Errors, Misunderstandings, and Attacks: Analyzing the Crowdsourcing Process of Ad-blocking Systems

Mshabab Alrizah
Sencun Zhu
Xinyu Xing
Gang Wang

PennState
PennState
PennState
Ad-Blocking System
Ad-Blocking System

Objectives
Datasets
FP & FN errors
Evasion
Methodology
Background
Ad Blocking System
Crowdsourcing and Ad-blocking Systems

Objectives

Datasets

FP & FN errors

Evasion

Methodology

Background

Want big impact? Use big image.

WWW

Internet Users 

Reports 

Filter List Editor

c. Community

a. Software

b. Filter Lists

List A
Filter 1
Filter 3

List B
Filter 1
Filter 3
Filter 2

List C
Filter 1
Filter 3
Filter 2

ABP

UO
Previous Work Studied ...

- Different problems or complementary solutions.
- Economic ramifications of the ad-blocking systems
- Specific cases of ad blocking.
  - *e.g.*, trackerblocking, anti-adblocking
- Relationships among Internet users, ad publishers, and ad blocker
Yet...

- Remains a lack of deep understanding on:
  - Filter list effectiveness
  - The crowdsourcing functionality and contribution
  - The potential pitfalls and security vulnerabilities
Objectives

Provide an in-depth study on the dynamic changes of the filter-list to answer the following key questions.

— Q1: How prevalent are the errors of missing real advertisements (false negative (FN) errors) and the errors of blocking legitimate content (false positive (FP) errors)?

— Q2: What are the primary sources of FP errors?

— Q3: How effective is crowdsourcing in detecting and mitigating FP and FN errors?

— Q4: How robust is the filter-list against evasion attacks?
Objectives

Q1: (prevalence of FN and FP errors)?

Q2: (primary sources of FP errors)?

Q3: (crowdsourcing effectiveness)?

Q4: (Robustness of the filter-list)?
Methodology

Datasets Collecting and Cleaning

- Collect and track **dynamic changes** of the filter list (EasyList)
  - Cleaning and extracting those versions created to correct FP and FN errors
- Extract filter **rules** added or removed and build a **record** for each rule.
  - Each record contains information about the rule (e.g. time of creation, deletion, EasyList versions).

**Dataset D1**

- Collect posts of FP and FN errors in EasyList forum.
  - 23,240 topics with at least one report.
- Extract the reports from the posts and build a record of each report.
  - The report record contains information such the contributor profile, webpage has the error, EasyList editor responses....

**Dataset D2**

Background

- Collect and track dynamic changes of the filter list (EasyList)
- Extract filter rules added or removed and build a record for each rule.
Q1: Error Prevalence

- To answer the question we need to know:
  1. Types of the errors
  2. Websites with the errors

Problem: Many reports do not have evidences of correction

Solution: Link Reports with EasyList
Q2: Primary Sources of FP Errors

• Required knowledge:
  – The web page that has the FP error(s).
  – The element impacted.
  – The filter that caused the error.
  – The EasyList versions created to fix the error(s).

• Reproduced FPs using:
Q3: Crowdsourcing Effectiveness

- Extracting from **Dataset D2** the crowdsourcing behaviors:
  - Reports
  - Type of Report
  - Reporter profile
  - EasyList editor response
  - EasyList editor profile
  - Time of correction
  - Reason of rejection
  - .......
Q4: Robustness of Filter-List

- Extract from **Dataset D1** the EasyList’s behavior:
  ✓ Reasons of adding rules.
  ✓ Syntax of rules.
  ✓ Ad server’s domains.
  ✓ Change of ad element attribute.
  ✓ …

- Extract from **Dataset D2** the websites’ behaviors:
  ✓ Reasons of FN errors.
  ✓ Responses of EasyList community.
  ✓ …

- Study the reaction of ad networks.
  ✓ Historical traffic information of the ad-severs.
  ✓ …
## Datasets

### Dataset D1

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleaned EasyList Versions</td>
<td>55,607</td>
<td>From November 30, 2009, to December 7, 2018</td>
</tr>
<tr>
<td>Added Filter Rules</td>
<td>534,020</td>
<td>In order to correct FP and FN errors</td>
</tr>
<tr>
<td>Removed Filter Rules</td>
<td>448,479</td>
<td>In order to correct FP and FN errors</td>
</tr>
</tbody>
</table>

### Dataset D2

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reports of FN errors</td>
<td>17,968</td>
<td>From November 30, 2009, to December 7, 2018</td>
</tr>
<tr>
<td>Reports of FP errors</td>
<td>5,272</td>
<td>From November 30, 2009, to December 7, 2018</td>
</tr>
</tbody>
</table>

### Dataset D2A: Linking EasyList Filter Rules with True Reports

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP and FN Reports linked with EasyList changes</td>
<td>5,284</td>
<td>From November 30, 2009, to December 7, 2018</td>
</tr>
</tbody>
</table>

### Dataset D2B: Reproducing FPs

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Instances of FP errors</td>
<td>570</td>
<td>2,203 webpages studied.</td>
</tr>
</tbody>
</table>

### Ad-servers: traffic information

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical traffic information of the ad-servers</td>
<td>567,293</td>
<td>Traffic information of 6903 ad server domains during 4-years.</td>
</tr>
</tbody>
</table>
Q1 Analysis: Error Prevalence

Websites with FN and FP errors
Q2 Analysis: Sources of FP Errors:

The responsibility (the source of the error) for the errors.

