

06. Privacy in Targeted advertising and Facebook ads explanations

Nataliia Bielova

[@nataliabelova](https://twitter.com/nataliabelova)

September 17th-21st, 2018

Web Privacy course

University of Trento

DON'T WORRY, IT'S ONLY
MARKETERS COLLECTING
OUR PERSONAL DATA
SO THEY CAN CREATE
MORE RELEVANT
ADVERTISING FOR US.



TOM
FISH
BURNIE

Today's class

- Targeted advertising
- Real-Time Bidding protocol (RTB)
- Cookie synching
- Attribute and PII-based targeting on Facebook
- Investigating Facebook Ad explanations



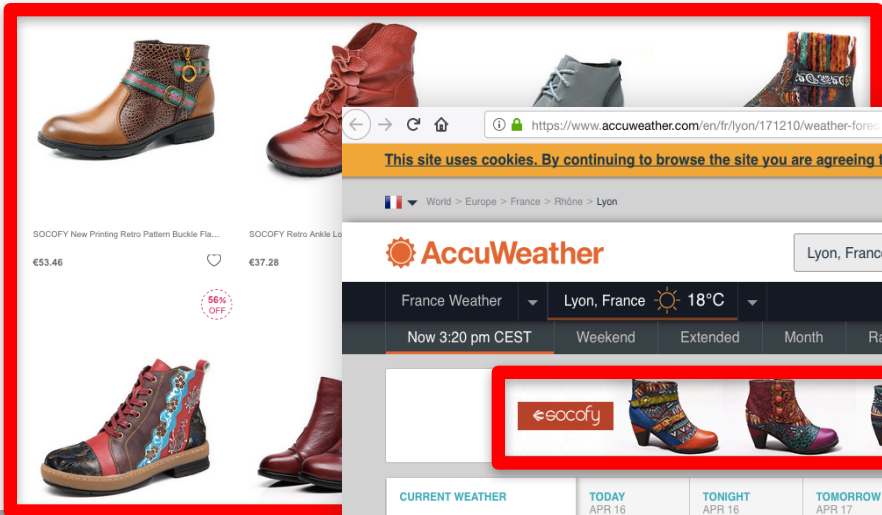
Browser: <https://www.newchic.com/fashion-collection/1385.html?utm> 80%

English €EUR Online Help

Newchic

Search Sign in

BOOTS FLATS AND PUMPS SANDALS



SOCOFOY New Printing Retro Pattern Buckle Fla... €53.46 56% OFF

SOCOFOY Retro Ankle Lo €37.28

Browser: <https://www.accuweather.com/en/fr/lyon/171210/weather-forec>

This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here

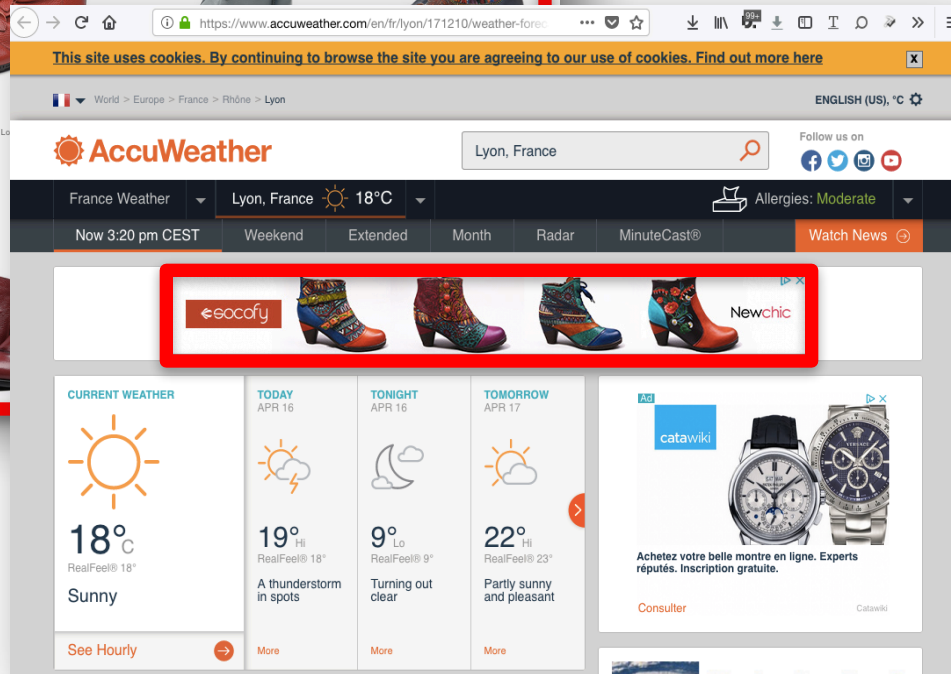
World > Europa > France > Rhône > Lyon ENGLISH (US), °C

AccuWeather

Lyon, France

France Weather Lyon, France 18°C Allergies: Moderate


Now 3:20 pm CEST Weekend Extended Month Radar MinuteCast® Watch News



←socofoy Newchic

CURRENT WEATHER	TODAY APR 16	TONIGHT APR 16	TOMORROW APR 17
 18°C RealFeel® 18° Sunny	 19° Hi RealFeel® 18° A thunderstorm in spots	 9° Lo RealFeel® 9° Turning out clear	 22° Hi RealFeel® 23° Partly sunny and pleasant
See Hourly →	More	More	More

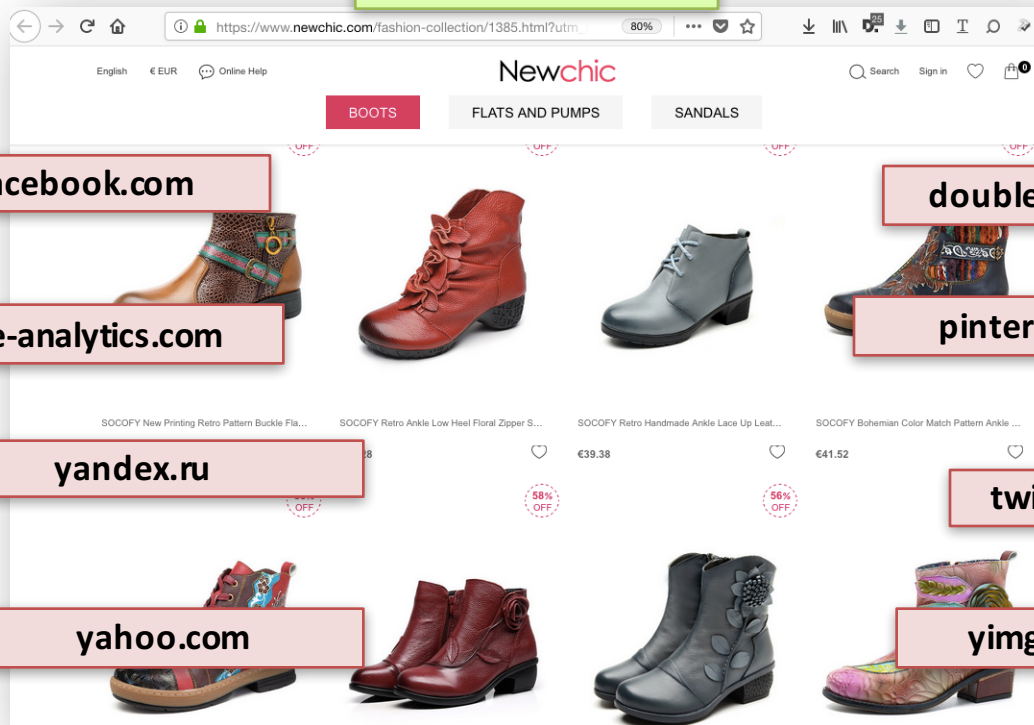
Ad **cataviki**



Achetez votre belle montre en ligne. Experts réputés. Inscription gratuite.

