Harpo: Learning to Subvert Online Behavioral Advertising

Jiang Zhang¹, Konstantinos Psounis¹, Muhammad Haroon², Zubair Shafiq²
¹ University of Southern California ²University of California, Davis

Introduction
Privacy-invasive tracking techniques for user profiling and subsequent ad targeting

Existing privacy-enhancing solutions

A real privacy threat

System Design

User persona

Environment:

Trickster

Content feature

Obluscation URL

Harpo extraction

Privacy metrics

User profiling

Content-based

Obluscation - Offensive - More ecosystem-friendly - But not principle/stealthy

Blocking - Defensive - Can be circumvented - Kill advertising ecosystem

Ad blocking

User persona

Ad targeting

Page content

User persona

Ad targeting

RL Agent

Modelled via Conv + LSTM Trained using A2C (Advantage Actor and Critic)

Surrogate Model

Replicate real-world tracker models

Why? Virtual environment for efficiently training RL

Harpo browser extension

Content Feature Extraction

Doc2vec embedding

User persona

Bid values

Threat Model

User URL

Obluscation URL

User persona

URL Agent

User profiling

Content-based

Tracker: - A third-party to provide advertising & tracking services

The obluscator should be:
- Seamless
- Principled
- Steady
- Low overhead

Main assumptions for tracker:
- Complete coverage of a user’s browsing history
- Substantial computational resources to train ML models for tracking

Two real-world tracker models:
- User profiling model:
  - User persona
  - Interest segments (user profiles)

- Ad targeting model:
  - User persona
  - User profiles
  - Bid values

Evaluation

Privacy:

<table>
<thead>
<tr>
<th>Approaches</th>
<th>( L_1 )</th>
<th>( L_2 )</th>
<th>( L_3 )</th>
<th>( L_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>1.00</td>
</tr>
<tr>
<td>AdNauseam</td>
<td>12.85%</td>
<td>1.53</td>
<td>2.70%</td>
<td>1.21</td>
</tr>
<tr>
<td>TrackThis</td>
<td>32.67%</td>
<td>2.81</td>
<td>-1.50%</td>
<td>0.89</td>
</tr>
<tr>
<td>Rand-intent</td>
<td>33.10%</td>
<td>3.18</td>
<td>8.40%</td>
<td>1.69</td>
</tr>
<tr>
<td>Bias-intent</td>
<td>31.72%</td>
<td>3.19</td>
<td>10.30%</td>
<td>2.07</td>
</tr>
<tr>
<td>Harpo</td>
<td>43.24%</td>
<td>5.22</td>
<td>43.30%</td>
<td>6.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Against user profiling</td>
</tr>
<tr>
<td>Up to 3x</td>
</tr>
<tr>
<td>Against ad targeting</td>
</tr>
<tr>
<td>Up to 16x</td>
</tr>
</tbody>
</table>

Summary

Contributions:
- Propose HARPO, a principled RL-based approach to obfuscate a user’s browsing history.
- Develop surrogate ML models to train HARPO’s RL agent with limited or no black-box access to real-world tracker models.
- Demonstrate the success of HARPO against real-world user profiling and ad targeting models in terms of privacy, overhead, and stealthiness.

Harpo’s Artifacts: https://github.com/bits2015/Harpo-Artifacts


Stealthiness:

Better privacy and stealthiness tradeoff

Minimal impact on overall user experience
- Increased page load time: 0.2 sec
- Increased CPU usage: 5.3%
- Increased memory usage: 3.9%

Note that tracker will run fraud detection to detect the usage of Harpo.