blocked sites (set C)

We also collected ad sets from blocked websites in the region of experimentation. We access these websites via relays as the direct path was censored.

Using Tor. Table 3 shows the ad relevance for blocked sites using Tor. On average, ~16% of the ads were in the language of the user while the rest used the language of the region of proxy; about 27% of ads belonged to the general category (e.g., `facebook.com`) while no advertisers were from the country of the user. According to the criteria for computing ad relevance for blocked sites (defined in §3.1.2), only 16% of ads shown to the user were relevant when using Tor.

Website	User lang.	User country	General	Relevance
Metro	22.23	0	22.23	22.23
Pamella	33.34	0	33.34	33.34
Zimbio	7.69	0	15.38	7.69
Blog	0	0	40	0
Malkin	16.67	0	25	16.67

Table 3: Ad relevance (%) for blocked sites with Tor.

Using Advention with IRS. Table 4 lists results for blocked sites using ADVENTION with IRS. As expected, all displayed ads were in the user's language; 39.8% of them were from the user's country, while others fell into the general category. No ads of advertisers from proxy's country were shown in this case. Thus, all ads received using ADVENTION with IRS on blocked websites passed our relevance criteria.

Website	User lang.	User country	General	Relevance
Metro	100	45.45	54.55	100
Pamella	100	28.57	71.42	100
Zimbio	100	25	75	100
Blog	100	83.33	16.66	100
Malkin	100	16.67	83.33	100

Table 4: Ad relevance (%) for blocked sites with Advention (w/ IRS).

5.5 Impact of Enabled Cookies

Although third-party cookies are blocked in the Tor browser by default, they are usually enabled when using static proxies or when another browser is used with Tor. To measure the impact of circumvention relays on ad relevance with enabled cookies, we setup a user profile (from Romania) by visiting top 10 Alexa websites with country-specific domains, serving content in the regional language. This profile was established to let the ad server infer user language correctly through cookies. Then we visited various uncensored websites without a relay; we observed that for almost all websites no ads were seen in the Romanian language and no advertiser from Romania was seen. On just one website, we received an ad in Romanian language. This establishes two things: 1) inferred language from cookies does not overwrite inferred language from IP. But inferred language can be used to serve ads, as happened in our case[7], and 2) inferred location from cookies does not overwrite inferred location from IP.

[7]With Google Adsense, users can see ads in the same language as the language of the recently viewed pages by the user [17].

6 RELATED WORK

Several studies have conducted measurements of online advertising systems. Guha et al. [18] identifies challenges in measuring advertising systems and presents an analysis of different classes of advertising. AdReveal [23] seeks to provide transparency into advertising systems by characterizing different types of ad targeting mechanisms used by advertisers. In [30], authors present an empirical study and a classification framework for third-party tracking on the Web. However, none of these works deal with ads shown over relay-based circumvention tools.

There is also a large body of work that focuses on providing *privacy* while serving targeted ads. Privad [19] and Adnostic [33] propose client side software to locally cache ads and generate user profiles to provide a privacy-preserving system. WIT [28] runs as a proxy and protects identity of users by manipulating cookies. We view these works as *complementary* to ADVENTION as they can enable more privacy-preserving circumvention systems.

We are not aware of any work that provides censorship circumvention while enabling relevant ads. Our work is the first to *identify* disruptions of information flow in serving ads and *propose* a mechanism for improving ad relevance.

7 DISCUSSION

Ad-blockers. Ad-blockers, such as Adblock Plus [2], have become popular in recent years. A survey by the Audience-Project [6] showed that 40% users employ ad-blockers to avoid irrelevant ads. Another study by HubSpot showed that 36% people use ad-blockers because ads affect PLTs and bandwidth usage [20]. ADVENTION can serve relevant ads as well as reduce PLTs Although, ADVENTION is not targeted towards users who use ad-blockers—if ads are blocked, it does not matter if they are relevant—it may lead to a decrease in ad-blocker usage.

Anonymous communication. Some users may prefer to use anonymous communication when accessing a censored website. Being a circumvention tool, ADVENTION does not target anonymity from publishers or advertisers. It just aims to avoid blocking by censors. To safeguard user anonymity, a Tor exit relay within a user's region may be used to route connections to the ad server. This will enable correct transfer of user location to the ad server but will disable the ad server from ascertaining user location, thereby preserving anonymity. A detailed analysis of the impact of ADVENTION on anonymity is left as future work.

8 CONCLUSION

This paper highlights the impact of censorship circumvention on online advertising and shows that it is feasible to design systems that can serve relevant ads without reducing the ability to bypass censorship. Our early experiments show such systems can also offer small page load times, thereby creating incentives for users to employ such tools.