Consulter Cataviki

newchic.com



facebook.com

doubleclick.net

google-analytics.com

pinterest.com

yandex.ru

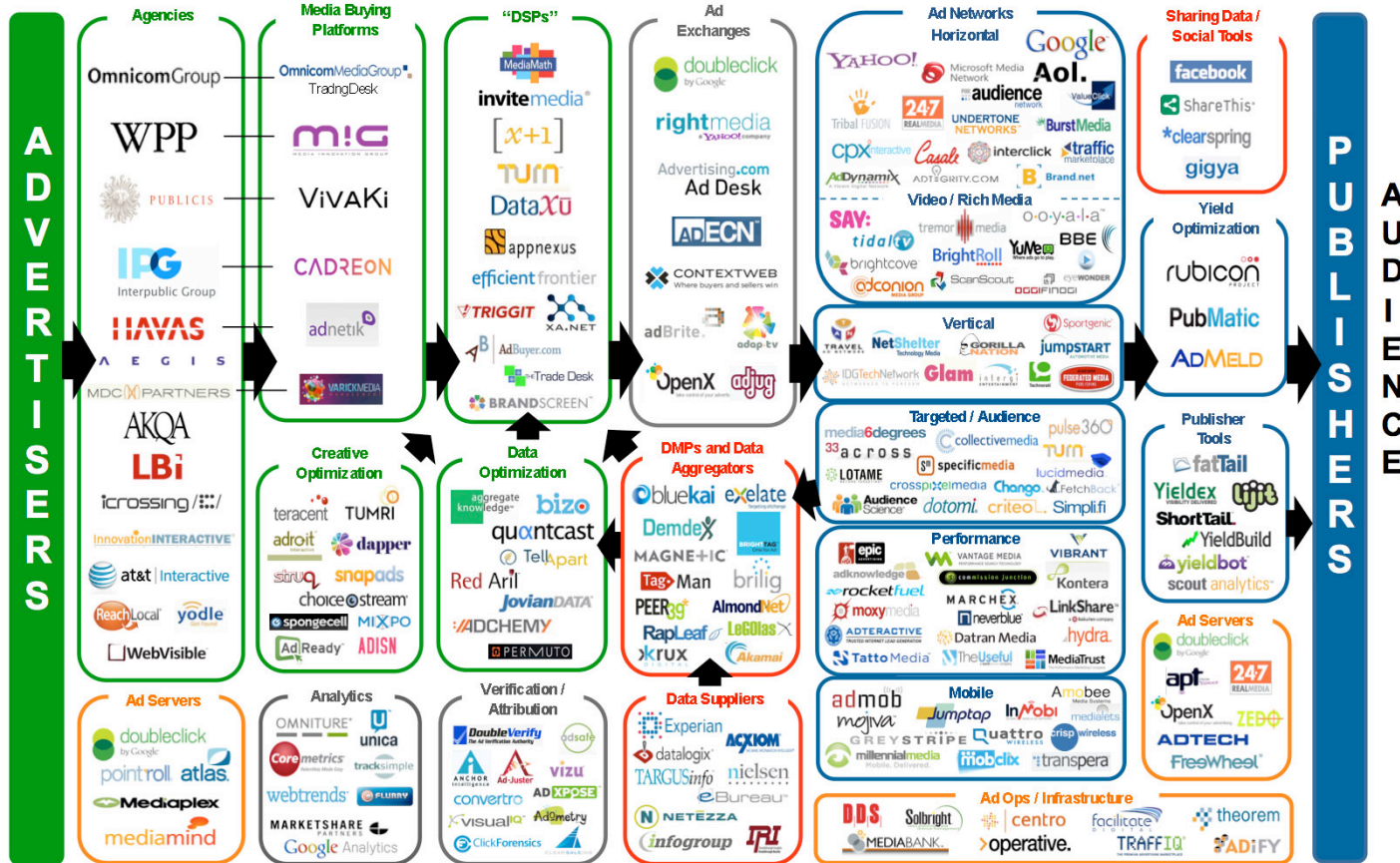
twitter.com

yahoo.com

yimg.com

Display Advertising Technology Landscape

2010



VIDEO LUMAscape

2017



Denotes acquired company

Denotes shuttered company

© LUMA Partners LLC 2017



Targeted Advertisement and Real-Time Bidding

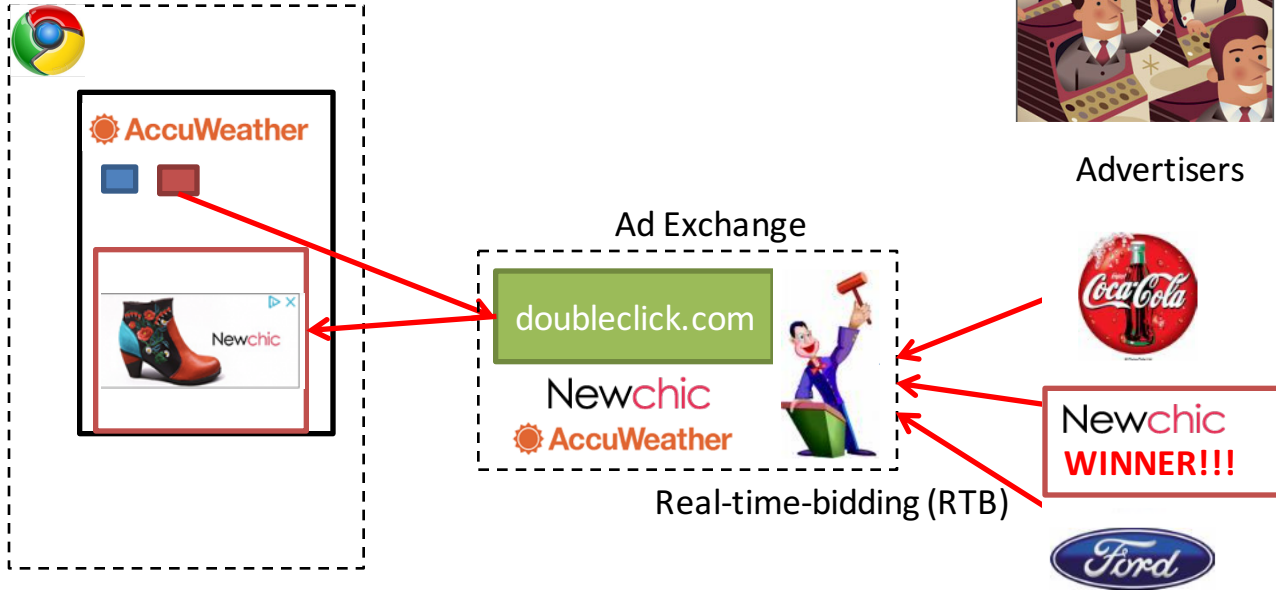
COLLABORATION BETWEEN PUBLISHERS, AD EXCHANGES AND ADVERTISERS

Targeted advertisement

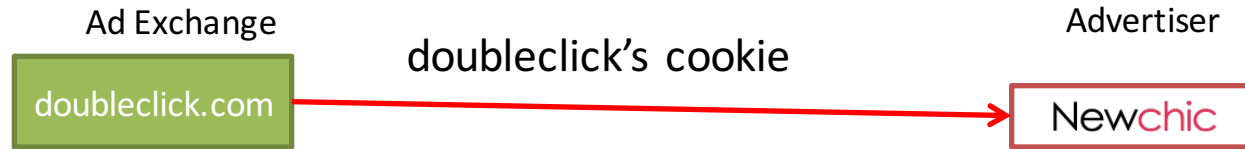
- Targeted advertisement allows to assign a concrete ad to an interested user
- **Real-Time Bidding (RTB)** is a protocol for transacting digital display ads in real time



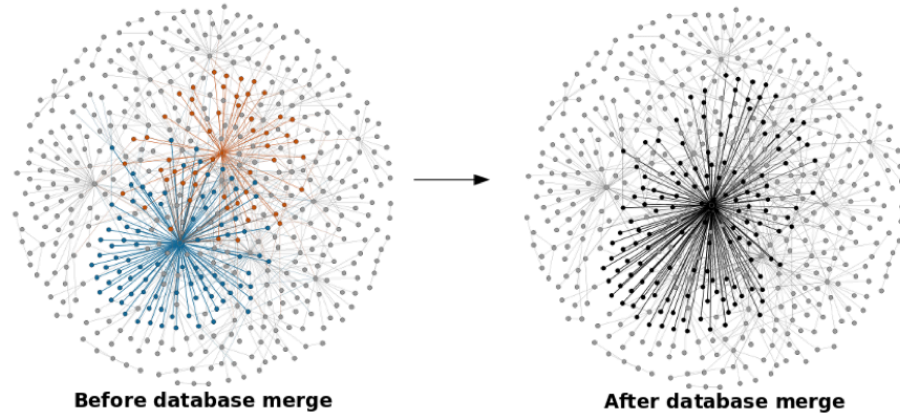
Real-Time Bidding (RTB)



Why would advertiser bid on a user?



Advertiser cannot read their own cookie for this user!

