The Inventory is Dark and Full of Misinformation:
Understanding Ad Inventory Pooling in the Ad-Tech Supply Chain

Paper: https://tinyurl.com/darkpooling
Code & Data: https://tinyurl.com/darkpooling-code-data

Yash Vekaria\textsuperscript{1}
Rishab Nithyanand\textsuperscript{2}
Zubair Shafiq\textsuperscript{1}

\textsuperscript{1}University of California, Davis
\textsuperscript{2}University of Iowa
The Democrats' War on Guns is continuing, with the newest battlegrounds being Minnesota and Pennsylvania.

U.S.A. — After pushing through gun control measures in Michigan and Washington, Democrats—dropping any pretense of being anything other than the “party of gun control”—are busy in Minnesota, where, according to the Twin Cities Pioneer Press, they moved a bill on a “party line vote” to expand background checks and a red flag law.

The newspaper notes in its coverage that “Democratic backers of these changes say they will help keep guns out of the hands of criminals and dangerous people. They say the new restrictions are ‘common sense’ measures that have broad support among voters and have been implemented in both liberal and conservative states.”
How does an ad for a reputed brand end up on a misinformation site?

Advertisers

Sample Inventory

Ad-exchange

Publishers

bad.com
okay.com
good.com
How does an ad for a reputed brand end up on a misinformation site?

Advertisers

Sample Inventory

Ad-exchange

intermediary.com

Publishers

bad.com

okay.com

good.com
Is brand safety effective against dark pooling?

Domain: good.com
SellerID: A1

PubID

bad.com

good.com

bad.com
good.com
Research Questions

- Compliance of ads.txt and sellers.json transparency standards
- Prevalence of dark pooling
- Effectiveness of brand safety in the face of dark pooling
Data Collection

- Visit Website
- Extract links
- Retain Ad URLs
- Click on Ads
- Extract brand from landing page
- Extract triplets from network traffic

- ads.txt
- sellers.json
- <pub, adx, owner>

- C_{100K}
- C_{rank}
- M_{full}
- D_{static}
- D_{crawls}
- D_{brands}
## Results (RQ1): Compliance

### Misrepresentations amongst our data:

<table>
<thead>
<tr>
<th>ads.txt Misrepresentation</th>
<th>Misinformation</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misrepresented direct relationships</td>
<td>64%</td>
<td>51%</td>
</tr>
<tr>
<td>Fabricated publisher/seller IDs</td>
<td>83%</td>
<td>65%</td>
</tr>
<tr>
<td>Conflicting relationships</td>
<td>49%</td>
<td>33%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sellers.json Misrepresentation</th>
<th>&gt;= 1 Misinformation Site</th>
<th>0 Misinformation Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invalid Seller Type</td>
<td>0%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Invalid Domain names</td>
<td>54.8%</td>
<td>0.8%</td>
</tr>
<tr>
<td>Confidential Sellers</td>
<td>46.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>INT w/o sellers.json</td>
<td>49.8%</td>
<td>13.3%</td>
</tr>
<tr>
<td>Non-unique IDs</td>
<td>95.3%</td>
<td>62.6%</td>
</tr>
</tbody>
</table>
Publisher IDs are shared by organizationally-unrelated publishers

publisherA & publisherB owned by the same Parent Company  

publisherA & publisherB are NOT owned by the same Parent Company
RQ2: Classifying Dark Pools

Based on Parent Organization of pooled publishers

1. Homogenous Pools

- A Times and B Times authorize 12345 as their direct seller. AdExchange recognizing 12345 as belonging to either A Times (or B Times) or AB Group are all valid cases of ID sharing since A Times and B Times are related (same parent).

2. Heterogenous Pools

- A Times and B Times authorize 12345 as their direct seller. AdExchange recognizes 12345 as belonging to A Times (or B Times). This type of ID pooling is incorrect since A Times and B Times are unrelated.
### Results (RQ2): Static Pooling

<table>
<thead>
<tr>
<th>Pool Type</th>
<th>Pools w/ Misinformation Sites</th>
<th>Pools w/o Misinformation Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># Pools</td>
<td>Pool Size</td>
</tr>
<tr>
<td>Homogenous</td>
<td>0.4%</td>
<td>2.6</td>
</tr>
<tr>
<td>Potential Homogenous</td>
<td>9.1%</td>
<td>18.8</td>
</tr>
<tr>
<td>Heterogenous</td>
<td>85.0%</td>
<td>482.5</td>
</tr>
<tr>
<td>Unknown</td>
<td>5.6%</td>
<td>4.3</td>
</tr>
</tbody>
</table>
Results (RQ2): Dynamic Pooling

- Unique misinformation pools: 297

- Majorly we observed syndication based ad-networks involved in pooling

- Pools with highest misinfo. sites sharing the same PubID
  - PubID owner: 33Across (30 Misinfo Domains) (issuer: Pubmatic)
    - Principia-scientific.org
    - ....
  - PubID owner: Gourmet Ads (23 Misinfo Domains) (issuer: Pubmatic)
    - Principia-scientific.org
    - ....

PubID owner domains*:
- confidential_seller: 198
- 33across.com: 39
- gourmetads.com: 33

PubID issuers*:
- pubmatic.com: 279
- google.com: 243
- rubiconproject.com: 54

*common across multiple pools (total across 2 crawls)
Results (RQ3): Brand Analysis

We collected a total of 4246 (2068 distinct brands) ads from 669 Misinformation websites.

- Most prevalent brands were: Amazon, Alibaba, Yahoo
- Top advertiser categories: business, shopping, computers, health, & finance
- We observed that reputable brands advertised more on the misinformation sites which were part of more pools.
1. Better standards and compliance

2. Trust delegation: Better vetting of publishers in ad syndication networks

3. Regulation
   a. Digital Services Act (DSA)
   b. Digital Services Oversight and Safety Act (DSOSA)
Research Paper
https://tinyurl.com/darkpooling

Code & Data
https://tinyurl.com/darkpooling-code-data

Yash Vekaria
@vekariayash
yash-vekaria
yash-vekaria.github.io

Rishab Nithyanand

Zubair Shafiq

UCD AVIS
UNIVERSITY OF CALIFORNIA

IOWA

UCD AVIS
UNIVERSITY OF CALIFORNIA