“Your Hashed IP Address: Ubuntu”
Perspectives on Transparency Tools for Online Advertising

Tobias Urban¹,², Martin Degeling², Thorsten Holz², and Norbert Pohlmann¹

¹ Institute for Internet Security – if(is); Westphalian University of Applied Sciences
² Horst Görtz Institute for IT Security; Ruhr University Bochum
INTRODUCTION
→ How Does Online Advertising Work?

ID: ABCD-1234

Segment
Location: Puerto Rico
Sex: Male
Age: 26-36
Interest: Football
Interest: Graph Theorie

White House Budget Official Said 2 Aides Resigned Amid Ukraine Aid Freeze

Mark Sandy, an official at the Office of Management and Budget, testified that two of his colleagues quit after expressing concerns about President Trump's decision to withhold military assistance.
Hi, my name is John. I am curious about personal data collected about me by ad-tech companies.

Research Question:
How beneficial is data provided to users by ad companies?
Research Question: What are the challenges companies face when they design a privacy concept?

Hi, my name is Clara. I am trying to build a GDPR-proof privacy concept for my company.
PERCEPTION OF TRANSPARENCY TOOLS

→ Method

Actually, I am the manifestation of an MTurk Survey (n=490).

<table>
<thead>
<tr>
<th></th>
<th>Amount</th>
<th>%</th>
<th>US pop.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>264</td>
<td>54%</td>
<td>49%</td>
</tr>
<tr>
<td>Female</td>
<td>224</td>
<td>46%</td>
<td>51%</td>
</tr>
<tr>
<td>Non-binary</td>
<td>1</td>
<td>0%</td>
<td>–</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–24</td>
<td>41</td>
<td>8%</td>
<td>16%</td>
</tr>
<tr>
<td>25–34</td>
<td>218</td>
<td>46%</td>
<td>22%</td>
</tr>
<tr>
<td>35–44</td>
<td>112</td>
<td>23%</td>
<td>20%</td>
</tr>
<tr>
<td>45–54</td>
<td>66</td>
<td>14%</td>
<td>21%</td>
</tr>
<tr>
<td>55–64</td>
<td>40</td>
<td>8%</td>
<td>21%</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>1</td>
<td>0%</td>
<td>12%</td>
</tr>
<tr>
<td>High School</td>
<td>161</td>
<td>33%</td>
<td>51%</td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>255</td>
<td>52%</td>
<td>18%</td>
</tr>
<tr>
<td>Pro./Master’s/Ph.D.</td>
<td>69</td>
<td>14%</td>
<td>11%</td>
</tr>
</tbody>
</table>

Conducted in February 2019
PERCEPTION OF TRANSPARENCY TOOLS → Attitude Towards Online Advertising

“ I honestly expect some ad companies to illegally collect my facial expressions and sounds in my environment through cameras and microphones. I always expect them to access other apps and histories of everything that I do ”

• Participants have a neutral view on ads

• 90% use some sort of privacy tool

• Participants think they understand online advertisement but most don’t.

• 65% do not trust companies to provide all collected data upon request
PERCEPTION OF TRANSPARENCY TOOLS

→ Identifying Data Collectors
I struggle to identify the ad network that provides the ad.

43% answered: “Outnet.com”
24% answered “reddit.com”
24% answered Google (correct)
PERCEPTION OF TRANSPARENCY TOOLS

Identifying Data Collectors

However, if the ad network’s name is present near the ad I have less problems to identify the network.

28% name the advertised company
17% answered “espn.com”
48% answered “Outbrain” (correct)
Once I identified the companies that collected my personal data, I want to request a copy of my data to understand what companies collected and how they use it.
Companies provide 3 types of data

- Technical (raw) data

- Tracking data

- Segment data
PERCEPTION OF TRANSPARENCY TOOLS

Assessment of data: Technical (Raw) Data

“ This is gibberish to me…”

“I can not think of anyway to use this information.”

<table>
<thead>
<tr>
<th>time</th>
<th>u_ip</th>
<th>ua</th>
<th>kvClob</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.09.18 09:29:25 UTC</td>
<td>127.0.0.0</td>
<td>Mozilla/5.0 (X11; Ubuntu; cs=;fst=;c=&lt;Country&gt;;st=&lt;state&gt;;</td>
<td></td>
</tr>
<tr>
<td>24.09.18 09:28:04 UTC</td>
<td>127.0.0.0</td>
<td>Mozilla/5.0 (X11; Ubuntu; cs=;fst=;ip=127.0.0.0/24;hb=1;de</td>
<td></td>
</tr>
<tr>
<td>24.09.18 09:24:19 UTC</td>
<td>127.0.0.0</td>
<td>Mozilla/5.0 (X11; Ubuntu; cs=;fst=;c=&lt;Country&gt;;st=&lt;state&gt;;</td>
<td></td>
</tr>
<tr>
<td>24.09.18 09:24:18 UTC</td>
<td></td>
<td>Mozilla/5.0 (X11; Ubuntu; cs=;DE;st=&lt;state&gt;;dma=&lt;metroKey&gt;</td>
<td></td>
</tr>
<tr>
<td>24.09.18 09:17:29 UTC</td>
<td>127.0.0.0</td>
<td>Mozilla/5.0 (X11; Ubuntu; cs=;fst=;c=&lt;Country&gt;;st=&lt;state&gt;;</td>
<td></td>
</tr>
<tr>
<td>24.09.18 09:16:19 UTC</td>
<td>127.0.0.0</td>
<td>Mozilla/5.0 (X11; Ubuntu; cs=;fst=;c=&lt;Country&gt;;st=&lt;state&gt;;</td>
<td></td>
</tr>
<tr>
<td>24.09.18 09:16:17 UTC</td>
<td>127.0.0.0</td>
<td>Mozilla/5.0 (X11; Ubuntu; cs=;fst=;c=&lt;Country&gt;;st=&lt;state&gt;;</td>
<td></td>
</tr>
<tr>
<td>24.09.18 09:10:44 UTC</td>
<td>127.0.0.0</td>
<td>Mozilla/5.0 (X11; Ubuntu; cs=;fst=;c=&lt;Country&gt;;st=&lt;state&gt;;</td>
<td></td>
</tr>
</tbody>
</table>
PERCEPTION OF TRANSPARENCY TOOLS

→ Assessment of data: Tracking Data

“I would use this if I wouldn't want to be tracked on certain sites. Or use an incognito browser for certain sites.”

