Adblocking and Counter-Blocking

Rishab Nithyanand
Sheharbano Khattak
Mobin Javed
Narseo Vallina-Rodriguez
Marjan Falahrastegar

Julia E. Powles
Emiliano De Cristofaro
Hamed Haddadi
Steven Murdoch
Now... our sponsors

- Open Technology Fund
- The Royal Society
- EPSRC
- Xerox

Skipping slide in 15 seconds
Or... Turn off your adblocker to proceed
In this paper, we ...

- Understand the current state of the content provider vs. adblocker arms race

Quick Facts:

- Ad blocking is estimated to cost publishers nearly $22 bil.
- There are now 198 million active adblock users around the globe.
- Ad blocking grew by 41% in the last 12 months.
- US ad blocking grew by 40% to reach 45 million active adblock users.
- UK ad blocking grew by 82% to reach 12 million active users.

Adobe and PageFair 2015 ad blocking report

Eyeo wins 5th straight case in Germany

Security and privacy activists fight back

And immediately face malvertising scares

Popular websites deploy anti-adblockers

August ‘15
March ‘16
March ‘16
April ‘16
In this paper, we ...

• Understand the current state of the content provider vs. adblocker arms race

  • How widespread are anti-adblockers?
    • Anti-adblockers: Adblocker blockers and detectors.

  • What sites are employing them?

  • Who are the most popular vendors of anti-adblockers?

  • How do the popular anti-adblockers work?

  • Can adblockers block anti-adblockers?
    • Or, can adblockers block adblocker blockers?
How widespread are anti-adblockers?

• Crawl and download all JSES in the Alexa T-5K.
  • Over 200K JS in Alexa T-5K.
  • Analyzing (functionality of) each one is not feasible.

• Observation: Most sites use the same JSES.
  • Bought from vendors, part of a library, etc.

• Cluster similar JSES and analyze each cluster.
Clustering similar JSes

• Formulate as the Clique Finding problem.

• Find all cliques with edge weight $\geq x$ (say, .80).

• JS 1, JS 2, and JS 4 are likely to have identical functionality and source for sufficiently high “$x$”.

• Gives a lower-bound of shared JSes. Fails with differently obfuscated code and composed JSes.

Similarity of code (TF-IDF)
- White-space insensitive
- Position insensitive
- Resistant to noise
Analyzing clusters of JSes

• Characteristics (script vendor and functionality) of each clique are identified by analyzing:

  • Sources of scripts.
  
  • External resources utilized by the scripts.
  
  • (TF-IDF) Keywords associated with the script.
  
  • Manually.
How widespread are anti-adblockers?

• Total number of cliques with size $\geq 5$ : 1882
  • 1882 JSes shared by 5+ websites.

• Total number of tracker JSes: 456 of 1882
  • 456 JSes used for tracking shared by 2741 websites.

• Total number of anti-adblocker JSes: 22 of 1882
  • 22 JSes used for detecting adblockers shared by 335 websites.
  • This is a lower bound.
What sites are employing anti-adblockers?

- McAfee’s URL categories over the 335 identified websites.

<table>
<thead>
<tr>
<th>%</th>
<th>Category</th>
<th>%</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.5%</td>
<td>General News</td>
<td>2.5%</td>
<td>Pornography</td>
</tr>
<tr>
<td>9.3%</td>
<td>Blogs/Wiki</td>
<td>2.5%</td>
<td>Forum/Bulletin Boards</td>
</tr>
<tr>
<td>8.5%</td>
<td>Entertainment</td>
<td>2.2%</td>
<td>Technical/Business Forums</td>
</tr>
<tr>
<td>4.3%</td>
<td>Internet Services</td>
<td>2.2%</td>
<td>Potential Illegal Software</td>
</tr>
<tr>
<td>3.7%</td>
<td>Sports</td>
<td>2.0%</td>
<td>Online Shopping</td>
</tr>
<tr>
<td>3.7%</td>
<td>Games</td>
<td>1.7%</td>
<td>Portal Sites</td>
</tr>
<tr>
<td>3.2%</td>
<td>Travel</td>
<td>1.7%</td>
<td>Humor/Comics</td>
</tr>
<tr>
<td>3.2%</td>
<td>Education/Reference</td>
<td>1.2%</td>
<td>Social Networking</td>
</tr>
<tr>
<td>2.7%</td>
<td>Business</td>
<td>1.2%</td>
<td>Provocative Attire</td>
</tr>
<tr>
<td>2.5%</td>
<td>Software/Hardware</td>
<td>1.2%</td>
<td>Marketing/Merchandising</td>
</tr>
</tbody>
</table>
Who are the popular anti-adblocking vendors?

