Selling Off Privacy at Auctions

Lukasz OLEJNIK, Minh-Dung TRAN, Claude CASTELLUCCIA
Online Advertising

Ad spaces

Ad sellers (Publishers)  Ad buyers (Advertisers)
Real-Time Bidding (RTB)

- RTB allows ad buyers (advertisers) and sellers (publishers) buy and sell ads in real time on a *per-ad-impression* basis through an Ad Exchange.
RTB Protocol

accuweather.com

(1) Load page

Web page

Publisher
RTB Protocol

1. Load page
2. Ad request

Publisher

Web page

DoubleClick Ad Exchange

accuweather.com
RTB Protocol

1. Load page
2. Ad request
3. Bid Request: User Information + Page Information

Publisher

Web page

Ad Exchange

DoubleClick

Accuweather.com

Bidders

Criteo

Turn

...
RTB Protocol

1. Load page

2. Ad request

3. Bid Request: User Information + Page Information

4. Bid response

accuweather.com

DoubleClick Ad Exchange

Criteo

Turn

Publisher

Ad Exchange

Bidders

Tuesday, February 25, 14
RTB Protocol

1. Load page
2. Ad request
3. Bid Request: User Information + Page Information
4. Bid response

Web page

accuweather.com

DoubleClick Ad Exchange

Criteo

Turn

Publisher

Ad Exchange

Bidders
RTB Protocol

1. Load page
2. Ad request
3. Bid Request: User Information + Page Information
4. Bid response
5. Ads

Publisher | Ad Exchange | Bidders

accuweather.com

DoubleClick Ad Exchange

Criteo

Turn
Cookie Matching

Bid request
- ADX’s cookie
- Visiting URL

Bidder

User profile (at Bidder’s DB)
- Bidder’s cookie | URL1
- | URL2
- | URL3
- | ....
Cookie Matching

- CM helps bidders recognize users at bid requests:

```
<table>
<thead>
<tr>
<th>Bid request</th>
<th>User profile (at Bidder’s DB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADX’s cookie</td>
<td>Bidder’s cookie</td>
</tr>
<tr>
<td>Visiting URL</td>
<td>URL1</td>
</tr>
<tr>
<td></td>
<td>URL2</td>
</tr>
<tr>
<td></td>
<td>URL3</td>
</tr>
<tr>
<td></td>
<td>....</td>
</tr>
</tbody>
</table>
```
RTB in Online Advertising

- RTB is rapidly growing, expectedly accounting for:
  - 27% of total display advertising sales in the US, 25% in United Kingdom, 21% in France, 20% in Germany.
  - 65% of indirect display ad sales revenue in the US and 55% in the most developed European markets.
  
  ...by 2015! (IDC, 2011)

- And privacy?
Product in RTB auction

User’s presence on a website

User X is now on website Y. Are you interested?

Is it worth showing ads to X? At what price?

Bid request

Bid response

Bidder
RTB and CM in Privacy Perspective

- Privacy problems in RTB:

  - User data leakage? (RTB Privacy Analysis)
  
  - User profiles can be linked among companies thanks to CM? (CM Privacy Analysis)
  
  - Users are evaluated differently? (Real-Time Price Analysis)

“The first impression seen by a high-value person on the opening page of a major newspaper first thing in the morning has a different value than a user from China who is 12 and has been on the Web all day long playing games,” says Frank Addante, the founder and chief executive of Rubicon.
CM and RTB Detection
CM Detection technique

- CM mechanism:

  User's cookie/id: aaa
  Ad Exchange cm.adexchange.com

  (1) Script or redirect instruction to load
      "http://cm.bidder.com/cmservice?ExchangeUserId=aaa"

  User's cookie: bbb
  Bidder cm.bidder.com

  (2) "http://cm.bidder.com/cmservice?ExchangeUserId=aaa"
      Cookie: UID = bbb

  (3) Matches cookies "aaa" with "bbb"

  ADX's cookie: aaa
  Bidder's cookie: bbb
CM Detection Technique

- **Step 1:** Detect all casual relationships between HTTP requests, i.e. A -> B
  - Based on HTTP header’s Referer & Location, Javascript...

- **Step 2:** Scan all cookies from A’s response and all HTTP parameters included in B to detect cookie matching.
RTB Detection Technique

* Based on **winning price** detection:
  
  * Winning price is notified to the winning bidder by ADX


  * The winning price is detectable in its encrypted form, which contains a timestamp **close to the real time** in the first 8 bytes.
Clear-Text Price Detection

- Many companies use **clear-text prices** in winning price notification
  - http://invitemedia.com?cost=U2VuZCByZWluZm9yY2VtZW50cwHgW5uY5GtM4t&...
  - http://invitemedia.com?cost=0.5&.....

- The values are very often floating-point numbers or integers in micro formats.