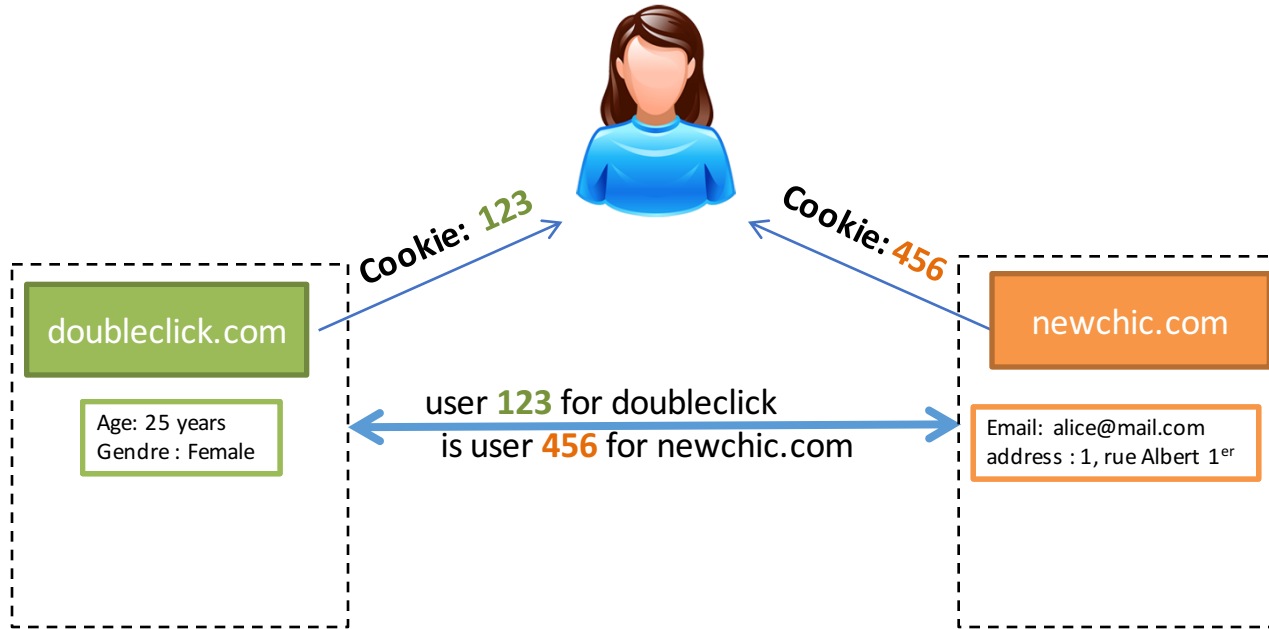


Cookie synchronization

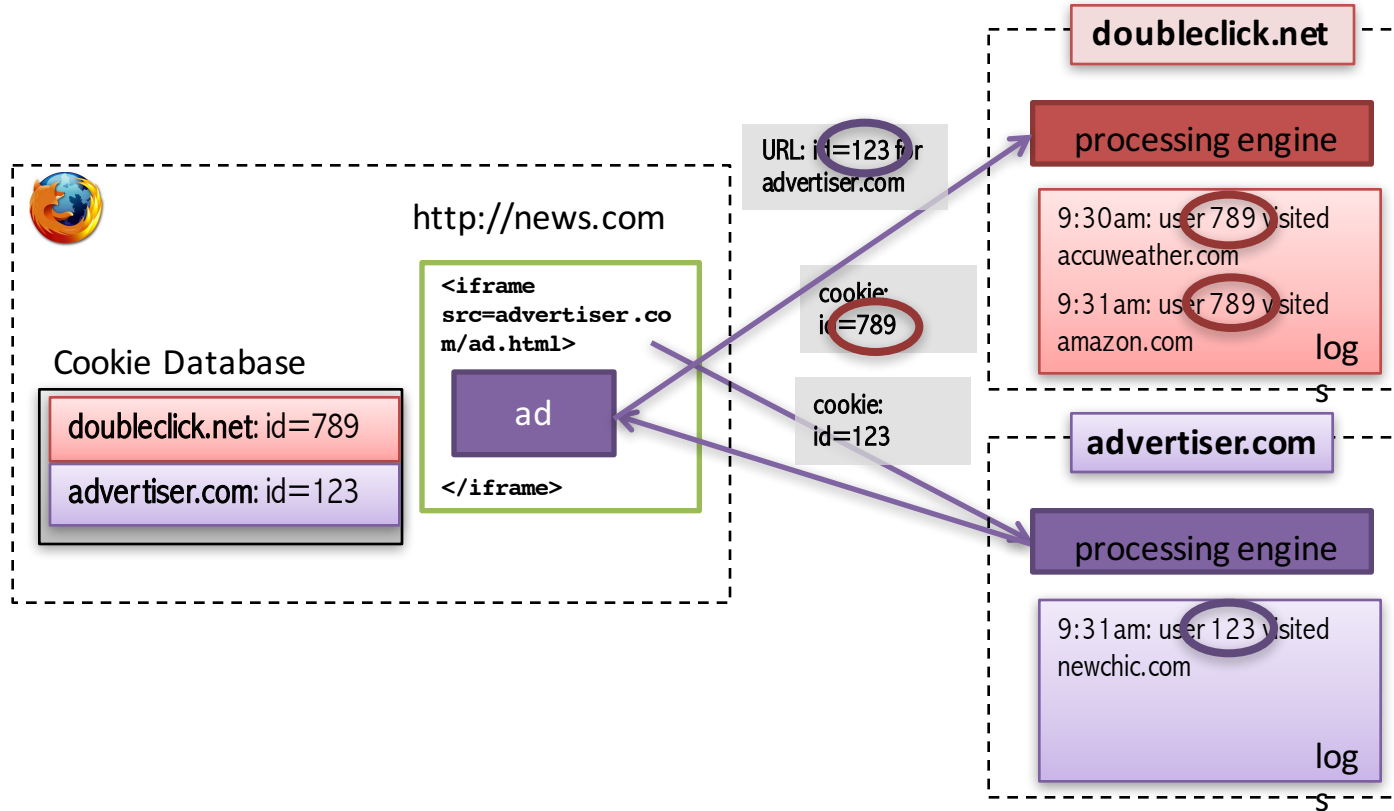
What is a Cookie synchronization?

- the process by which two different trackers **link the IDs** they've given to the same user

Cookie synchronization



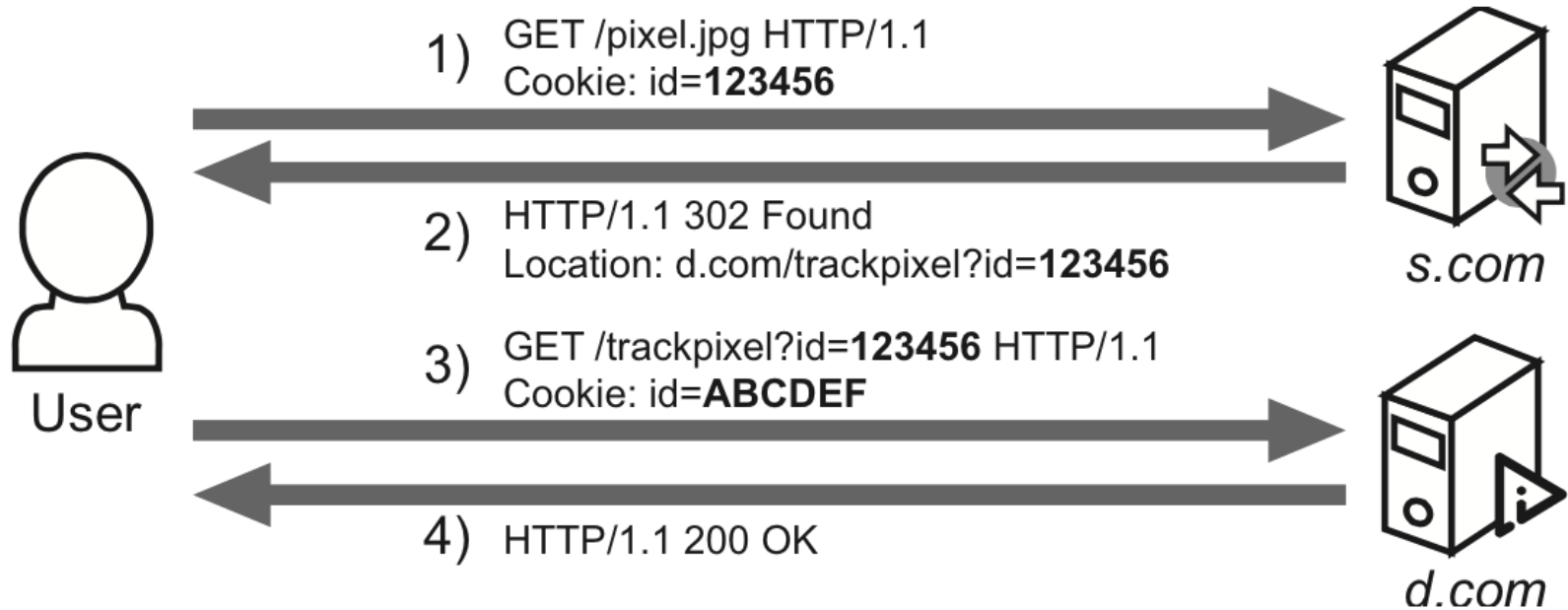
Example of cookie synchronization



*Hypothetical relations

Cookie Syncing

- *s.com* matches their cookie with *d.com* using an HTTP redirect



What if I delete all my cookies?