<table>
<thead>
<tr>
<th>Domain</th>
<th>Description</th>
<th>#Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>pagefair.com</td>
<td>Anti-adblocking</td>
<td>20</td>
</tr>
<tr>
<td>googleadservices.com</td>
<td>Ads</td>
<td>61</td>
</tr>
<tr>
<td>google syndication.com</td>
<td>Ads</td>
<td>13</td>
</tr>
<tr>
<td>taboola.com</td>
<td>Ads</td>
<td>36</td>
</tr>
<tr>
<td>outbrain.com</td>
<td>Ads</td>
<td>10</td>
</tr>
<tr>
<td>ensighten.com</td>
<td>Ads</td>
<td>6</td>
</tr>
<tr>
<td>hotjar.com</td>
<td>Analytics</td>
<td>9</td>
</tr>
<tr>
<td>doublepimp.com</td>
<td>Pornography</td>
<td>8</td>
</tr>
<tr>
<td>tacdn.com</td>
<td>Travel</td>
<td>8</td>
</tr>
<tr>
<td>cloudflare.com</td>
<td>CDN</td>
<td>50</td>
</tr>
<tr>
<td>cloudfront.net</td>
<td>CDN</td>
<td>6</td>
</tr>
<tr>
<td>ytimg.com</td>
<td>Content/Ads</td>
<td>108</td>
</tr>
</tbody>
</table>
How do anti-adblockers work?

• Key idea: Bait objects.
• Example 1: Using a bait DIV element

```javascript
var test = document.createElement('div'); //Create a new DIV element
test.innerHTML = '&nbsp;';
test.className = 'adsbox'; //Set bait

document.body.appendChild(test); //Add element to document body

if (test.offsetHeight === 0) //Check element height
{
    document.body.classList.add('adblock');
}
```
How do anti-adblockers work?

• Key idea: Bait objects.
• Example 2: Using bait JSes

```javascript
// ads.js
var foo = False

// main.js
var foo = True

// main.html
if (foo is True)
{
    document.body.classList.add('adblock');
}
```
Can adblockers block anti-adblockers?

<table>
<thead>
<tr>
<th>Domain</th>
<th>Description</th>
<th>#Sites</th>
<th>ABP</th>
<th>Gh</th>
<th>PB</th>
</tr>
</thead>
<tbody>
<tr>
<td>pagefair.com</td>
<td>Anti-adblocking</td>
<td>20</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>googleadservices.com</td>
<td>Ads</td>
<td>61</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>googlesyndication.com</td>
<td>Ads</td>
<td>13</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>taboolaa.com</td>
<td>Ads</td>
<td>36</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>outbrain.com</td>
<td>Ads</td>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ensighten.com</td>
<td>Ads</td>
<td>6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>hotjar.com</td>
<td>Analytics</td>
<td>9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>doublepimp.com</td>
<td>Pornography</td>
<td>8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>tacidn.com</td>
<td>Travel</td>
<td>8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>cloudflare.com</td>
<td>CDN</td>
<td>50</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>cloudfront.net</td>
<td>CDN</td>
<td>6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>yting.com</td>
<td>Content/Ads</td>
<td>108</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
TL; DR

- Anti-adblockers are employed by at least 6.7% of the Alexa T-5K.

- Anti-adblockers are mostly employed by “News” websites.

- Anti-adblockers are being supplied mostly by ad vendors.
  - PageFair report should be taken with a grain of salt (conflict of interest).

- The arms race has entered the next phase.
  - Escalation seems likely unless a compromise is found (e.g., Eyeo AAI).
Upcoming work

• Automated categorization of JSes.

• Longitudinal measurements.