



Privacy measurement and protection in Web applications

Nataliia Bielova

INDES project-team

Inria Sophia Antipolis – Méditerranée

Invited talk at Nice DSmeetup

24/10/2018

Targeted Advertising

The image illustrates targeted advertising by showing two overlapping web pages. The background page is Newchic, a fashion retailer, displaying various styles of boots. The foreground page is AccuWeather, showing a weather forecast for Lyon, France. A red rectangular box highlights a banner advertisement on the AccuWeather page for socofy boots, which are the same boots featured on the Newchic website. This demonstrates how advertising can be targeted to users based on their browsing history or interests.

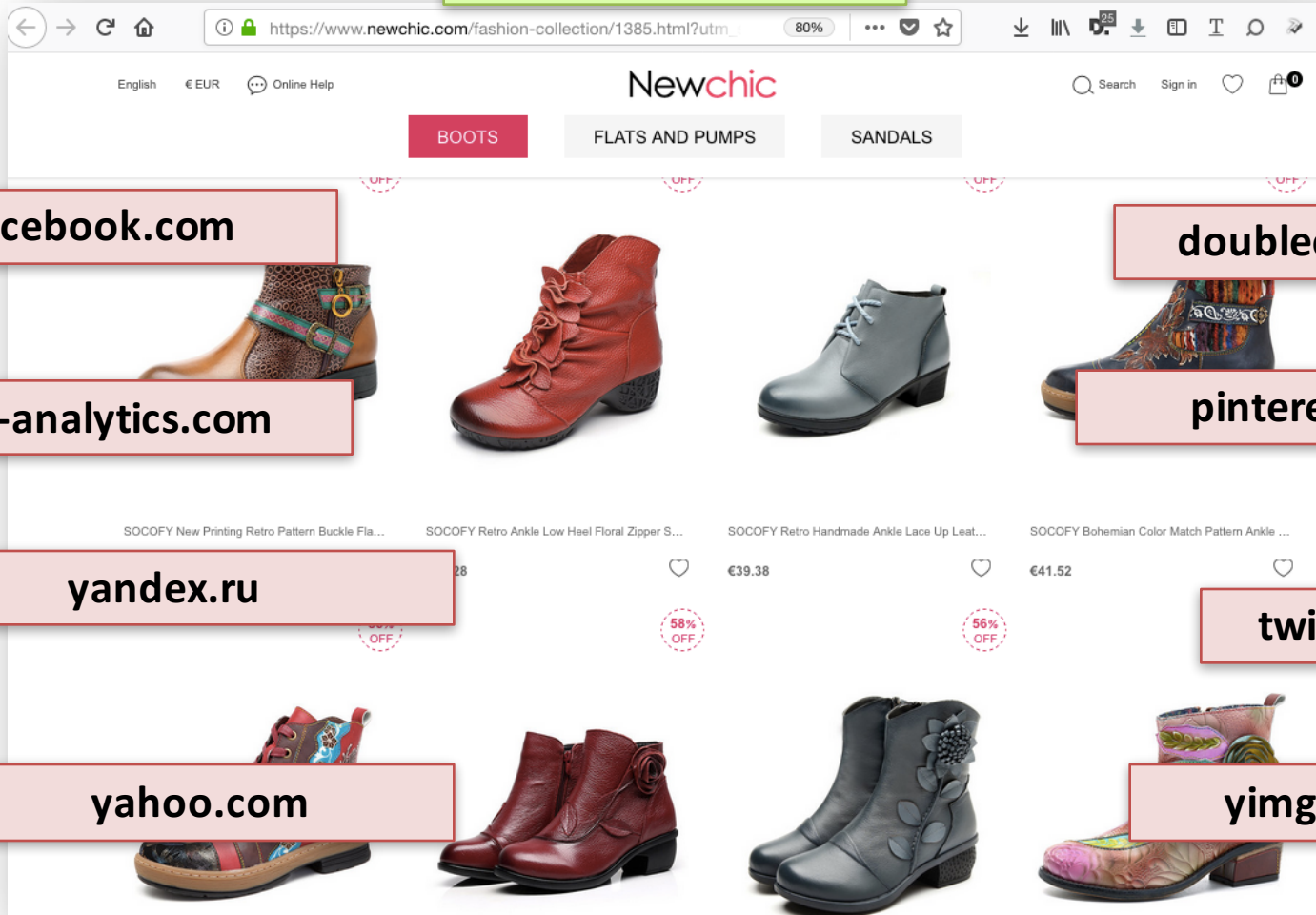
Newchic Website:

