

Privacy Risks with Facebook's PII-based Targeting

Auditing a data broker's advertising interface

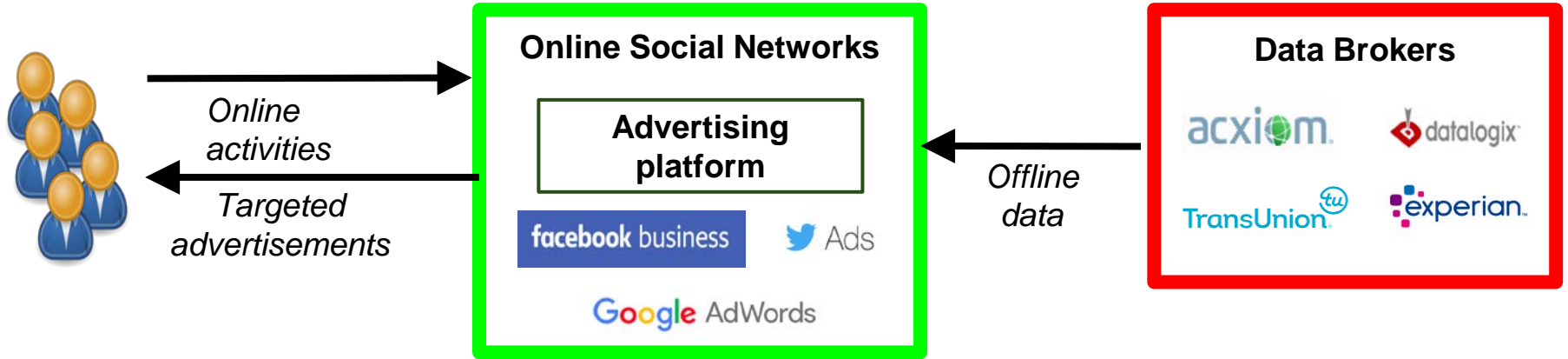
Giridhari Venkatadri¹, Athanasios Andreou², Yabing Liu¹, Alan Mislove¹,
Krishna P. Gummadi³, Patrick Loiseau^{3,4}, Oana Goga⁴

*¹Northeastern University, ²EURECOM, ³MPI-SWS
⁴Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG*

Traditional Data Brokers



21st Century Data Brokers



Attribute-based Targeting

1. Attribute-based targeting

Advertiser specifies attributes to create *audience* of users

2. Audience size

Facebook provides *obfuscated* number of selected users

3. Custom (PII) audiences

Advertiser can select via *PII of target users instead*

The screenshot displays the Facebook Audience Definition interface. At the top, there are two tabs: "Create New" (selected) and "Use a Saved Audience". Below this, the "Custom Audiences" section is highlighted with a yellow border, showing a custom audience named "Customer List" with "list_of_9444_records" and an option to "Add Custom Audiences or Lookalike Audiences". The "Locations" section is highlighted with a red border, showing "Everyone in this location" with an "Include" dropdown and "Add locations" button. Below that, the "Age" section is set to "18 - 65+", "Gender" is set to "Men", and "Languages" has a text input field. The "Detailed Targeting" section is also highlighted with a red border, showing "INCLUDE people who match at least ONE of the following" with a text input field and "Suggestions" and "Browse" buttons. On the right side, the "Audience Definition" section shows a gauge indicating the audience is defined, with "Specific" on the left and "Broad" on the right. Below this, the "Audience Details" section lists: "Custom Audience: list_of_9444_records", "Age: 18 - 65+", "Gender: Male", and "Placements: Facebook Feeds and Instagram Feed". A green box highlights the "Potential Reach: 2,600 people" section. Below that, the "Estimated Daily Reach" section shows "520 - 1,400 people on Facebook" and "270 - 530 people on Instagram".

Custom Audiences

The screenshot displays the Facebook Custom Audiences creation interface. At the top, there are tabs for 'Create New' and 'Use a Saved Audience'. The main area shows a list of custom audiences under the heading 'Custom Audiences'. One audience, 'Customer List', is highlighted with a blue border and contains two entries: 'list_of_40k_records' and 'list_of_9444_records'. Below this list, there are options to 'Exclude' and 'Create New'. The 'Locations' section is set to 'Everyone in this location' with 'United States' selected. The 'Age' range is set to '18 - 65+' and 'Gender' is set to 'All'. On the right side, a summary panel shows 'Audience Size' with a gauge indicating the audience is defined, 'Potential Reach: 11,000 people', and 'Estimated Daily Results Reach' of '540 - 1,000'. A disclaimer note explains that estimates are based on past campaign data and budget.

1. **Select fields to use**

Advertiser specifies which fields they have on users

2. **Upload CSV file with user data**

Advertiser uploads file to Facebook

3. **Facebook matches users**

Advertiser is provided with an audience to advertise to

4. **Facebook provides statistics**

Advertiser can obtain obfuscated size for audience combinations

What PII can be Used?

Site	Name	Email	Phone number	City or ZIP	State or Province	Birthday, Gender	Employer	Site user ID	Mobile advertiser ID
Facebook	✓	✓	✓	✓	✓	✓	✗	✓	✓
Instagram	✓	✓	✓	✓	✓	✓	✗	✓	✓
Twitter	✗	✓	✓	✗	✗	✗	✗	✓	✓
Google	✓	✓	✓	✓	✗	✗	✗	✓	✓
Pinterest	✗	✓	✗	✗	✗	✗	✗	✗	✓
LinkedIn	✗	✓	✗	✗	✗	✗	✓	✗	✓

Raises Concerns

Sites have **detailed data** from service, data brokers

First time allow anyone to **link against that database**

Allow advertisers to upload data, obtain aggregate information

Question: Are attacks that leak user information possible?

