

# Privacy Risks with Facebook's PII-based Targeting

Auditing a data broker's advertising interface

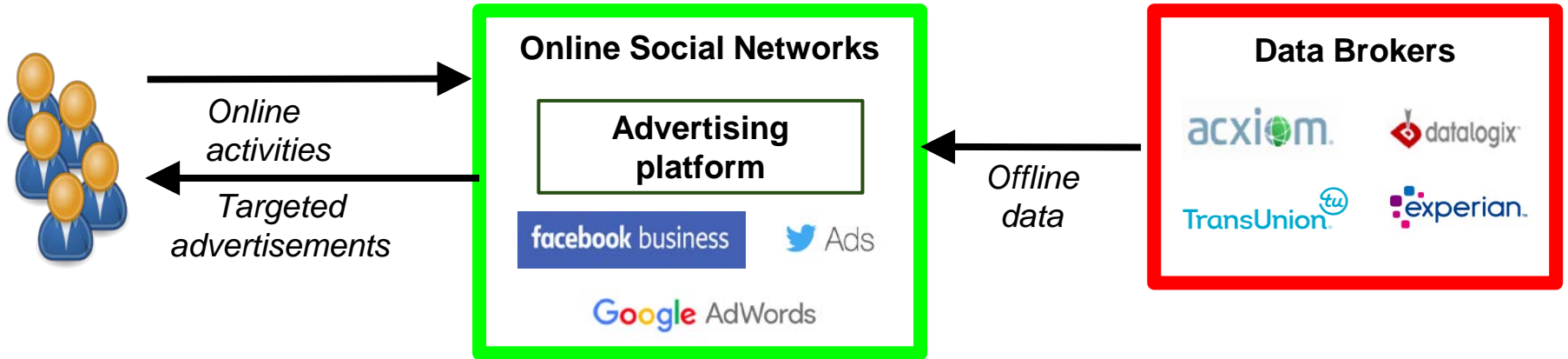
Giridhari Venkatadri<sup>1</sup>, Athanasios Andreou<sup>2</sup>, Yabing Liu<sup>1</sup>, Alan Mislove<sup>1</sup>,  
Krishna P. Gummadi<sup>3</sup>, Patrick Loiseau<sup>3,4</sup>, Oana Goga<sup>4</sup>

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

# Traditional Data Brokers



# 21<sup>st</sup> 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".

**Custom Audiences:** A yellow box highlights the "Custom Audiences" section, which includes a text input field containing "Customer List" and "list\_of\_9444\_records", and a button "Add Custom Audiences or Lookalike Audiences". Below this are "Exclude" and "Create New" options.

**Targeting Options:** A red box highlights the targeting options section, which includes:

- Locations:** "Everyone in this location" dropdown, "Include" dropdown, and "Add locations" button.
- Age:** "18" and "65+" dropdowns.
- Gender:** "All", "Men" (selected), and "Women" buttons.
- Languages:** "Enter a language..." input field.

**Detailed Targeting:** A section with the heading "INCLUDE people who match at least ONE of the following" and a "Browse" button.

**Audience Definition Summary:** On the right side, a gauge shows the audience is "Specific" (narrower) rather than "Broad". Below this, "Audience Details" lists:

- Custom Audience: list\_of\_9444\_records
- Age: 18 - 65+
- Gender: Male
- Placements: Facebook Feeds and Instagram Feed

**Reach Estimates:** A green box highlights the "Potential Reach: 2,600 people" estimate. Below it, "Estimated Daily Reach" shows:

- Facebook: 520 - 1,400 people on Facebook (of 1,700)
- Instagram: 270 - 530 people on Instagram (of 530)

# 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 is divided into two columns. The left column contains the audience configuration options: 'Custom Audiences' (with a list of 'Customer List' containing 'list\_of\_40k\_records' and 'list\_of\_9444\_records'), 'Locations' (set to 'United States'), 'Age' (set to '18 - 65+'), and 'Gender' (set to 'All'). The right column shows the 'Audience Size' section with a gauge indicating the audience is defined, and a 'Potential Reach' of 11,000 people. Below this is the 'Estimated Daily Results' section, showing a reach of 540 - 1,000. A blue box highlights the 'Potential Reach' value.

## 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). A yellow box highlights a record (Record 52) that is linked to a 'Victim' label. Arrows indicate the flow of records from the central column to the 'Victim' label and then to a decision point: 'Victim PII matches FB account?' with 'No' and 'Yes' options.

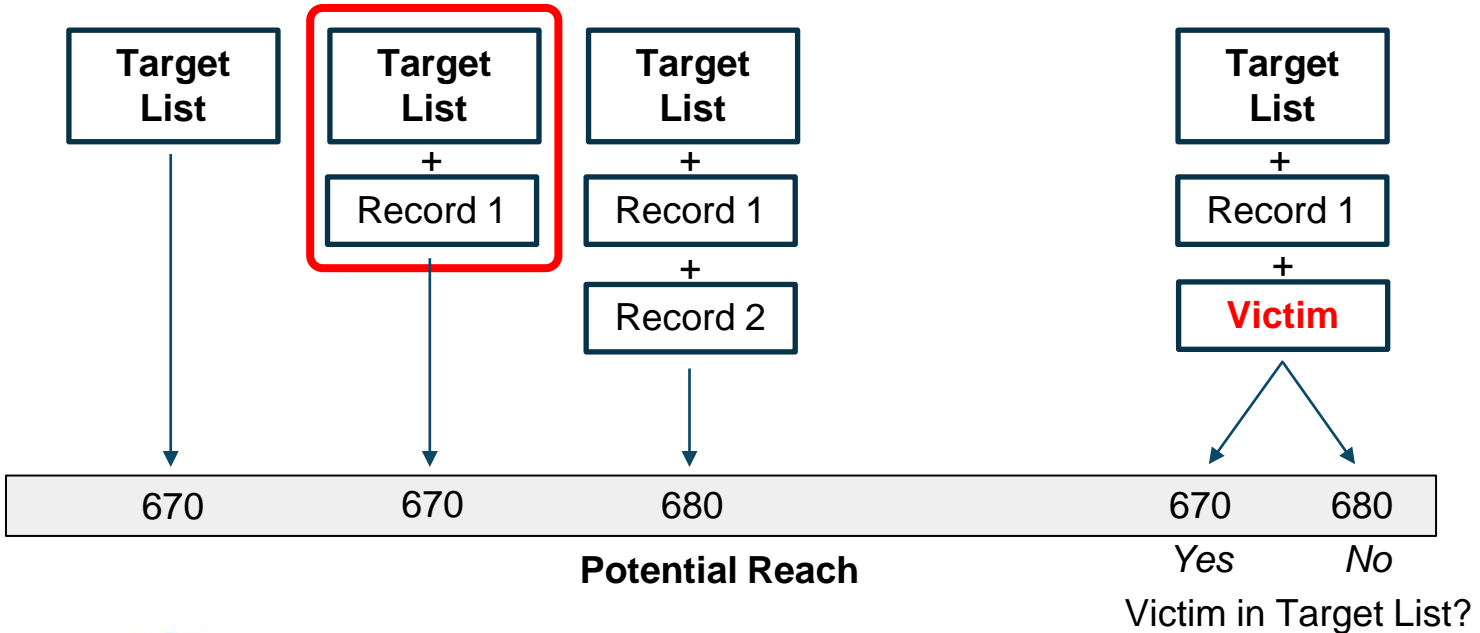
**Record 1**  
**Record 2**  
 ...  
**Record** 50  
**Record** 51  
**Record** 52  
 ...  
**Record** 53

**Victim**

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>