- URL: https://www.newchic.com/fashion-collection/1385.html?utm_
- Categories: BOOTS, FLATS AND PUMPS, SANDALS
- Product: SOCOFY New Printing Retro Pattern Buckle Fla... (€53.46) and SOCOFY Retro An... (€37.28)
- Discount: 56% OFF

AccuWeather Website:

- URL: <https://www.accuweather.com/en/fr/lyon/171210/weather-forec>
- Location: Lyon, France
- Current Weather: 18°C, Sunny
- Forecast:
 - TODAY APR 16: 19° Hi, RealFeel® 18°, A thunderstorm in spots
 - TONIGHT APR 16: 9° Lo, RealFeel® 9°, Turning out clear
 - TOMORROW APR 17: 22° Hi, RealFeel® 23°, Partly sunny and pleasant
- Advertisements:
 - socofy boots banner
 - catawiki watches

newchic.com



facebook.com

doubleclick.net

google-analytics.com

pinterest.com

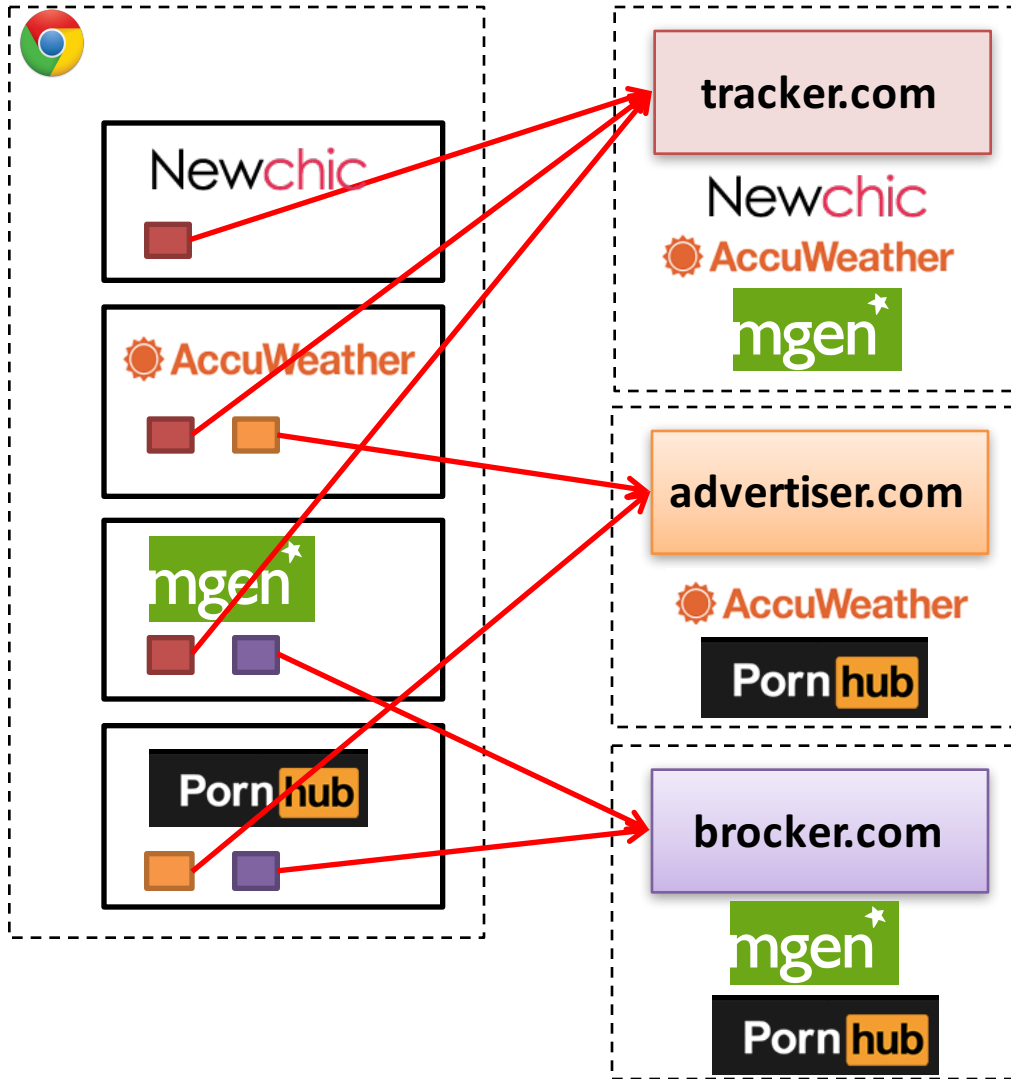
yandex.ru

twitter.com

yahoo.com

yimg.com

Web Tracking

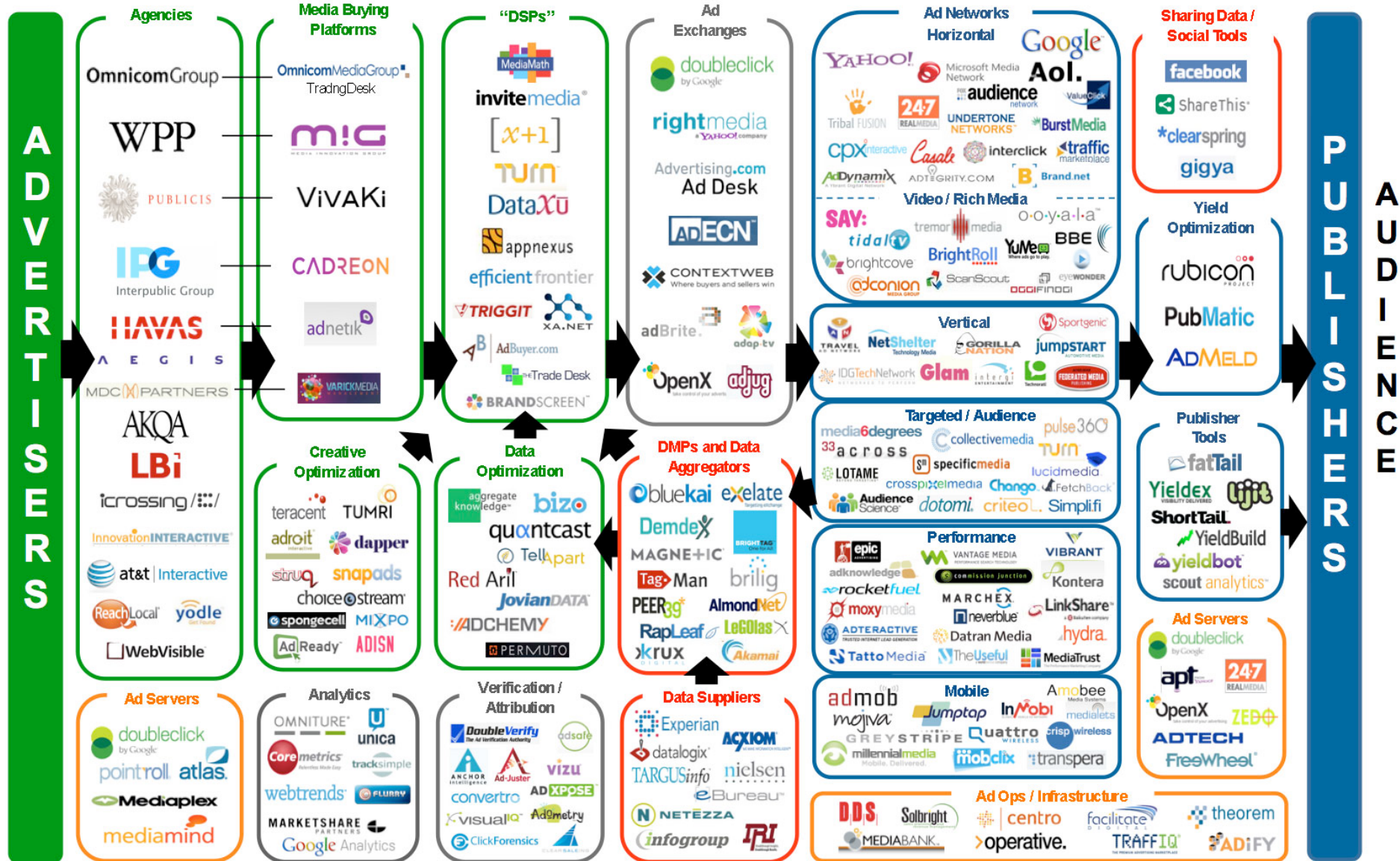


Tracking companies build bigger browsing profiles
= **increased value** for trackers
= **reduced privacy** for users