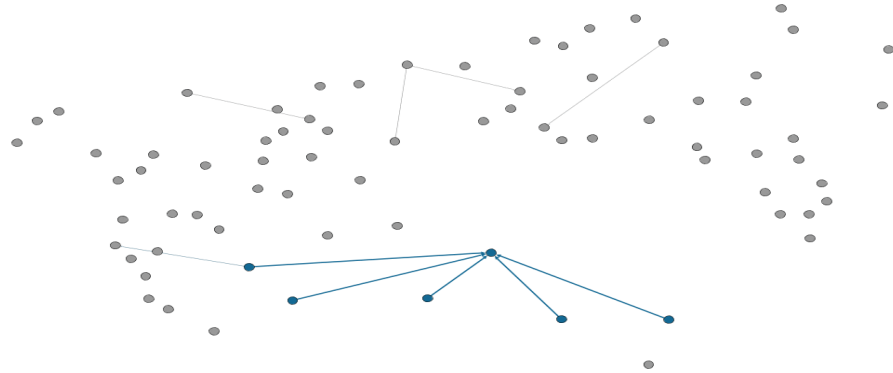
- **Important detail:**

- If at least one tracker respawns one cookie, he passes it to other trackers

- **Privacy problem:**

- Thus, even trackers that don't employ respawning gain the ability to continually track users who clear cookies!

Cookie syncing graphically



Large-scale study of cookie syncing

- 3,000 domains, top parties involved in cookie syncing:

All Cookies Allowed		No 3P Cookies	
Domain	# IDs	Domain	# IDs
gemius.pl	33	gemius.pl	36
doubleclick.net	32	2o7.net	27
2o7.net	27	omtrdc.net	27
rubiconproject.com	25	cbsi.com	26
omtrdc.net	24	parse.ly.com	16
cbsi.com	24	marinsm.com	14
adnxs.com	22	gravity.com	14
openx.net	19	cxense.com	13
cloudfront.net	18	cloudfront.net	10
rlcdn.com	17	doubleclick.net	10

Even more IDs get synced if the user blocks third-party cookies!



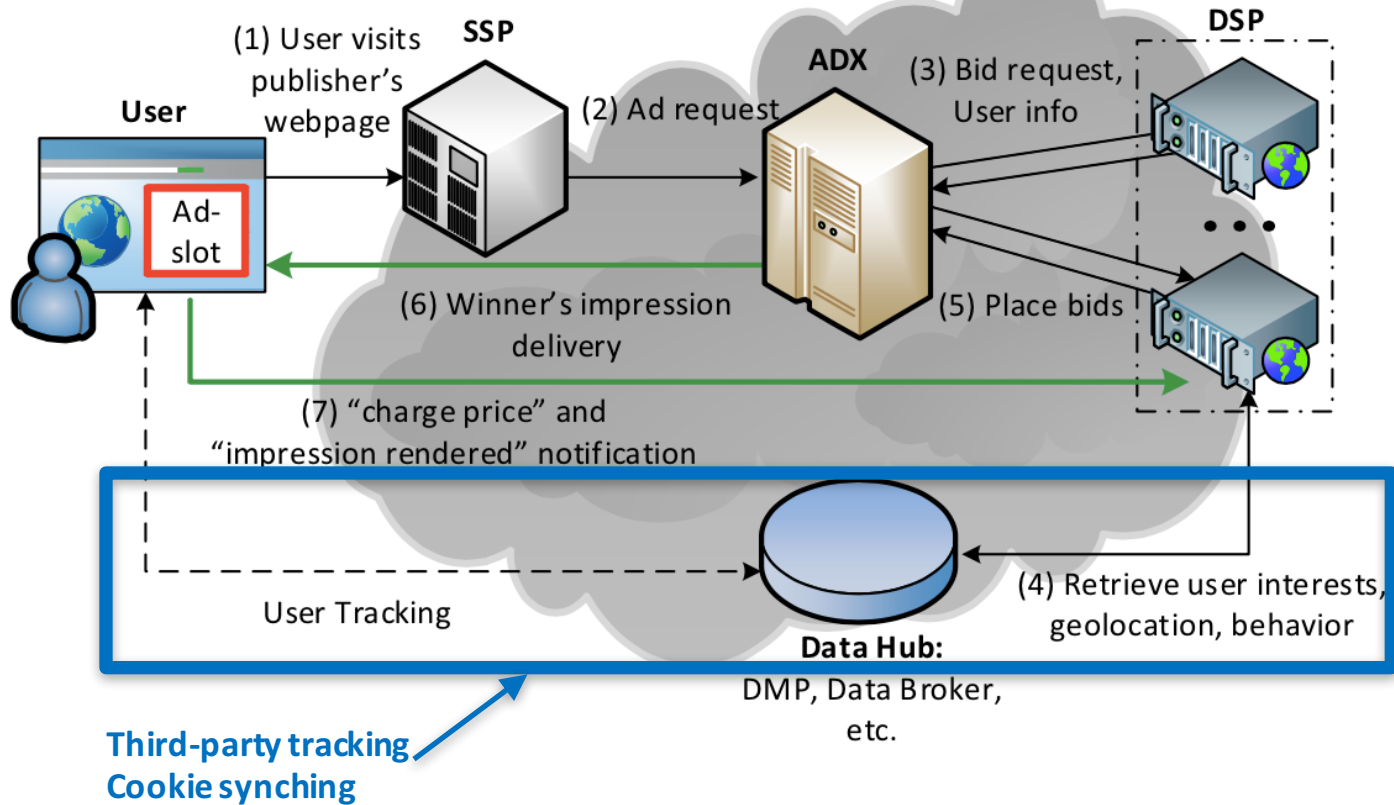
Large-scale study of cookie syncing

- 100,000 domains
- doubleclick.net is syncing the most:
 - 108 cookies to 118 partners
- The vast majority of top third parties sync cookies with at least one other party :
 - 45 out of top 50
 - 85 out of top 100
 - 157 out of top 200
 - 460 out of top 1,000

Limitations on detection of synched IDs

- Only detect matched user IDs in clear – not detected if IDs are obfuscated
- Two studies examined the behavior of DoubleClick only.
- Imprecise detection of information flows between companies because they rely on HTTP headers
 - Inclusion of dynamic content is not linked to the party that created it

Real-Time Bidding (RTB)



Retargeted ads

- a user is targeted with ads related to the exact products she has previously browsed



Cisco-Linksys AE1000 High-Performance Wireless-N Adapter
by Linksys
★★★★☆ 207 customer reviews | 10 answered questions

Price: **\$15.99** ✓Prime

Only 1 left in stock.
Want it Tuesday, June 14? Order within **33 hrs 50 mins** and choose **One-Day Shipping** at checkout. [Details](#)

Sold by [Home Sweet Home Direct](#) and [Fulfilled by Amazon](#).

Eligible for [amazon smile](#) donation.

 **Want to hire a computer technician?**
Buy professional computer technician services directly on Amazon. Backed by our Happiness Guarantee.
[Learn more](#)

• Networking Equipment Features: WEP Security, WPA Security, Easy Setup, WPA2



amazon.com [Shop now](#)

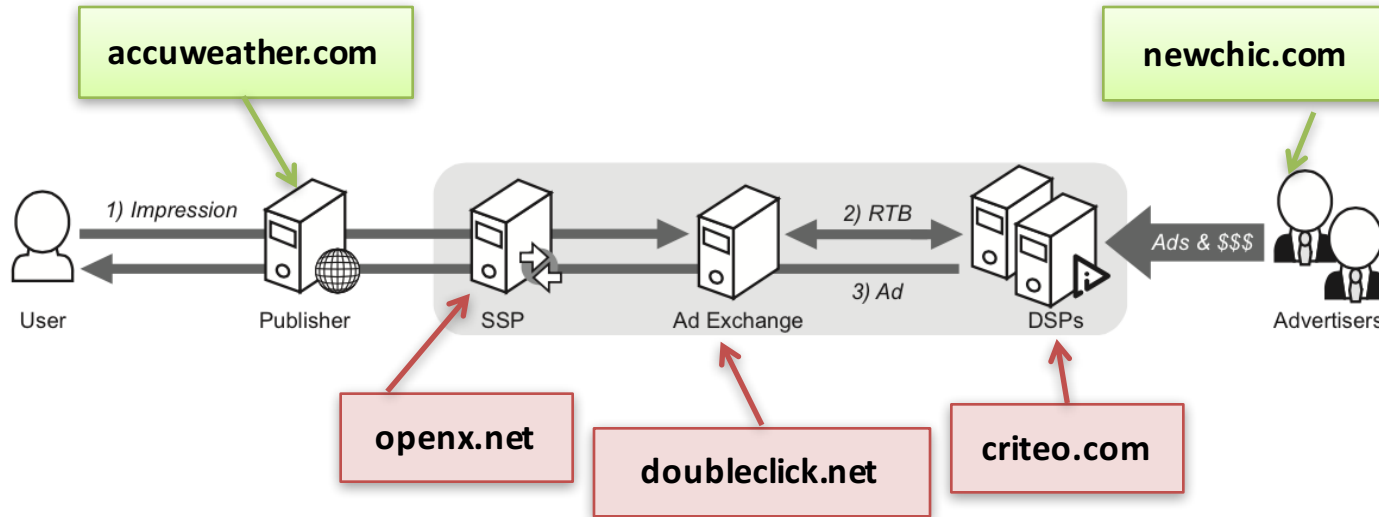
Linksys/Cisco AE1000
300Mbps 802.11n
Dual-Band Wireless...
~~\$39.99~~ **\$15.99**

Privacy

Key insight: use retargeted ads to detect information flows

- Two conditions for retargeted ads served for user u :
 - the company must know that u browsed a specific product on a specific e-commerce site
 - the company must be able to uniquely identify u during an auction
- [see [an experiment setup](#)]

Real-Time Bidding (RTB)



Who is Cookie Matching?