Objectives

Datasets

FP & FN errors

Evasion

Methodology

Background

Time

t1 < t2 < t3

Web Designer's create

Non-Ad Content

Ad Blocker create

Filter Rule

Ad blocker's Fault

Web Designer's Fault

0%

30%

60%

90%

Block Request

Hide Element

The responsibility (the source of the error) for the errors.
## Q3 Analysis: Crowdsourcing Effectiveness

<table>
<thead>
<tr>
<th>Title</th>
<th># Reports</th>
<th>Avg. of days</th>
<th>SD.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FP</td>
<td>FN</td>
<td></td>
</tr>
<tr>
<td>Anonymous</td>
<td>530</td>
<td>853</td>
<td>2.37</td>
</tr>
<tr>
<td>New Member</td>
<td>371</td>
<td>307</td>
<td>3.94</td>
</tr>
<tr>
<td>Senior Member</td>
<td>160</td>
<td>749</td>
<td>2.31</td>
</tr>
<tr>
<td>Developer</td>
<td>83</td>
<td>99</td>
<td>1.80</td>
</tr>
<tr>
<td>Other Lists Editor</td>
<td>105</td>
<td>603</td>
<td>1.65</td>
</tr>
<tr>
<td>Veteran</td>
<td>255</td>
<td>751</td>
<td>1.95</td>
</tr>
<tr>
<td>Editor</td>
<td>80</td>
<td>338</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,584</strong></td>
<td><strong>3,700</strong></td>
<td><strong>2.09</strong></td>
</tr>
</tbody>
</table>

FP and FN error reports submitted by different categories of users.
Contributions by Different Types of Users

**FN Reports**
- Anonymous: 23%
- Veteran: 20%
- Senior Member: 20%
- New Member: 8%
- Editor: 9%
- Other Lists: 17%
- Developer: 3%

**FP Reports**
- Anonymous: 34%
- Veteran: 16%
- New Member: 23%
- Editor: 5%
- Other Lists: 5%
- Senior Member: 10%
- Developer: 7%
Contributions of Different Types of Users

- To Anonymous, New Member, Senior Member, and Veteran classes, the error type and website popularity dependent.
  - Anonymous and New Members contributed more on correcting FP errors than FN errors for lower-rank websites.
    - Expert members tend to the opposite side.
Delay in Reporting FP Errors

Delay of reporting FP errors
Q4 Analysis: Robustness of Filter-list against Evasion Attacks

15 different evasion attacks:
  • More-Studied Attacks (4),
  • Less-Studied Attacks (3), and
  • Nonstudied Attacks (8).
# More-Studied Attacks

<table>
<thead>
<tr>
<th>Attacks</th>
<th>Our Findings</th>
</tr>
</thead>
</table>
| WebSockets.                                  | Since 2016, EasyList had blocked:  
• 291 websites.  
• 137 ad servers.                              |
| Anti-ad Blocker.                             | Reaction:  
• Restricting content on the sites (paywalls, blocking the websites)  
• Redirecting the users to different websites or content.                 |
| Randomization of Ad Attributes and URLs.     |  
• 15 websites using randomization.  
• Facebook appeared most frequently. |
| Factoring Acceptable Ads List Sitekeys.       | Our datasets do not show any of this attack.                                |
## Less-Studied Attacks

<table>
<thead>
<tr>
<th>Attacks</th>
<th>Our Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changing Ad-Server Domains.</td>
<td>• 52% of the ad servers’ traffic activities disappeared in three days.</td>
</tr>
<tr>
<td></td>
<td>• 84% of these 52% ad servers were blocked shortly after they were used.</td>
</tr>
<tr>
<td></td>
<td>• Ad servers with long life: 61% were significantly influenced by the blocking.</td>
</tr>
<tr>
<td></td>
<td>• The EasyList community ran code to monitor the changes of ad servers (limited)</td>
</tr>
<tr>
<td>Changing Ad-Element Attributes.</td>
<td>• EasyList did not have the capability to automatically trace the changes of ad elements.</td>
</tr>
<tr>
<td></td>
<td>• Manually detected by EasyList, 553 instances changed the filters in response to this type of evasion.</td>
</tr>
<tr>
<td>Changing the Path of Ad Source.</td>
<td>• 644 websites changed their the ad URL’s paths.</td>
</tr>
</tbody>
</table>
Nonstudied Attacks

1. Exploiting **Obsolete Whitelist Filters**.
2. Using **Generic Exception Rules** (Whitelist Filters).
3. Exploiting **False Positive** Errors.
4. **First-Party** Content and Inline Script.
5. **ISP** Injecting Ads.
6. **Background** Redirection.
7. Exploiting **WebRTC**.
8. **CSS Background Image Hack**.

Domains in the whitelist filters were not monitored by EasyList.

EasyList counters the anti-ad blocker or solve FP error by GER. So ..?
Limitations and Future Work

• Limitations:
  – The dataset covered the historical dated back to 2009. We could not find any data before November 2009.
  – Conservative approach was used to link the reported errors to the EasyList updates.
    • Trade-off between the scale of the data and the accuracy of the analysis.
  – The Internet Archive data was limited.

• Future work
  – Crowdsourcing mechanisms.
  – Dynamic analysis.
  – And more...
Conclusions

• An in-depth measurement study to reveal
  — Q1: Prevalence of FP and FN errors
  — Q2: Primary sources of FP errors
  — Q3: Effectiveness of crowdsourcing in detecting and mitigating FP and FN errors
  — Q4: Robustness of filter-list against evasion attacks?

• Our findings are expected to help shed light on any future work to evolve ad blocking and/or to optimize crowdsourcing mechanisms.
THANK YOU!

Mshabab Alrizah
maa25@psu.edu