Yes! With a few insights...

Insight 1: Overcoming Obfuscation

The image shows a Facebook Audience Definition interface with several overlaid boxes and arrows. The interface includes sections for Custom Audiences, Locations, Age, Gender, Languages, and Detailed Targeting. A red box highlights a central column of records (Record 1, Record 2, ..., Record 52). Blue boxes highlight other records (Record 1, Record 2, ..., Record 53) and a 'Victim' record. Arrows indicate the flow of information from the records to the 'Victim' record, which is highlighted in red. The 'Victim' record is associated with a 'Potential Reach' of 2,600 people and an 'Estimated Daily Reach' of 520 - 1,400 people on Facebook. A question 'Victim PII matches FB account?' is posed at the bottom right, with 'No' and 'Yes' options.

Record 1
Record 2
...
Record 50
Record 51
Record 52

Record 1
Record 2
...
Record 51
Record 52
Record 53

Record 1
Record 2
...
Record 51
Record 52

Victim

Potential Reach: 2,600 people

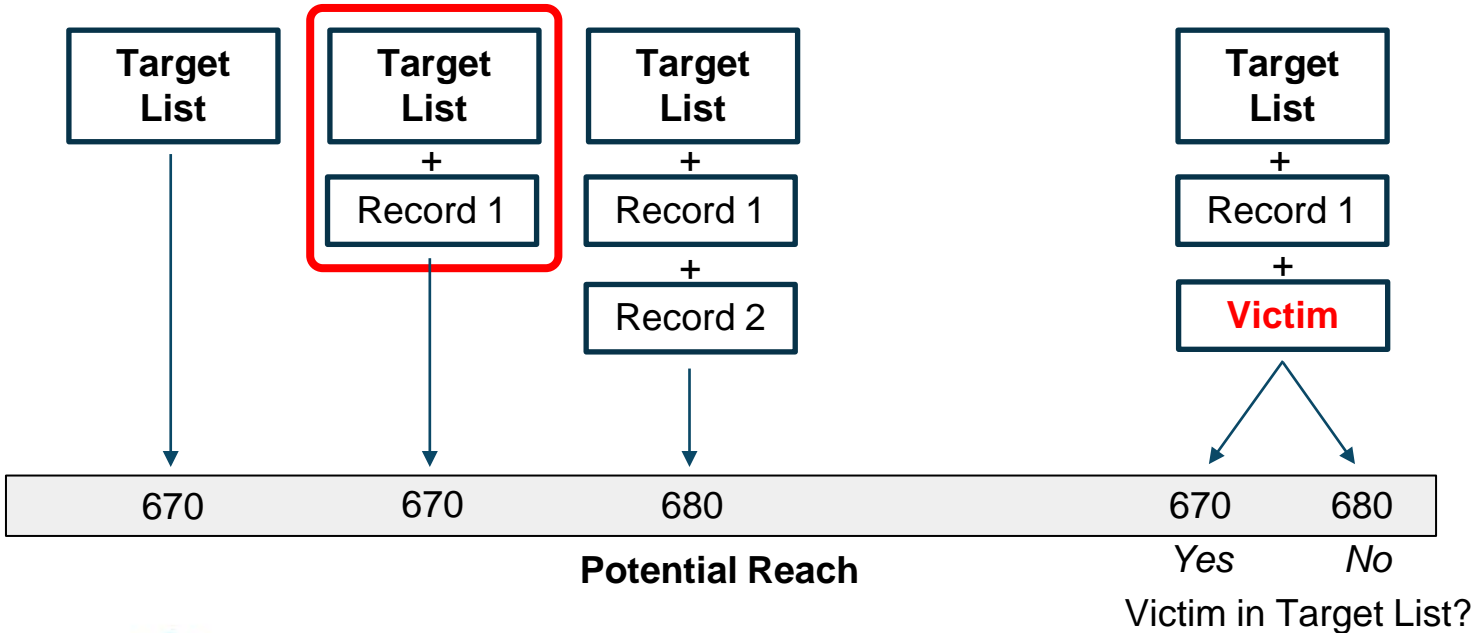
Estimated Daily Reach: 520 - 1,400 people on Facebook

270 - 530 people on Instagram

The accuracy of estimates is based on factors like past campaign data, the budget you entered and market data. Numbers are

Victim PII matches FB account?
No Yes

Insight 2: Determine if (matched) Victim is in List



Attack: Learning User's Phone Numbers

Can ask: Is **Victim** in Target List ?

Is **Victim** in

1	01-000-0000
1	01-000-0001
1	01-000-0002
	...
1	99-999-9998
1	99-999-9999

If No: First digit is not 1

If Yes: First digit is 1

Is **Victim** in

2	01-000-0000
2	01-000-0001
2	01-000-0002
	...
2	99-999-9998
2	99-999-9999

If No: First digit is not 2

If Yes: First digit is 2

Is **Victim** in

1	01-000-0000
1	01-000-0001
1	01-000-0002
	...
9	09-999-9998
9	09-999-9999

If No: Second digit is not 0

If Yes: Second digit is 0

Resulting Attacks

The attacks we discover allow attacker to:

- Link multiple pieces of PII to a single user

- Infer any active Facebook user's phone number

- De-anonymize visitors to attacker's website

No ads placed, no victim interaction, no way for victim to detect attack

Reported to Facebook, **mitigation in place**

Implications

Many online services now de-facto data brokers; use data for advertising

Anyone can be an advertiser!

Interfaces allow advertisers to query this database, in obfuscated form

Our work shows that **interfaces can inadvertently leak user data**

Highlights the need to audit advertising interfaces for privacy leaks

More info? Check out our IEEE S&P'18, NDSS'18, and FAT*'18 papers

<https://mislove.org>