

Annoyed Users: Ads and Ad-Block Usage in the wild

Enric Pujol, Oliver Hohlfeld, Anja Feldmann

Presented by Melanie Klerer

Motivation

- Monetization of free content via ads is rampant throughout the internet
- Ad blocking software is popular with users, Why?
- Is Ad Blocking “stealing” from the “free” content providers?



Major Contributions

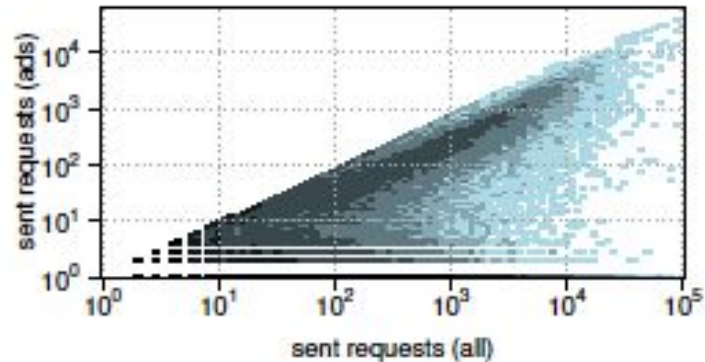
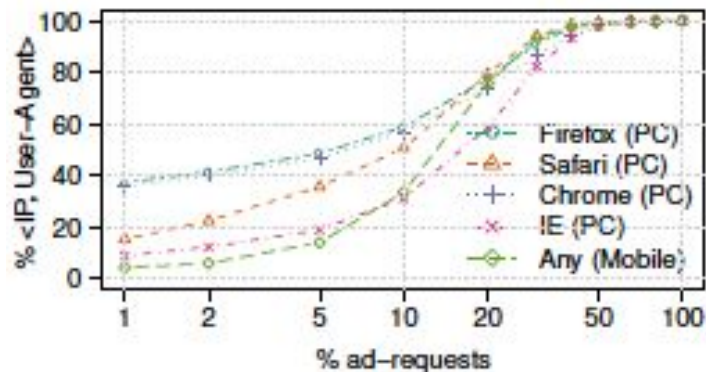
- Measuring ads on a website passively by leveraging AdBlock Plus
- Measuring usage of AdBlock by users via ratio of add requests and connections to AdBlock server (22% of users use Adblock)
- Ad traffic contributes to 18% of requests, 11% of which match non-intrusive ads

Methodology

- Goal: identify ad-blocker users by using TCP/HTTP header traces captured via passive measurements
- Utilizing Bro to get HTTP information
- Adblock needs information about DOM tree to classify ads/non ads
 - Infer the needed information from HTTP header info
 - Set of related URLs via referer header
 - Content type via using generalized Content-Type from header
- Identify users with adblock by seeing ones that view less ads

Active Measurement

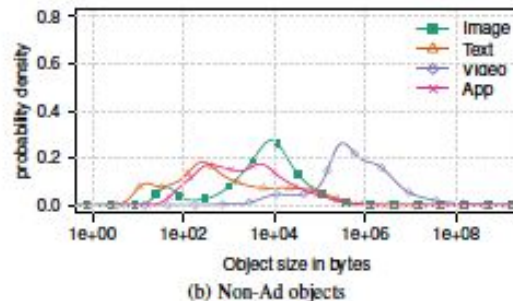
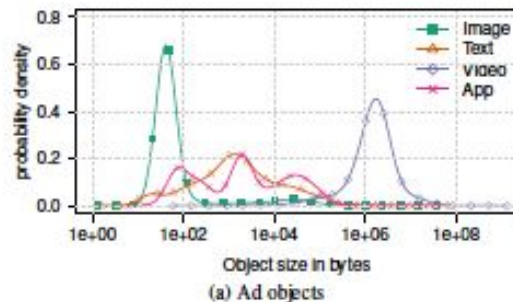
- Adblock reduces number of HTTP request significantly
- Firefox and safari users have a low ratio of ad request
- Users of adblock will have a lower ratio of ad requests and download filter lists



Ads in the wild

- Ad requests follow a diurnal pattern
- Most ads are text/html or images
- Whitelisting non-intrusive ads would benefit sites like Google, which has a large percentage of non-intrusive ads, and not benefit “adult” sites, which tend to have intrusive ads

Content-type	Ads		Non-Ads	
	Reqs	Bytes	Reqs	Bytes
image/gif	35.1%	14.1%	3.5%	0.7%
text/plain	28.7%	34.2%	14.3%	2.1%
text/html	14.4%	11.8%	7.6%	1.1%
-	11.8%	5.4%	28.7%	63.4%
app./xml	4.5%	2.2%	1.8%	0.2%
image/png	1.9%	1.6%	5.1%	0.8%
image/jpeg	1.8%	5.2%	19.8%	3.4%
app./x-shock.	1.4%	8.1%	0.2%	0.1%
video/mp4	0.0%	10.9%	0.3%	8.6%
video/x-flv	0.0%	5.4%	0.1%	3.1%



Ad infrastructure

- Concentrated in a few ASes
- Ad-tech agency deploy infrastructure, rely on CDNs, manage ASes
- Same infrastructure serves ads and regular content



Thoughts on the Paper?