Tracking and Targeting: State-of-the-art Advertisements

Erik Sy
About me

Research on security, privacy, online tracking, and geolocation methods

2016 Master of Physics
University of Rostock

Since 2016 researcher
University of Hamburg

Coordinator of the research project AppPETs
Outline

- Extend of the data collection
- Motivation for data collection
- Ad exchanges and real-time bidding
- Targeting methods
- Tracking methods
Extend of the data collection: browsing history

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Most Recent Visit</th>
<th>Visit Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>eBay Kleinanzeigen</td>
<td>Kostenlos. Einfach, Lokal. Anzeigen gratis inserieren...</td>
<td><a href="https://www.ebay-kleinanzeigen.de/">https://www.ebay-kleinanzeigen.de/</a></td>
<td>11:17 AM</td>
</tr>
<tr>
<td>Twitter. That's what's happening...</td>
<td><a href="https://twitter.com/">https://twitter.com/</a></td>
<td>11:17 AM</td>
<td>1</td>
</tr>
<tr>
<td>10 News, E-Mailing und Shopping bei online.de</td>
<td></td>
<td><a href="https://www.ebay.de/">https://www.ebay.de/</a></td>
<td>11:18 AM</td>
</tr>
<tr>
<td>WEB.DE - E-Mail-Adresse kostenlos, Free-mail, De-Mail &amp; Nachrichten</td>
<td><a href="http://www.web.de/">http://www.web.de/</a></td>
<td>11:16 AM</td>
<td>1</td>
</tr>
<tr>
<td>Wikipedia</td>
<td></td>
<td><a href="https://www.wikipedia.org/">https://www.wikipedia.org/</a></td>
<td>11:16 AM</td>
</tr>
<tr>
<td>Wikipedia.org</td>
<td></td>
<td></td>
<td>11:16 AM</td>
</tr>
<tr>
<td>Amazon.de: Günstige Preise für Elektronik &amp; Foto, Filme, Musik, Büch...</td>
<td><a href="http://www.amazon.de/">http://www.amazon.de/</a></td>
<td>11:18 AM</td>
<td>1</td>
</tr>
<tr>
<td>Electronic, Autos, Mode, Sammervideos, Möbel und mehr Online-Shopp...</td>
<td></td>
<td>11:18 AM</td>
<td>1</td>
</tr>
<tr>
<td>Facebook - Log In or Sign Up</td>
<td></td>
<td><a href="https://www.facebook.com/">https://www.facebook.com/</a></td>
<td>11:15 AM</td>
</tr>
<tr>
<td>Facebook.com</td>
<td></td>
<td></td>
<td>11:15 AM</td>
</tr>
<tr>
<td>YouTube</td>
<td></td>
<td></td>
<td>11:15 AM</td>
</tr>
<tr>
<td>YouTube.com</td>
<td></td>
<td></td>
<td>11:15 AM</td>
</tr>
<tr>
<td>heise online - IT-News, Nachrichten und Hintergründe</td>
<td></td>
<td></td>
<td>11:16 AM</td>
</tr>
<tr>
<td>Wetter - Wettervorhersage - Wetterbericht - wetter.de</td>
<td></td>
<td></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>Aktuelle Nachrichten - Inland Ausland Wirtschaft Kultur Sport - ARD Tag...</td>
<td></td>
<td></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>Startseite - ZDFmediathek</td>
<td></td>
<td></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>SPIEGEL ONLINE - Aktuelle Nachrichten</td>
<td></td>
<td></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>Google</td>
<td></td>
<td></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>UHH - Universität Hamburg</td>
<td></td>
<td></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>taz.de - taz.de</td>
<td></td>
<td></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>ZEIT ONLINE</td>
<td>Nachrichten, Hintergründe und Debatten</td>
<td><a href="http://www.zeit.de/index">http://www.zeit.de/index</a></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>The world's leading software development platform - GitHub</td>
<td></td>
<td></td>
<td>11:14 AM</td>
</tr>
<tr>
<td>NZZ News - Aktuelle Nachrichten - Neue Zürcher Zeitung</td>
<td></td>
<td></td>
<td>11:13 AM</td>
</tr>
<tr>
<td>ARTE in English</td>
<td></td>
<td></td>
<td>11:13 AM</td>
</tr>
<tr>
<td>Yahoo</td>
<td></td>
<td></td>
<td>11:13 AM</td>
</tr>
<tr>
<td>Trending stories on Indian Lifestyle, Culture, Relationships, Food, Travel...</td>
<td></td>
<td></td>
<td>11:13 AM</td>
</tr>
<tr>
<td>News, sport and opinion from the Guardian’s global edition</td>
<td>The Guardian</td>
<td></td>
<td>11:12 AM</td>
</tr>
<tr>
<td>reddit: the front page of the Internet</td>
<td></td>
<td></td>
<td>11:12 AM</td>
</tr>
</tbody>
</table>
Extend of the data collection: web searches

Google search for "how long are web searches stored"

About 42,700,000 results (0.36 seconds)

How long does Google retain my web history? - Quora
2: Server logs which log all searches and may be stored indefinitely. From what I gather, these are logged without Gmail account details but do include IP addresses. IPs were anonymized after 9 months from 2007/8. It is now unclear whether this is the case.

Sweeping Away a Search History - The New York Times
Apr 3, 2014 - Unfortunately, your web searches are carefully tracked and saved in ... "Nobody understands the long-term impact of this data collection," said ...