Participant 1		Participant 2	Chains	Ads	Heuristics
criteo	↔	googlesyndication	9090	1887	↔ P
criteo	↔	doubleclick	3610	1144	→ E, P ← DC, P
criteo	↔	adnxs	3263	1066	↔ E, P
criteo	↔	rubiconproject	1586	749	↔ E, P
criteo	↔	servedbyopenx	707	460	↔ P
doubleclick	↔	steelhousemedia	362	27	→ P ← E, P
mathtag	↔	mediaforge	360	124	↔ E, P
netmng	↔	scene7	267	119	→ E ← ?
googlesyndication	↔	adsvr	107	29	↔ P
rubiconproject	↔	steelhousemedia	86	30	↔ E
googlesyndication	↔	steelhousemedia	47	22	?
adtechus	→	adacado	36	18	?
atwola	→	adacado	32	6	?
adroll	↔	adnxs	31	8	?

Heuristics Key (used by prior work)

E – share exact cookies

P – special URL parameters

DC – DoubleClick URL parameters

? – Unknown sharing method

Who is Cookie Matching?

Participant 1		Participant 2	Chains	Ad	Heuristics
criteo	↔	googlesyndication	9090	1887	↔ P
criteo	↔	doubleclick	3610	1144	→ E, P ← DC, P
criteo	↔	adnxs	3263	1066	↔ E, P
criteo	↔	rubiconproject	1586	749	↔ E, P
criteo	↔	servedbyopenx	707	460	↔ P
doubleclick	↔	steelhousemedia	362	27	→ P ← E, P
mathtag	↔	mediaforge	360	124	↔ E, P
netmng	↔	scene7	267	119	→ E ← ?
googlesyndication	↔	adsvr	107	29	↔ P
rubiconproject	↔	steelhousemedia	86	30	↔ E
googlesyndication	↔	steelhousemedia	47	22	?
adtechus	→	adacado	36	18	?
atwola	→	adacado	32	6	?
adroll	↔	adnxs	31	8	?

Heuristics Key (used by prior work)

E – share exact cookies

P – special URL parameters

DC – DoubleClick URL parameters

? – Unknown sharing method

Who is Cookie Matching?

Participant 1		Participant 2	Chains	Ad	Heuristics
criteo	↔	googlesyndication	9090	1887	↔ P
criteo	↔	doubleclick	3610	1144	→ E, P ← DC, P
criteo	↔	adnxs	3263	1066	↔ E, P
criteo	↔	rubiconproject	1586	749	↔ E, P
criteo	↔	servedbyopenx	707	460	↔ P
doubleclick	↔	steelhousemedia	362	27	→ P ← E, P
mathtag	↔	mediaforge	360	124	↔ E, P
netmng	↔	scene7	267	119	→ E ← ?
googlesyndication	↔	adsvr	107	29	↔ P
rubiconproject	↔	steelhousemedia	86	30	↔ E
googlesyndication	↔	steelhousemedia	47	22	?
adtechus	→	adacado	36	18	?
atwola	→	adacado	32	6	?
adroll	↔	adnxs	31	8	?

Heuristics Key (used by prior work)

E – share exact cookies

P – special URL parameters

DC – DoubleClick URL parameters

? – Unknown sharing method

Who is Cookie Matching?

Participant 1		Participant 2	Chains	Ads	Heuristics
criteo	↔	googlesyndication	9090	1887	↔ P
criteo	↔	doubleclick	3610	1144	→ E, P ← DC, P
criteo	↔	adnxs	3263	1066	↔ E, P
criteo	↔	rubiconproject	1586	749	↔ E, P
criteo	↔	servedbyopenx	707	460	↔ P
doubleclick	↔	steelhousemedia	362	27	→ P ← E, P
mathtag	↔	mediaforge	360	124	↔ E, P
netmng	↔	scene7	267	119	→ E ← ?
googlesyndication	↔	adsvr	107	29	↔ P
rubiconproject	↔	steelhousemedia	86	30	↔ E
googlesyndication	↔	steelhousemedia	47	22	?
adtechus	→	adacado	36	18	?
atwola	→	adacado	32	6	?
adroll	↔	adnxs	31	8	?

Heuristics Key (used by prior work)

E – share exact cookies

P – special URL parameters

DC – DoubleClick URL parameters

? – Unknown sharing method

31% of cookie matching partners would be missed.

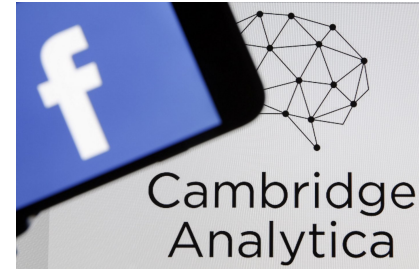
RTB and Tracking ecosystem

- **Important detail:**
 - Targeted advertisement and RBT is based on personal data: cookies, browsing history, preferences...
- **Transparency studies:**
 - Hard to analyse tracking and RTB together
 - Only a part of the process is visible in the user's browser
 - Hard to establish what data is collected by trackers

Privacy Risks in Targeted Advertising

Cambridge Analytica scandal

- A lot of user's data collected
- Political ads that target users
- Reactions
 - How did they collect all that data?
 - Why collect this data? What did it allow them to do?



Why collect this data? What did it allow them to do?

*“Cambridge Analytica generally utilizes **custom audiences**, some of which were created from contract lists and other identifiers... and in some instances, refined those audiences with **additional targeting attributes**.”*

- Facebook (in response to U.S. Senate questions)

How many of you understand what this means?

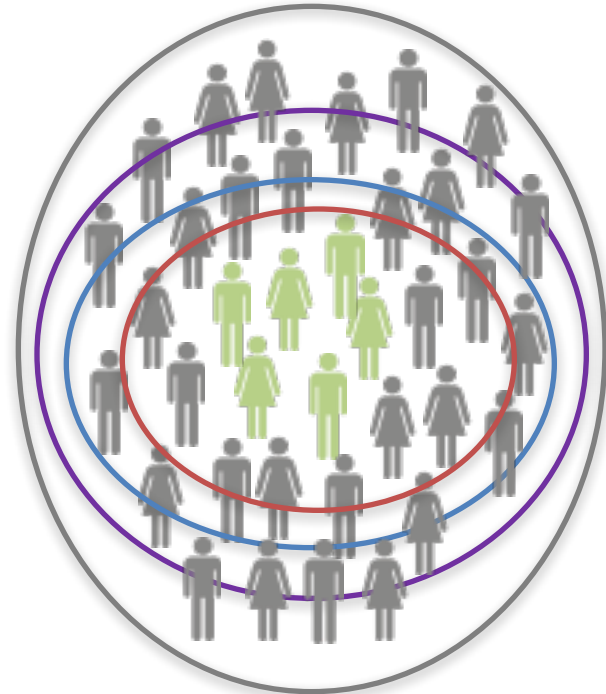
Targeted advertising

- Users who see ad called the *audience*
 - How can advertiser select their audience?

Ad on a broadcast medium (TV)

Ad targeting *search keywords*

Ad targeting *user attributes*



Ads on Facebook

The image displays three overlapping Facebook advertisements. The top-left ad is for Coralie Masson, featuring a brown suede boot with a silver glitter stripe. The middle ad is for Fluent Forever, showing a hand holding a smartphone displaying the app interface. The bottom-right ad is for Swatch, showing a person's wrist wearing a blue Swatch watch. Each ad includes the brand name, a 'Sponsored' label, a 'Like Page' button, and a 'Shop Now' button. The Fluent Forever ad also includes a 'Suggested Post' label and a 'Like Page' button. The Swatch ad includes a 'Suggested Post' label and a 'Like Page' button. The Coralie Masson ad has a 'Like Page' button and a 'Shop Now' button. The Fluent Forever ad has a 'Shop Now' button. The Swatch ad has a 'Shop Now' button. The Coralie Masson ad has a 'Like Page' button and a 'Shop Now' button. The Fluent Forever ad has a 'Like Page' button and a 'Shop Now' button. The Swatch ad has a 'Like Page' button and a 'Shop Now' button.

