Measuring Online Behavioural Advertising

A tale of Transparency & Human-Centric Economics

Nikolaos Laoutaris

Research Professor
IMDEA Networks Institute
I am not a secretive person
I am not a privacy nut
real name

Muhammad Ali

pose

silly comment

Exploding cigar! Really? Did they buy it from Coyote Wilie?

10 Ways the CIA Tried to Kill Castro

There are only so many different ways you can ambush someone with a sharpshooter, so some of the ways the CIA plotted to kill Castro were pretty wild.
About Me

I am a research professor at IMDEA Networks Institute in Madrid. Prior to that I was director of data science at Eurecat and chief scientist of the Data Transparency Lab which I co-founded in 2014, during my 10 year tenure as a researcher and senior researcher of Telefonica Research in Barcelona. Before Telefonica, I was a postdoc fellow at Harvard University and Marie Curie postdoc fellow at Boston University. I got my PhD in computer science from the University of Athens in 2004. My interests include: privacy/transparency/data protection, economics of networks and information, intelligent transportation, distributed systems, protocols, and network measurements.
in 2012 I gave up on all my previous research
to work exclusively on privacy
the web economy could collapse
due to **Tragedy of the commons** around privacy

Internet company in
Web Economy … crossing privacy **red lines**

The “commons”: consumer trust on the web and it’s business models
Big Idea #1 - Obvious in retrospect

The importance of Transparency (Software)

“Sunlight is the best disinfectant.”
-Louis Brandeis

“Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman.”

A first of its kind Transparency Tool

Sheriff
Detecting Price Discrimination

1. Select price

2. Check it

3. Examine differences

Jakub Mikians
UPC (now Amazon)

Kostas Iordanou
Telefonica-UC3M
**Canon EOS 7D Mark II Body**

Fuel Your Creative Passion

- 20.2 MP APS-C CMOS sensor
- 3.0 inch (1040k) Clear View II LCD monitor

**Price:** €1,229.01

In Stock (Ships within 24 hours)

**Add to Basket**

### All Prices Results

<table>
<thead>
<tr>
<th>Variant</th>
<th>Converted Value</th>
<th>Original Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada, Ontario</td>
<td>€1,409.60</td>
<td>€2,049.00</td>
</tr>
<tr>
<td>Czech Republic, Praha</td>
<td>€1,395.52</td>
<td>US$ 1,558.80</td>
</tr>
<tr>
<td>France, Champagne-ardenne, Troyes</td>
<td>€1,249.84</td>
<td>€1,249.84</td>
</tr>
<tr>
<td>Ireland, Westmeath</td>
<td>€1,249.84</td>
<td>€1,249.84</td>
</tr>
<tr>
<td>Portugal, Praga</td>
<td>€1,249.84</td>
<td>€1,249.84</td>
</tr>
<tr>
<td>You</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
<tr>
<td>Windows 7, Chrome, Spain</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
<tr>
<td>Mac OS, Safari, Spain</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
<tr>
<td>Linux, Firefox, Spain</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
<tr>
<td>United States, Washington</td>
<td>€1,162.94</td>
<td>US$ 1,299.00</td>
</tr>
<tr>
<td>Poland, Warsaw</td>
<td>€1,430.41</td>
<td>US$ 1,597.77</td>
</tr>
<tr>
<td>Singapore, Singapore</td>
<td>€1,250.16</td>
<td>US$ 1,396.43</td>
</tr>
<tr>
<td>Australia, Clayton</td>
<td>€1,175.79*</td>
<td>AU$ 1,739.00</td>
</tr>
<tr>
<td>Brazil</td>
<td>€1,453.67</td>
<td>US$ 1,623.75</td>
</tr>
</tbody>
</table>

### Results from local users

<table>
<thead>
<tr>
<th>Source ID</th>
<th>Converted Value</th>
<th>Original Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local User 0</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
<tr>
<td>Local User 1</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
<tr>
<td>Local User 2</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
<tr>
<td>Local User 3</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
<tr>
<td>Local User 4</td>
<td>€1,229.01</td>
<td>€1,229.01</td>
</tr>
</tbody>
</table>
Which retailers?
Which products?
Which countries?
Detection of personalized PD

1. Collect some info

2. Do P2P checks
Solving the profile pollution problem

Doppelgänger

From Wikipedia, the free encyclopedia

For other uses, see Doppelgänger (disambiguation).

A doppelgänger (ˈdɒpəl ˈgænər or ˈdʒænər; German: [ˈdɔpl ˈɡeːnɐ] (listen), literally "double-goer") is a look-alike or double of a living person, sometimes portrayed as a ghostly or paranormal phenomenon and usually seen as a harbinger of bad luck. Other traditions and stories equate a doppelgänger with an evil twin. In modern times, the term twin stranger is occasionally used.

[1][2]
Enough with PD ... let's get to tracking & advertising

What we search – The sites we visit – Who we befriend – What we buy ... everything is tracked
How can you tell if an ad is targeted?
This goes beyond curiosity ...
First approach

Detection via content-based analysis
Main idea of our methodology

CLEAN (PhantomJS)

PROFILE/BROWSER

NEUTRAL WEBPAGE

WEBPAGES

CoNEXT 2015
How frequent is OBA?
Are some personas more targeted than others?