How much of my search history could be recovered? - Ask Leo!
https://askleo.com/how_much_of_my_search_history_could_be_recovered/
Jul 30, 2011 - Does CCleaner or even manually erasing history actually remove the history from the hard drive? ... Unless you have spyware installed on your computer, "every bloody keystroke" is not being recorded. ... A file deleted to the Recycle Bin can be recovered from that very simply, but I ...
Extend of the data collection: spendings

- Enterprises share transaction data for conversion tracking
Extend the data collection: location and social data

Turn on Location History?

Location History helps you get useful information, such as commute predictions, improved search results, and more useful ads on and off Google. It does this by creating a private map of where you go with your signed-in devices.

To create this map, Google regularly obtains location data from devices for which you’ve turned on Location History. This data is collected even when you aren’t using a specific Google product. Learn more

Remember, location data may be saved from any signed-in devices for which you’ve turned on Location History. You can always control and review your activity at My Account.

CANCEL    TURN ON
The data broker Acxiom

- data sets on approx. 700 million consumers
- over 3000 propensities for U.S. consumer

Risk of data leakage!

1 Acxiom Corporation Annual Report 2014
Motivation for data collection

Evidence for effectivity

Evidence for efficiency

Detection of click fraud

Example for a campaign in Google AdWords

<table>
<thead>
<tr>
<th></th>
<th>Budget</th>
<th>Cost</th>
<th>Impressions</th>
<th>Clicks</th>
<th>Conver.</th>
<th>Ø Cost per click</th>
<th>Ø Cost per conver.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 €/ day</td>
<td>120,02 €</td>
<td>7,149</td>
<td>246</td>
<td>140</td>
<td>0,49 €</td>
<td>0,86 €</td>
</tr>
<tr>
<td>2</td>
<td>5 €/ day</td>
<td>11,57 €</td>
<td>1,250</td>
<td>47</td>
<td>23</td>
<td>0,25 €</td>
<td>0,50 €</td>
</tr>
</tbody>
</table>

Validation of statistics difficult for advertiser
Ad exchanges and real-time bidding

1. Advertising space
   - User
   - Publisher
   - Supply-side platform
   
2. Information flow
   - Ad exchange

3. Auction
   - Demand-side platform
   - Advertiser

4. Advertisement
Behavioral targeting

- Intent to increase the conversions-rate
- Statistical methods such as A/B-test lead to an identification of relevant characteristics
- More precise user profiles allow to test more characteristics
- Higher bids, if a higher conversion-rate is assumed

Example:
Correlation between gender and conversions-rate?

Test group 1

Test group 2
Email addresses are used to define the audience of an ad campaign (see AdWords Customer Match).

User 1. visits YouTube Publisher 2. Advertisement Advertiser defines audience by their email addresses
Online tracking

Customer data are shared with trackers
### Tracking based on cookies

The following cookies are stored on your computer:

<table>
<thead>
<tr>
<th>Site</th>
<th>Cookie Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>github.com</td>
<td></td>
</tr>
<tr>
<td>go.sonobi.com</td>
<td></td>
</tr>
<tr>
<td>google.com</td>
<td></td>
</tr>
<tr>
<td>google.com</td>
<td>NID</td>
</tr>
<tr>
<td>google.com</td>
<td>NID</td>
</tr>
<tr>
<td>google.de</td>
<td>NID</td>
</tr>
<tr>
<td>google.de</td>
<td>CONSENT</td>
</tr>
<tr>
<td>googlefeedsyndication.com</td>
<td></td>
</tr>
<tr>
<td>qualist.com</td>
<td></td>
</tr>
</tbody>
</table>

**Name:** NID

**Content:** 102-JM3GgFyb_bDQERHVm2ZxItlHgrGQPXzrhAxlJe39jyfMucFbiasXNYSk82qcAhWGXNO9gH-Mmat

**Domain:** .google.de

**Path:** /

**Send For:** Any type of connection

**Expires:** October 27, 2017, 4:01:50 PM
„Here are some examples of the device information we collect:”¹
Cross-device and cross-user tracking
Behavior-based tracking

### 8th June

- Website visits:
  - google.de: 7
  - tagesschau.de: 5
  - sparkasse.de: 1
  - wetter.de: 1
  - bauernzeitung.de: 5

### 9th June

- Website visits:
  - google.de: 4
  - tagesschau.de: 7
  - sparkasse.de: 1
  - wetter.de: 2
  - bauernzeitung.de: 3

### Experimental results

- Up to 85% of all browsing sessions can be linked to specific users\(^1\)

---

Tracking header

HTTP-request of a user

User profiles

Third-parties such as VPN services, ISP, ...

Tracking Header

Online service

http://www.example.com
Offline tracking
Offline tracking: Calls

- **Online Campaign 1**: 040123...
- **Online Campaign 2**: 040456...
- **Flyer Campaign**: 040789...

- User
- Multiple marketing campaigns
- Unique trackable phone numbers
- Advertiser such as locksmith or delivery service

- Call statistic allows an campaign-level performance analysis
Offline tracking: store visits

1. Online advertisement
   - Google AdWords
   - Facebook for business

2. Report of store visits
   - Data broker

User (location tracked)
- Data broker tracks geolocation of user
- Data broker matches store visits with displayed ads

Advertiser with local store
Offline tracking: store sales

1. Online advertisement

2. Analysis of customer transactions

- Data broker matches customer transactions with displayed ads
Conclusion

- Precise user profiles increase the performance of advertisements
- Selection of single users by identifiers such as email addresses
- Tracking arrives in local stores
Universität Hamburg
Fachbereich Informatik
Arbeitsbereich Security and Privacy
Erik Sy
Vogt-Kölln-Straße 30
D-22527 Hamburg

E-Mail sy@informatik.uni-hamburg.de