Coralie Masson Sponsored ·

Parce qu'il vous faut des Coralie Masson sous le sapin, n'attendez plus et passez vos commandes avant le 20 décembre!

Boots Josy cognac et glitter

29

Like Comment Share

Fluent Forever - Learn Any Language Sponsored ·

Suggested Post

The Secret to Learning a New Language is Sticking To It. Fluent Forever offers exercises and programs to keep you engaged at any level. The BEST part: what you learn, you KEEP for life. Pre-Order for E

Fluent Forever
The App

Unlock The Secret of Language Learning.
The Most Successful App in Kickstarter History. Now Live on

INDIEGOGO.COM

3

Like Comment Share

Swatch Sponsored ·

Suggested Post

Soyez au top de la tendance minimaliste avec les nouveaux modèles SKIN.

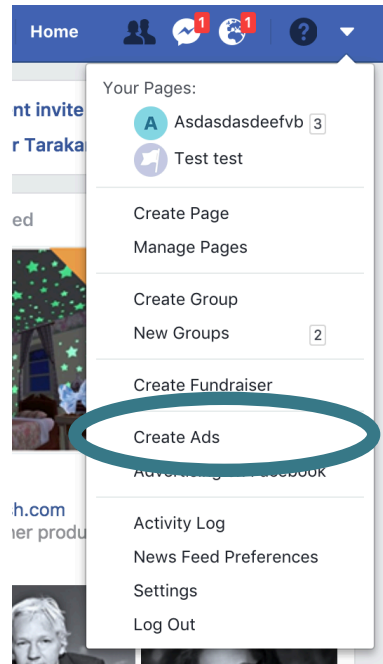
#YOURMOVE, votre style
Imposez votre style

SWATCH.COM

9

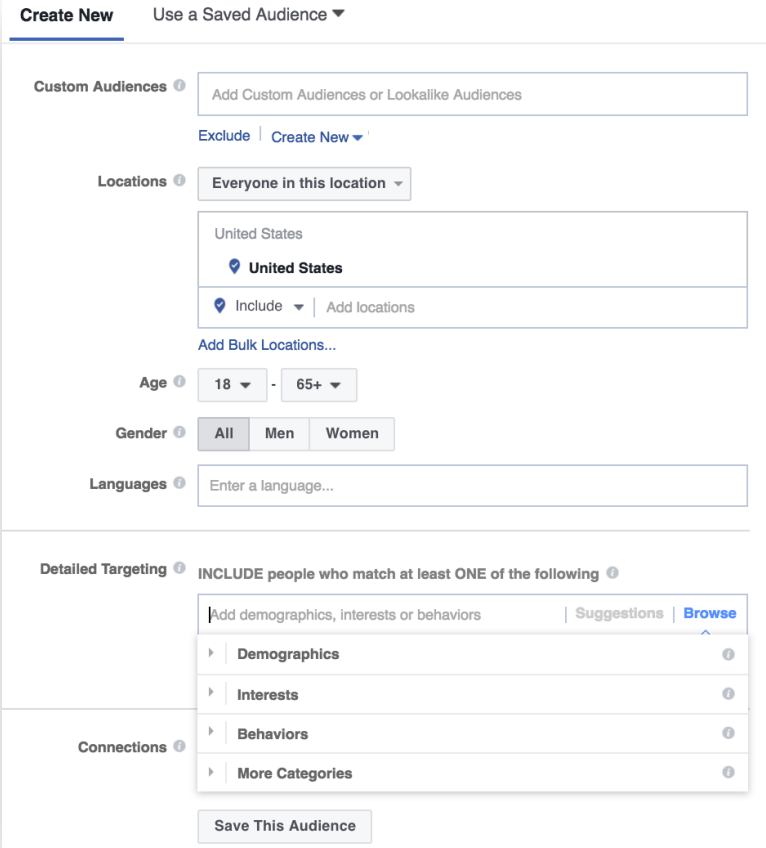
Like Comment Share

Anyone can be an advertiser in just a few clicks!



Targeting method 1: *Attributes*

- Advertiser can **include/exclude attributes** to create audience
- Attributes attached to users based on
 - Facebook activity
 - Third-party Web browsing
 - “Partner” companies
- >1,200 well-defined attributes
- >250,000 free-form attributes



The screenshot shows the Facebook Audience Targeting interface. At the top, there are two tabs: "Create New" (selected) and "Use a Saved Audience". Below the tabs, there are several sections for defining the audience:

- Custom Audiences:** A text input field with the placeholder "Add Custom Audiences or Lookalike Audiences". Below it are "Exclude" and "Create New" options.
- Locations:** A dropdown menu set to "Everyone in this location". Below it is a list of locations, with "United States" selected. There are "Include" and "Add locations" options.
- Age:** Two dropdown menus set to "18" and "65+".
- Gender:** Three buttons: "All" (selected), "Men", and "Women".
- Languages:** A text input field with the placeholder "Enter a language...".

Below these sections is the "Detailed Targeting" section, which is set to "INCLUDE people who match at least ONE of the following". It features a search bar with the placeholder "Add demographics, interests or behaviors" and "Suggestions" and "Browse" links. Below the search bar is a list of categories:

- Demographics
- Interests
- Behaviors
- More Categories

At the bottom of the interface is a "Save This Audience" button.

How can advertisers target users?

Traditional targeted advertising: boolean formula

Audience = married + new mover + lives in Grenoble

Many attributes available for targeting

Detailed Targeting ⓘ INCLUDE people who match at least ONE of the following ⓘ

Add demographics, interests or behaviors | [Suggestions](#) | [Browse](#)

- ▶ Multicultural affinity
- ▶ Purchase behavior
- ▼ Residential profiles
 - Likely to move
 - New mover
 - Recent homebuyer
 - Recent mortgage borrower
- ▶ Travel
- ▶ More Categories

- ▶ Financial
- ▶ Football
- ▼ Job role
 - Corporate executives
 - Financial professionals
- ▶ Media
- ▶ Mobile Device User
- ▶ Mobile Device User/Device Use Time

Connections ⓘ

Demographics

Show your ads to the right people

Attributes can be invasive

The image shows a search filter interface with three overlapping panels. The top-left panel is titled "INCLUDE people who match at least ONE of the following" and contains a search input with "Fascism" and a dropdown menu with "Fascism" and "Anti-fascism". The top-right panel is titled "INCLUDE people who match at least ONE of the following" and contains a search input with "Homosexuality" and a dropdown menu with "Homosexuality" and "Interests". The bottom panel is titled "INCLUDE people who match at least ONE of the following" and contains a search input with "Income >" and a dropdown menu with "Income > 2. \$50,000 - \$74,999", "Income > 3. \$75,000 - \$99,999", "Income > 6. \$150,000 - \$249,999", "Income > 4. \$100,000 - \$124,999", "Income > 5. \$125,000 - \$149,999", and "Income > 1. \$40,000 - \$49,999".

INCLUDE people who match at least ONE of the following

Fascism | Suggestions | Browse

Fascism

Anti-fascism

INCLUDE people who match at least ONE of the following ⓘ

Homosexuality | Suggestions | Browse

Homosexuality

Interests

Interests

INCLUDE people who match at least ONE of the following ⓘ

Income > | Suggestions | Browse

Income > 2. \$50,000 - \$74,999	Demographics
Income > 3. \$75,000 - \$99,999	Demographics
Income > 6. \$150,000 - \$249,999	Demographics
Income > 4. \$100,000 - \$124,999	Demographics
Income > 5. \$125,000 - \$149,999	Demographics
Income > 1. \$40,000 - \$49,999	Demographics

Data on everyone

- Out of 230M users (US):
 - Politics (US): 179M
 - Financial/Net worth/Liquid Assets: 74M
 - Family & Relationships: 138M

Where these attributes come from?

Collected all targeting attributes form 8 countries

Country	Facebook	Acxiom	Epsilon	Experian	DLX	Data broker (unspecified)	Total
US	614	128	14	5	350	10	1121
UK	614	103	0	17	19	0	753
France	614	21	0	0	0	0	635
Germany	614	60	0	0	0	0	674
Australia	614	24	0	34	0	0	672
Brazil	614	0	0	20	0	0	634
Japan	614	17	0	0	0	0	631
India	614	0	0	0	0	0	614

PII-based targeting

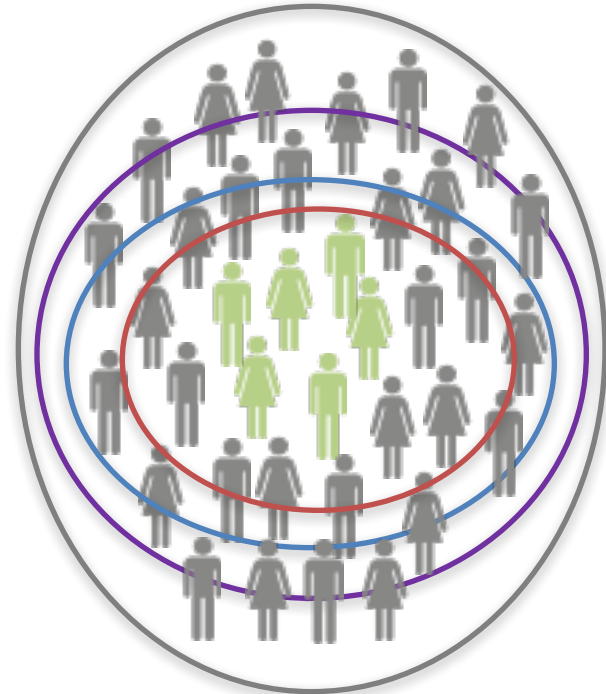
Targeted advertising

- Users who see ad called the *audience*
 - How can advertiser select their audience?