- Many parameters are contextual and meaningful: “win_price”, “cost”, “price”, “rtbwinprice”.....
CM and RTB Privacy Analysis
Methodology

• Tools
  • RTBAnalyzer: Firefox plugin implementing CM and RTB detection techniques

• Dataset:
  • 100 volunteers using RTBAnalyzer during several days
  • User data is anonymized
CM Privacy Analysis

Company A

| Cookie 1 | URL1 | URL2 | URL3 | .... |

Company B

| Cookie 2 | URL4 | URL5 | URL6 | .... |

Combine

| Cookie 1 & 2 | URL1 | URL2 | URL3 | URL4 | URL5 | URL6 | .... |

What is the extent of such potential combination?
CM Privacy Analysis

<table>
<thead>
<tr>
<th>Pair of domains</th>
<th>CM frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>facebook - adnxs.com</td>
<td>91%</td>
</tr>
<tr>
<td>turn.com - admeld.com</td>
<td>87%</td>
</tr>
<tr>
<td>doubleclick.net - rfihub.com</td>
<td>86%</td>
</tr>
<tr>
<td>doubleclick.net - adnxs.com</td>
<td>85%</td>
</tr>
<tr>
<td>doubleclick.net - mathtag.com</td>
<td>85%</td>
</tr>
<tr>
<td>adnxs.com - admeld.com</td>
<td>84%</td>
</tr>
<tr>
<td>doubleclick.net - turn.com</td>
<td>80%</td>
</tr>
<tr>
<td>doubleclick.net - yieldmanager.com</td>
<td>77%</td>
</tr>
<tr>
<td>invitemedia.com - admeld.com</td>
<td>73%</td>
</tr>
<tr>
<td>mathtag.com - admeld.com</td>
<td>71%</td>
</tr>
</tbody>
</table>

CM prevalence

Top pairs of domains performing CM the most
CM Privacy Analysis

Pair of domains | Profile combination
---|---
doubleclick.net - adnxs.com | 52.43%
doubleclick.net - yieldmanager.com | 52.01%
facebook.com - adnxs.com | 39.35%
.... | ....
To which extent a user history could be leaked through RTB?
Example of Google’s bid request

id: "Mv\2005\000\017.\001\n\345\177\307X\200M8"
ip: "\314j\310"
user_agent: "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US) AppleWebKit/534.13 (KHTML, like Gecko) Chrome/9.0.597.107 Safari/534.13,gzip"
url: "http://www.example.com/"
detected_language: "en"
detected_vertical {
  id: 22
  weight: 0.67789277
}
detected_vertical {
  id: 355
  weight: 0.32210726
}
cookie_version: 1
googel_user_id: "CAESElcs1pC2TBvb-4SLDjMqsY9"
seller_network_id: 1
publisher_settings_list_id: "\357\237V\206\231\3125\%$\032"
vertical_dictionary_version: 2
timezone_offset: -300
cookie_age_seconds: 7685804
RTB Privacy Analysis

The number of RTB events is about 10% of the number of visited sites.
RTB Privacy Analysis

The leakages in case of these companies are about **11%** of user browsing history on average, but can be as high as **27%** with some user profiles!
Real-Time Price Analysis
Real-Time Price Analysis

- Context dependence:
  - Site, Category, Time, Location...?

- Profile dependence:
  - History category (the category of user visited sites)?
  - Intents (looking for a commercial products)?
Context Analysis

- Dataset: 5K RTB-enabled Alexa sites

- Experiments:
  - Visit sites with empty user profile
  - Different locations: France, the US, and Japan (using Planet Lab)
Prices tend to be highest in the early morning.

<table>
<thead>
<tr>
<th>Time division</th>
<th>The US</th>
<th>France</th>
<th>Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 8h</td>
<td>0.75 (3,246)</td>
<td>0.39 (10,621)</td>
<td>0.28 (729)</td>
</tr>
<tr>
<td>8 - 16h</td>
<td>0.68 (2,772)</td>
<td>0.36 (11,375)</td>
<td>0.22 (732)</td>
</tr>
<tr>
<td>16 - 24h</td>
<td>0.62 (2,520)</td>
<td>0.31 (7,675)</td>
<td>0.19 (516)</td>
</tr>
</tbody>
</table>

Prices in the US are highest in the three analyzed countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>Average</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>The US</td>
<td>0.69</td>
<td>0.15</td>
<td>0.33</td>
<td>1.00</td>
<td>8,538</td>
</tr>
<tr>
<td>France</td>
<td>0.36</td>
<td>0.11</td>
<td>0.24</td>
<td>0.47</td>
<td>29,671</td>
</tr>
<tr>
<td>Japan</td>
<td>0.24</td>
<td>0.04</td>
<td>0.07</td>
<td>0.22</td>
<td>1,977</td>
</tr>
</tbody>
</table>
## Context Analysis

Some site categories are more “worthy” than others.