“They can track me all they want ;) I have browser history that does the same thing.”

<table>
<thead>
<tr>
<th>Last seen on</th>
<th>Site</th>
<th>IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-09-24 08:51:01</td>
<td><a href="https://www.01net.com/">https://www.01net.com/</a></td>
<td>127.0.0.0</td>
</tr>
<tr>
<td>2018-09-24 08:53:11</td>
<td><a href="https://ad3.adserver01.de/www/delivery/af">https://ad3.adserver01.de/www/delivery/af</a></td>
<td>127.0.0.0</td>
</tr>
<tr>
<td>2018-09-24 09:05:06</td>
<td><a href="https://www.farfetchplus.com/">https://www.farfetchplus.com/</a></td>
<td>127.0.0.0</td>
</tr>
<tr>
<td>2018-09-24 09:15:50</td>
<td><a href="https://www.wetter.de/">https://www.wetter.de/</a></td>
<td>127.0.0.0</td>
</tr>
<tr>
<td>2018-09-24 09:16:31</td>
<td><a href="https://www.pistonheads.com/">https://www.pistonheads.com/</a></td>
<td>127.0.0.0</td>
</tr>
<tr>
<td>2018-09-24 09:18:59</td>
<td><a href="https://ad6.ad-srv.net/request_content.php">https://ad6.ad-srv.net/request_content.php</a>?</td>
<td>127.0.0.0</td>
</tr>
</tbody>
</table>
PERCEPTION OF TRANSPARENCY TOOLS

Assessment of data: Segment Data

“"I would use this in court if I feel they have too much of my information without my consent."”

“I already know my own tastes and preferences better than any website.”

<table>
<thead>
<tr>
<th>Category</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>nightwish</td>
<td></td>
</tr>
<tr>
<td>jay-z</td>
<td></td>
</tr>
<tr>
<td>brooklyn</td>
<td></td>
</tr>
<tr>
<td>diamond (jewellery)</td>
<td></td>
</tr>
<tr>
<td>power metal</td>
<td></td>
</tr>
<tr>
<td>symphonic metal and opera heavy</td>
<td></td>
</tr>
</tbody>
</table>
PERCEPTION OF TRANSPARENCY TOOLS

→ Evaluation of Profiles

Generally, I understand ‘Interest Data’ and think it is helpful.

‘Technical’ and ‘Tracking’ data is less helpful, tough.
I think that “segment” data is most useful to understand privacy implications of a company.
However, I prefer to receive “tracking” data.

Participants’ ranking of profiles.
Before, you simply blame the “bad” ad companies, please give me an opportunity to explain the situation from my point of view.
Conducted in the first quarter of 2019

I am the manifestation of an online survey (n=24) and of eight in-person interviews.

- Invitation via “bulk” email (4% participated)
- Companies of all sizes participated (some with more than 1k employees)
- Participants are DPOs, Consultants, or “C”-Level
- Company come from different layers in the ad ecosystem
BUSINESS PERSPECTIVE
→ Impact of the GDPR

“So I represented advertising in the GDPR groups, so we kind of created a group of champions, if you want, so I was the champion on advertising, and it proved that advertising was the most challenging one”

• Compliance is harder to achieve in online advertisement than in other areas

• The GDPR became a marketing point

• Some companies claimed that they were well prepared while others were overwhelmed by the new regulation

“[...] suddenly, any identifier [sic.] would become PII. [...]”
BUSINESS PERSPECTIVE

→ Subject Access Requests

I expected a “five-digit figure of requests per month.” But we only got 10 in a whole year.

- Half use a standardized process and the other half handles each request individually
- 92% of companies got the expected amount or less requests
- 62% of companies expressed that they want unified guidelines how access requests should be handled
Authentication of individuals creates more sensitive data.

Identification of users
- several different forms of identification
  - screenshots, affidavits, cookie-id, …
BUSINESS PERSPECTIVE

→ Challenges

Identifying data flows
• Understanding external and internal data flows
• Understanding if consent was given

“...because I’m quite sure that they do not know how much data they have. I’m positive on this.”
BUSINESS PERSPECTIVE
→ Opinions

“[The GDPR] is not useless, but users—and that’s why they are not contacting us—that do not want to be tracked already have an app [to block tracking]”

• High fines brought more attention to privacy

• The GDPR only provides few benefits to users

• One company was puzzled by the distrust of users while others agreed that companies probably do not share all data upon request
CONCLUSION

Initially, we asked ourselves:

**How beneficial is data provided to users by ad companies?**

- Users struggle to identify companies collecting data
- Displaying high level information ("segment data") is preferred by users
- Users do not trust ad companies

**What are the main challenges companies face when designing a privacy concept?**

- (Small) companies want more regulation
- Compliance has higher priority than transparency
- Right to Access is less popular than expected
“Your Hashed IP Address: Ubuntu”
Perspectives on Transparency Tools for Online Advertising

Thank you for your attention!
Are there questions?

Tobias Urban
✉️ urban (at) internet-sicherheit [dot] de
🐦 @_tobiasu