In summary,
• TTK measures if OBA is happening
• BAiLP captures what percentage is due to OBA.

CORRELATION → received OBA (BAiLP) and its value for the advertisers
Is OBA applied to sensitive topics?

- Same methodology → 21 sensitive personas
Limitations of content-based analysis detection

- Slow
- Not scalable
- Intrusive
- Cannot detect implicit targeting
Count-based detection & crowdsourcing
Targeted ads follow you around

• Detection via simple counting
• No need for content analysis
• No need to inject traffic
• Real-time
• No prob with indirect targeting
Detects & annotates all rendered ads
Algorithm 1: The count-based algorithm for ad $\alpha$ seen by user $u$

**Require:**

**Counters:**
- $\#\text{Users}_\alpha$ → Number of other users that observe ad $\alpha$
- $\#\text{Domains}_{u,\alpha}$ → Number of domains that user $u$ observe ad $\alpha$

**Thresholds:**
- $\text{Users}_{th}$ → Users threshold based on all users
- $\text{Domains}_{th,u}$ → Domains threshold for a specific user $u$

1. if $\#\text{Users}_\alpha \leq \text{Users}_{th}$ AND $\#\text{Domains}_{u,\alpha} \geq \text{Domains}_{th,u}$ then
2. Targeted ad
3. else
4. Non-targeted ad
Works pretty fine

Figure 3: False Negatives % Vs. Frequency Cap using two different thresholds (Mean, Mean+Median) for both variables ($\#\text{Users}_a$, $\#\text{Domains}_{u,a}$)
Crowd-sourced Users statistics

Similarities with other users

Per Country Ad Impressions

adwords cloud
Launch of Data Transparency Lab

A community of technologists, researchers, policymakers and industry representatives working to advance online personal data transparency through scientific research and design.

Kick-off Workshop : Nov’14, Bcn
Building a community
18 grants & lots of collaborations
DATA TRANSPARENCY LAB - EXAMPLE 1: FACEBOOK DATA VALUATION TOOL
Permits users to estimate how much money Facebook is making on them

INFORMS INDIVIDUAL USERS

A plugin for your browser that combines your online activity with Facebook’s Public APIs to estimate your advertising value

REVEALS SOME MORE GENERAL TRENDS

Demonstrates how factors like country, status, studies, etc. impact on a user’s advertising value

DEMO VIDEO: https://youtu.be/QPfc-gXGdjI

LIVE DEMO
https://acrumin.cartodb.com/viz/75d6d052-0648-11e6-8923-0e3ff518bd15/public_map
A CENSUS THAT IDENTIFIES WEBSITES THAT TRACK USERS TO SHED LIGHT ABOUT HOW TRACKING IS USED AND BY WHOM
Big Idea #2 - NOT obvious even in retrospect

Most problems of the web are due to its broken economics model
You and online services

• Free data
• For free service
  – payment “in kind”
  – no cash
BAD for privacy!

(tracking is cheap) + (no payments to users) = (collect ... everything)

Moore's Law - The number of transistors and resistors on a chip doubles every 24 months
— Gordon Moore

You Online Service

<table>
<thead>
<tr>
<th>data</th>
<th>maps email search</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td>Online Service</td>
</tr>
</tbody>
</table>

ALL YOU CAN EAT
BUFFET

ALL YOU CAN EAT
BUFFET

SPAM

SPAM

SPAM
BAD for sustainability!
A Human-Centric Data Economy

- explicit monetary compensation for data based on their value for online services
  - e-commerce
  - media streaming
  - location services
  - ...

[Image: A human-centric data economy illustration showing data exchange between a person and an online service, with a monetary transaction symbol]
Tons of great questions to ask
### HCDE vs. Data Marketplaces 1/4

<table>
<thead>
<tr>
<th>DMs</th>
<th>Pricing Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td>Ad hoc</td>
</tr>
<tr>
<td>Personal Data</td>
<td>Auction-based</td>
</tr>
</tbody>
</table>

Source: [https://about.datarade.ai](https://about.datarade.ai)
HCDE vs. Data Marketplaces 2/4

New commodities go through periods of very high volatility

-- Until people learn to trade

-- Oil 19th century, cryptocurrency now
HCDE vs. Data Marketplaces 3/4

● Data is **not** really a commodity
● Two different liters of oil are almost identical
● But what about
  -- Browsing behavior famous investor vs. average person?
  -- Mobility data from a taxi driver vs. a weekend driver?
  -- Shopping cart of a teenager vs. middle-ager?
HCDE vs. Data Marketplaces 4/4

Digital goods / Information
-- Don’t decay
-- Cost to copy = 0
-- Data provenance = hard

Without Clear Context
-- Hard to price
-- Hard to auction

X euros
-- Infinite days
-- Infinite Km
-- Infinite drivers

* Consumption 0 lt per 100Km
* Service every infinite years
To probe further
Almost done
people don’t care about privacy
(some say)
some other things people didn’t care about
flight security
kids playing with melted glass
kids playing with melted iron
kids playing with power tools

still available at ebay
kerosene train

haven’t located one yet
societies evolve
Won’t be long before we look back and shake our head.
Thank you!