Ad on a broadcast medium (TV)

Ad targeting *search keywords*

Ad targeting *user attributes*



Targeted advertising

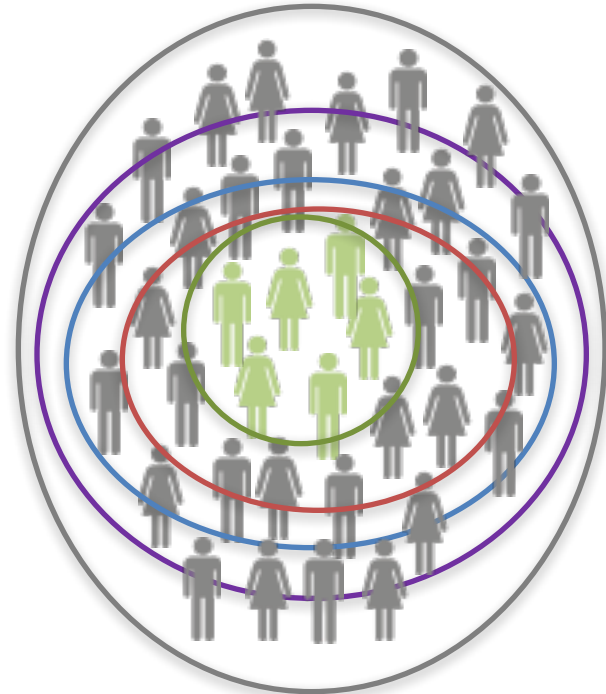
- Users who see ad called the *audience*
 - How can advertiser select their audience?

Ad on a broadcast medium (TV)

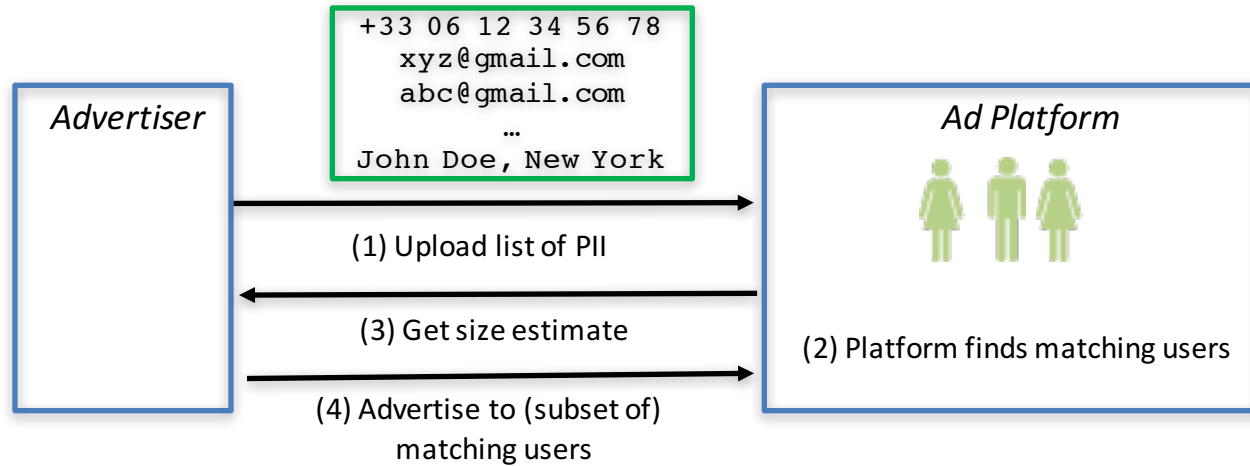
Ad targeting *search keywords*

Ad targeting *user attributes*

Ad targeting *identified users*



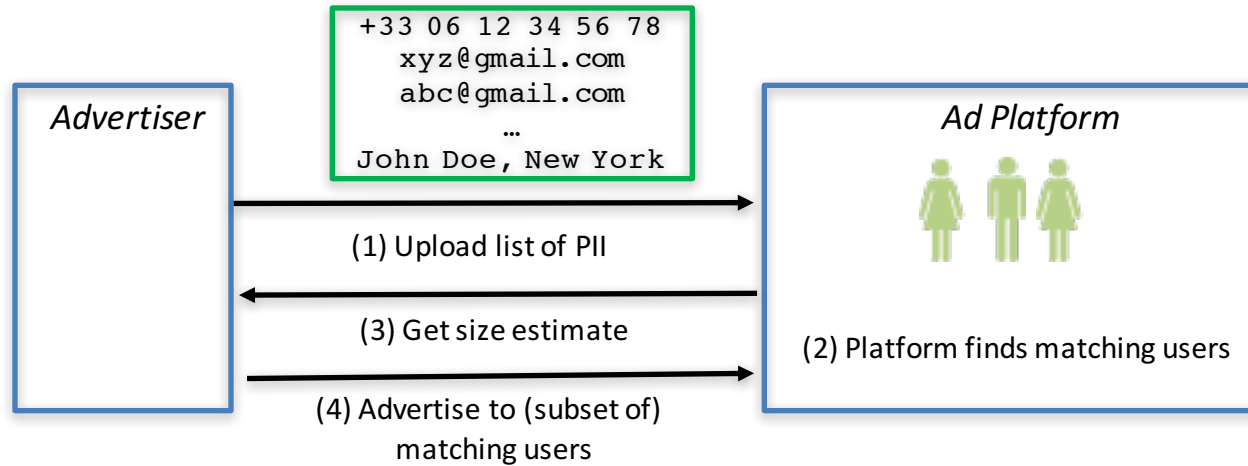
Targeting method 2: *PII*



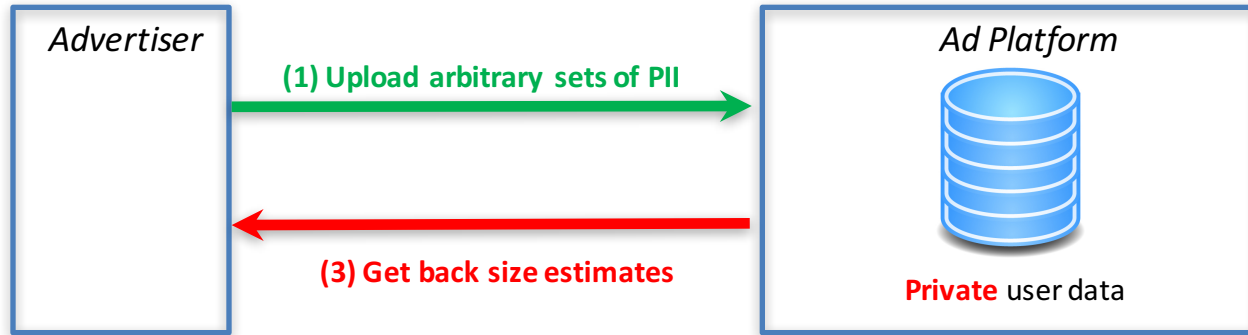
Advantages for advertiser:

1. Pay only **for users** you want to reach
2. Exploit different **external data** sources

PII-based targeting is common



PII-based targeting is common



This is a **query** to the user database!

Could these statistics **intentionally** leak user information?

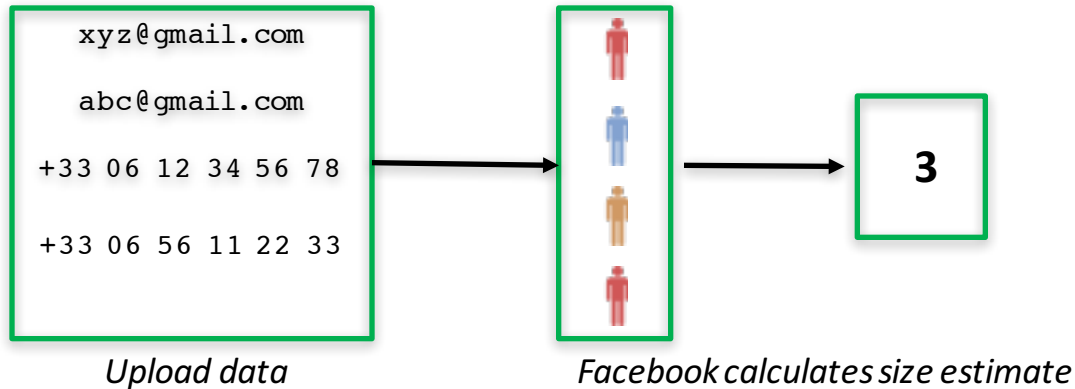
Anybody can be an advertiser...