REFERENCES

[1] 2016. U.S. digital ad spending to surpass TV this year: Digital will represent 37% of U.S. total media ad spending. In *eMarketer*. https://www.emarketer.com/Article/US-Digital-Ad-Spending-Surpass-TV-this-Year/1014469

[2] ABP. 2017. *Adblock Plus*. https://adblockplus.org/

[3] Alexa. 2017. *The top 500 sites on the web*. http://www.alexa.com/topsites

[4] An OpenX whitepaper. 2013. Ad Networks vs. Ad Exchanges: How They Stack Up. (2013). http://openx.com/whitepapers/.

[5] An OpenX whitepaper. 2015. Ad Exchanges Are (not) All The Same. (2015). http://openx.com/whitepapers/.

[6] AudienceProject. 2017. *Smarter ads can help curb ad blocking*. https://tinyurl.com/yb6vvz8u

[7] Muhammad Ahmad Bashir, Sajjad Arshad, William Robertson, and Christo Wilson. 2016. Tracing Information Flows Between Ad Exchanges Using Retargeted Ads. In *USENIX Security Symposium*.

[8] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. 2015. I always feel like somebody's watching me. In *ACM CoNEXT*.

[9] Farah Chanchary and Sonia Chiasson. 2015. User Perceptions of Sharing, Advertising, and Tracking. In *SOUPS*.

[10] Roger Dingledine, Nick Mathewson, and Paul Syverson. 2004. Tor: The Second-generation Onion Router. In *USENIX Security Symposium*.

[11] DoubleClick. 2017. *Websites using DoubleClick.Net*. http://trends.builtwith.com/websitelist/DoubleClick.Net

[12] S. Englehardt and A. Narayanan. 2016. Online Tracking. In *ACM CCS*.

[13] A. Filasto and J. Appelbaum. 2012. OONI: Open Observatory of Network Interference. In *FOCI*.

[14] Electronic Frontier Foundation. 2017. *HTTPS Everywhere*. https://www.eff.org/https-everywhere

[15] Phillipa Gill, Masashi Crete-Nishihata, Jakub Dalek, Sharon Goldberg, Adam Senft, and Greg Wiseman. 2015. Characterizing Web Censorship Worldwide: Another Look at the OpenNet Initiative Data. *ACM Trans. Web* 9, 1, Article 4 (Jan. 2015), 4:1–4:29 pages.

[16] Phillipa Gill, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, Konstantina Papagiannaki, and Pablo Rodriguez. 2013. Follow the Money: Understanding Economics of Online Aggregation and Advertising. In *ACM IMC*.

[17] Google. 2017. *Ad targeting by language*. https://support.google.com/adsense/answer/2753586

[18] Saikat Guha, Bin Cheng, and Paul Francis. 2010. Challenges in Measuring Online Advertising Systems. In *ACM IMC*.

[19] Saikat Guha, Bin Cheng, and P. Francis. 2011. Privad: Practical Privacy in Online Advertising. In *NSDI*.

[20] HubSpot. 2017. *Why People Block Ads (And What It Means for Marketers and Advertisers)*. https://research.hubspot.com/reports/why-people-block-ads-and-what-it-means-for-marketers-and-advertisers

[21] Marc Juarez, Sadia Afroz, Gunes Acar, Claudia Diaz, and Rachel Greenstadt. 2014. A Critical Evaluation of Website Fingerprinting Attacks. In *ACM CCS*.

[22] Lantern. 2017. *Lantern: Faster than a VPN*. https://getlantern.org/

[23] Bin Liu, Anmol Sheth, Udi Weinsberg, Jaideep Chandrashekar, and Ramesh Govindan. 2013. AdReveal: improving transparency into online targeted advertising. In *ACM HotNets*.

[24] Marco Lui and Timothy Baldwin. 2012. langid.py: An off-the-shelf language identification tool. In *ACL*.

[25] Muhammad Haris Mughees, Zhiyun Qian, and Zubair Shafiq. 2017. Detecting Anti Ad-blockers in the Wild. In *PETS*.

[26] Aqib Nisar, Aqsa Kashaf, Zartash Afzal Uzmi, and Ihsan Ayyub Qazi. 2015. A Case for Marrying Censorship Measurements with Circumvention. In *ACM HotNets*.

[27] PAC. 2017. *Proxy Auto-Config*. https://en.wikipedia.org/wiki/Proxy_auto-config

[28] Fotios Papaodyssefs, Costas Iordanou, Jeremy Blackburn, Konstantina Papagiannaki, and Nikolaos Laoutaris. 2015. Web Identity Translator. In *ACM HotNets*.

[29] Pubmatic. 2017. *Websites using Pubmatic*. http://trends.builtwith.com/websitelist/Pubmatic

[30] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. 2012. Detecting and defending against third-party tracking on the web. In *NSDI*.

[31] Selenium. 2017. *Selenium: Automating Web Browsers*. http://www.seleniumhq.org/about

[32] Hotspot Shield. 2017. *Hotspot Shield*. https://www.hotspotshield.com/

[33] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas. 2010. Adnostic: Privacy Preserving Targeted Advertising. In *NDSS*.

[34] uProxy. 2017. *uProxy: Your private access to the open internet.* https://www.uproxy.org/

[35] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *SOUPS*.