<table>
<thead>
<tr>
<th>Category</th>
<th>Avg. price</th>
<th>Std</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult / Mature Content</td>
<td>0.25</td>
<td>0.15</td>
<td>0.22</td>
</tr>
<tr>
<td>Humor</td>
<td>0.25</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>Sports</td>
<td>0.29</td>
<td>0.18</td>
<td>0.36</td>
</tr>
<tr>
<td>Games</td>
<td>0.32</td>
<td>0.16</td>
<td>0.42</td>
</tr>
<tr>
<td>Blogs / Web Communications</td>
<td>0.33</td>
<td>0.25</td>
<td>0.32</td>
</tr>
<tr>
<td>Entertainment</td>
<td>0.33</td>
<td>0.23</td>
<td>0.35</td>
</tr>
<tr>
<td>Streaming Media / MP3</td>
<td>0.36</td>
<td>0.27</td>
<td>0.42</td>
</tr>
<tr>
<td>Computers / Internet</td>
<td>0.38</td>
<td>0.24</td>
<td>0.38</td>
</tr>
<tr>
<td>News / Media</td>
<td>0.38</td>
<td>0.26</td>
<td>0.43</td>
</tr>
<tr>
<td>Society / Lifestyle</td>
<td>0.38</td>
<td>0.27</td>
<td>0.46</td>
</tr>
<tr>
<td>Vehicles</td>
<td>0.41</td>
<td>0.34</td>
<td>0.37</td>
</tr>
<tr>
<td>Reference</td>
<td>0.48</td>
<td>0.21</td>
<td>0.61</td>
</tr>
<tr>
<td>Restaurants / Food</td>
<td>0.59</td>
<td>0.31</td>
<td>0.73</td>
</tr>
<tr>
<td>Shopping</td>
<td>0.68</td>
<td>0.38</td>
<td>1.10</td>
</tr>
</tbody>
</table>
Profile Analysis

- **Dataset:**
  - Top Alexa sites belonging to 14 categories *(to build user history category)*
  - Commercial products on sites maty.com, hotels.com, fnac.com *(to build user intents)*

- **Experiment:**

  ![](image)

  *P_{ij}* is a site chosen from Alexa sites or retargeting products

  *S_1, S_2, ..., S_k* are the sites with prices. These sites are the same in all Price collection processes.
Profile Analysis

- Users with history category are evaluated higher than new users.
- Users with specific intents are evaluated higher than users with history category.

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>New user</td>
<td>0.41</td>
<td>0.10</td>
</tr>
<tr>
<td>Only history category</td>
<td>0.58</td>
<td>0.26</td>
</tr>
<tr>
<td>History category + products from fnac.com</td>
<td>0.64</td>
<td>0.29</td>
</tr>
<tr>
<td>History category + products from hotels.com</td>
<td>0.69</td>
<td>0.26</td>
</tr>
<tr>
<td>History category + products from maty.com</td>
<td>1.20</td>
<td>0.25</td>
</tr>
<tr>
<td>Only products from maty.com</td>
<td>1.17</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Users with history category are evaluated higher than new users.

Users with specific intents are evaluated higher than users with history category.
Profile Analysis

- Among different “Only category” profiles, some are more valuable than others: **Games, Sport, Health, Kids and Teens**

<table>
<thead>
<tr>
<th>Category</th>
<th>Avg. price</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult</td>
<td>0.44</td>
<td>0.20</td>
</tr>
<tr>
<td>Arts</td>
<td>0.51</td>
<td>0.17</td>
</tr>
<tr>
<td>Business</td>
<td>0.55</td>
<td>0.22</td>
</tr>
<tr>
<td>Business - Financial Services</td>
<td>0.59</td>
<td>0.20</td>
</tr>
<tr>
<td>Computers</td>
<td>0.48</td>
<td>0.21</td>
</tr>
<tr>
<td>Games</td>
<td>0.80</td>
<td>0.35</td>
</tr>
<tr>
<td>Health</td>
<td>0.67</td>
<td>0.47</td>
</tr>
<tr>
<td>Home</td>
<td>0.58</td>
<td>0.21</td>
</tr>
<tr>
<td>Kids and Teens</td>
<td>0.64</td>
<td>0.33</td>
</tr>
<tr>
<td>News</td>
<td>0.50</td>
<td>0.12</td>
</tr>
<tr>
<td>Recreation</td>
<td>0.55</td>
<td>0.21</td>
</tr>
<tr>
<td>Science</td>
<td>0.50</td>
<td>0.19</td>
</tr>
<tr>
<td>Shopping</td>
<td>0.53</td>
<td>0.22</td>
</tr>
<tr>
<td>Sports</td>
<td>0.71</td>
<td>0.47</td>
</tr>
</tbody>
</table>
Conclusion

- CM is prevalent and might help companies significantly increase the size of their tracked user profiles.

- RTB can leak up to 27% of a user history to a bidder in auction

- Each user is being evaluated differently by bidders depending on their profile and visiting contexts.

- User’s presence on a website is often sold off for less than $0.0005 (an equivalence of $0.5 CPM), much less than that from users’ perspective (EUR 7)
Thank you for your attention!