Features of Facebook's size estimates

1. Size estimates obfuscated by **sample rounding**
(20, 30, 40, ..., 1000, 1100, 1200, ..., 10000, 11000, ...)
2. Records matching same user are **de-duplicated**

Obfuscation

De-duplication



Exploiting these features

Is

Victim's email

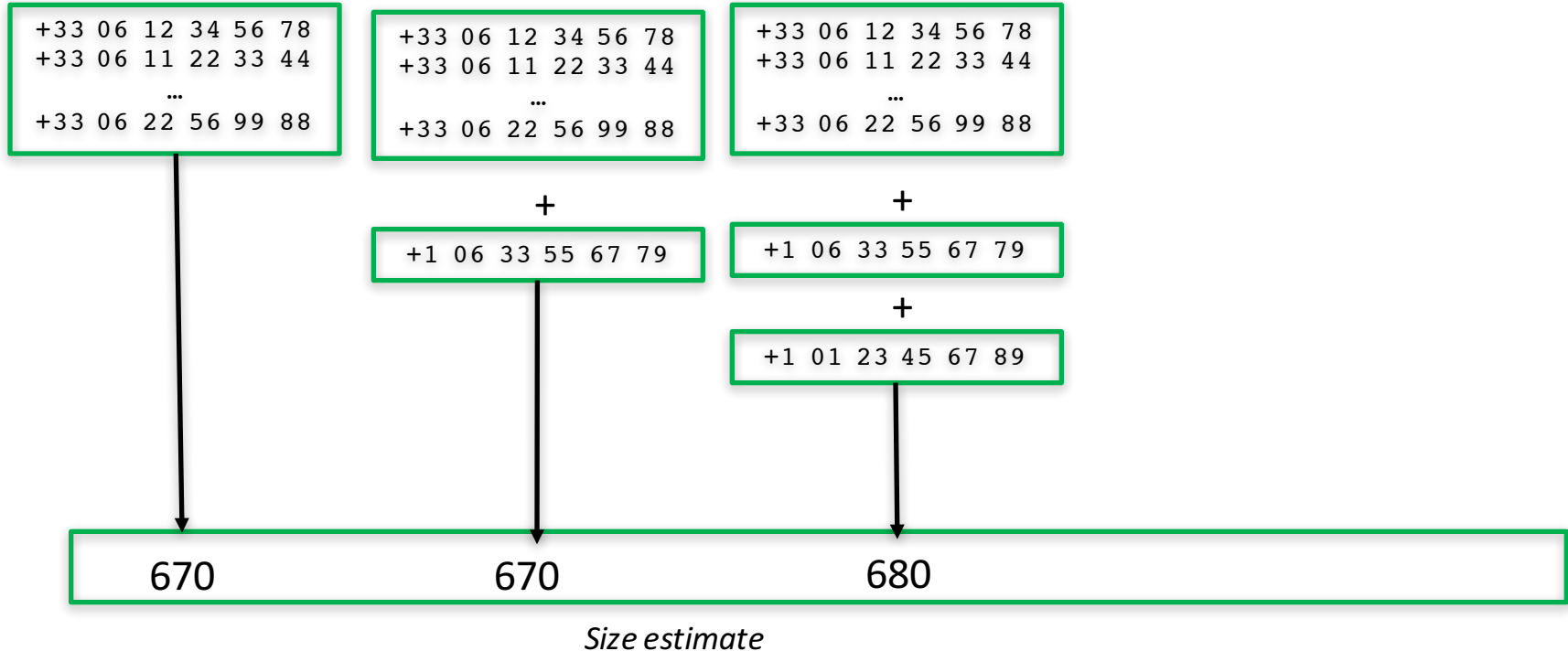
in

```
+33 06 12 34 56 78  
+33 06 11 22 33 44  
...  
+33 06 22 56 99 88
```

?

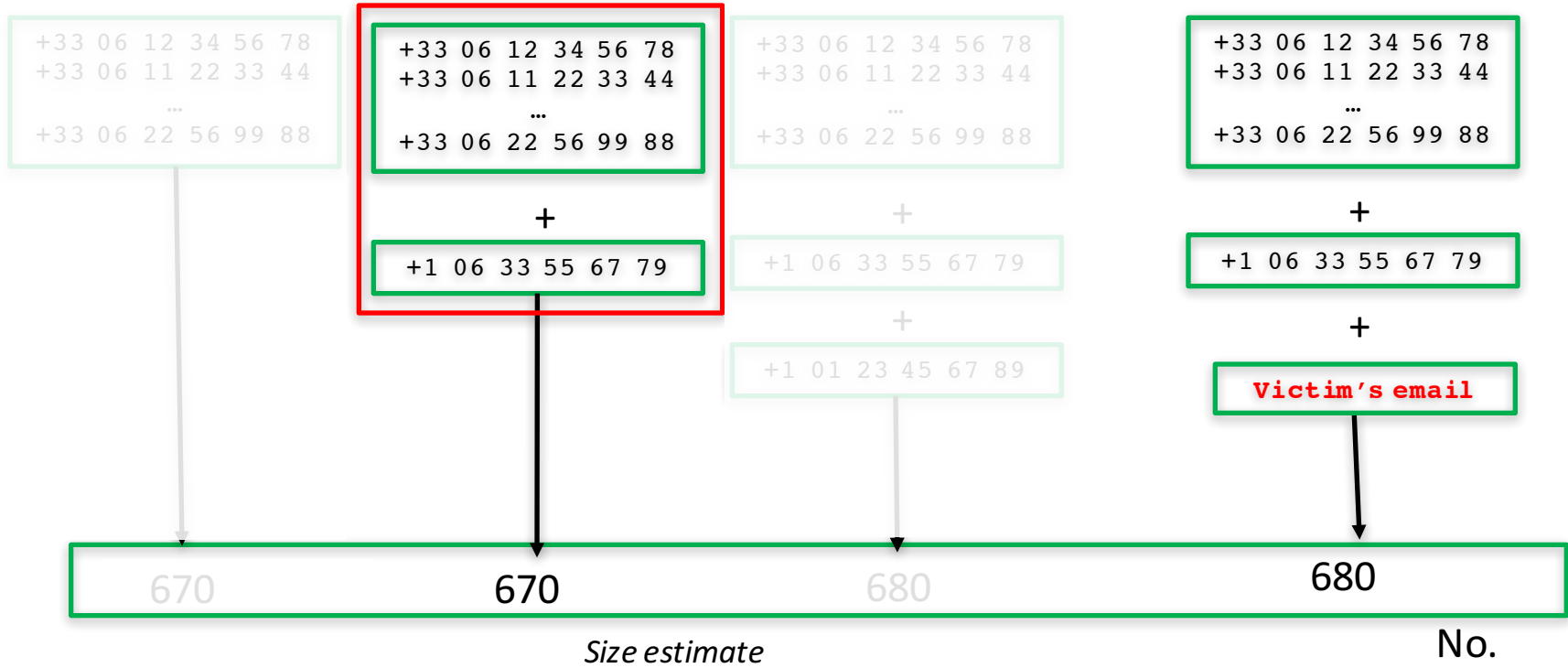
Exploiting these features

Goal: given victim's email address, find if victim is in a given **list of phone numbers**



Exploiting these features

Goal: given victim's email address, find if victim is in a given **list of phone numbers**



Attack: Inferring user's phone numbers

Can ask: Is **Victim** in Target list

Is **victim** in

100-000-0000
100-000-0001
100-000-0002
...
199-999-9999

If **No**: First digit is not 1
If **Yes**: First digit is 1

Is **victim** in

200-000-0000
200-000-0001
200-000-0002
...
299-999-9999

If **No**: First digit is not 2
If **Yes**: First digit is 2

Is **victim** in

010-000-0000
010-000-0001
010-000-0002
...
919-999-9999

If **No**: Second digit is not 1
If **Yes**: Second digit is 1

Attacks discovered

- Other attacks:
 - De-anonymise individual visitors to attacker's website
 - Infer PII for visitors to attacker's website en masse
- All attacks are easy to launch and have a large-scale impact:
 - Anyone can conduct an attack
 - Any Facebook user can be a potential victim
 - No interaction with the victim
- Responsibly disclosed with Facebook:
 - Temporarily removed size estimates for audiences with multi PII

Investigating ad transparency mechanisms in social media

[SEE SLIDES OF OANA GOGA]