*Hypothetical relations

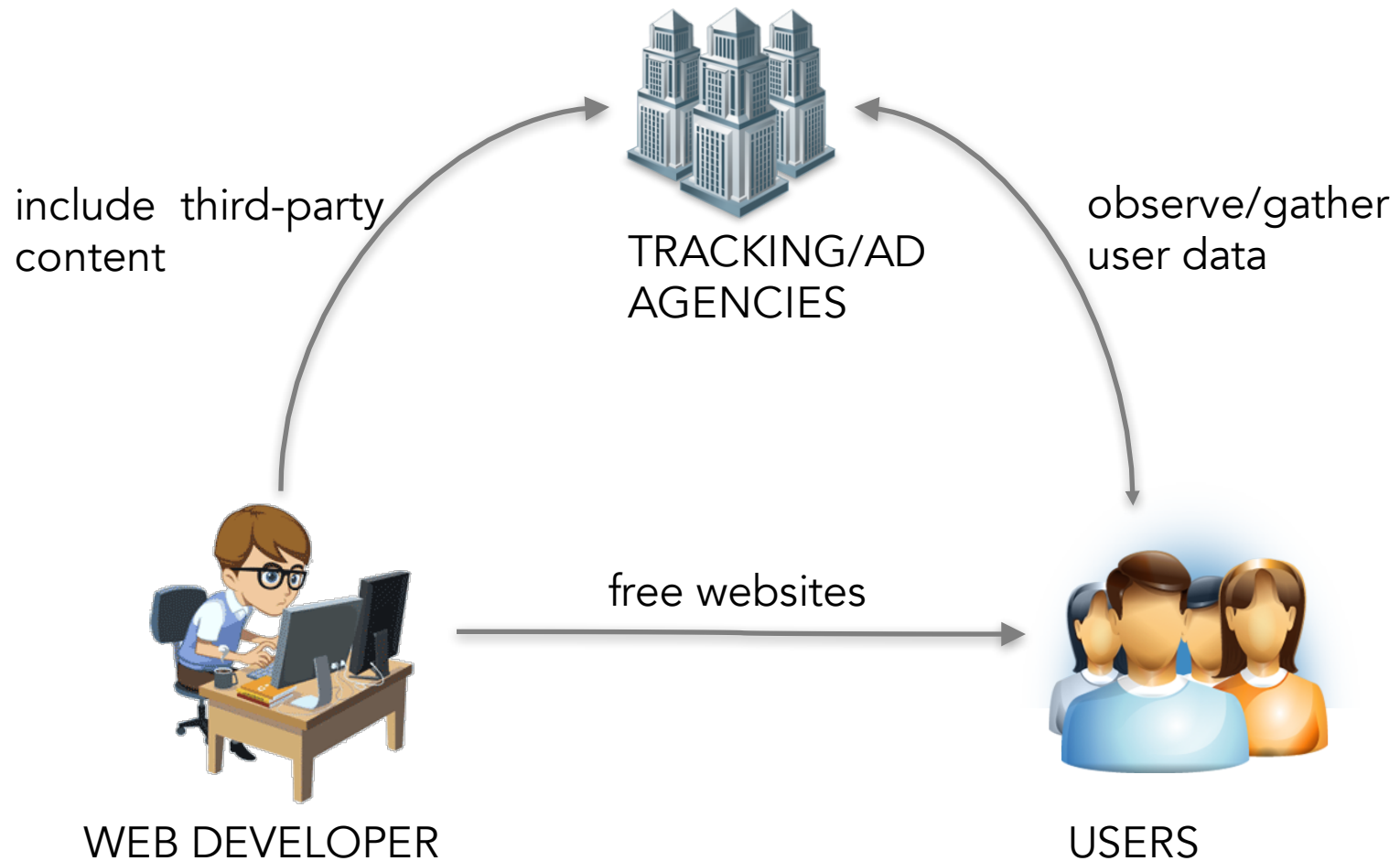
Display Advertising Technology Landscape

2010





Business model of the Web



Business model of the Web



include third-party
content



TRACKING/AD

observe/gather
user data

GDPR in force since May 25, 2018
ePrivacy Regulation under discussion



WEB DEVELOPER

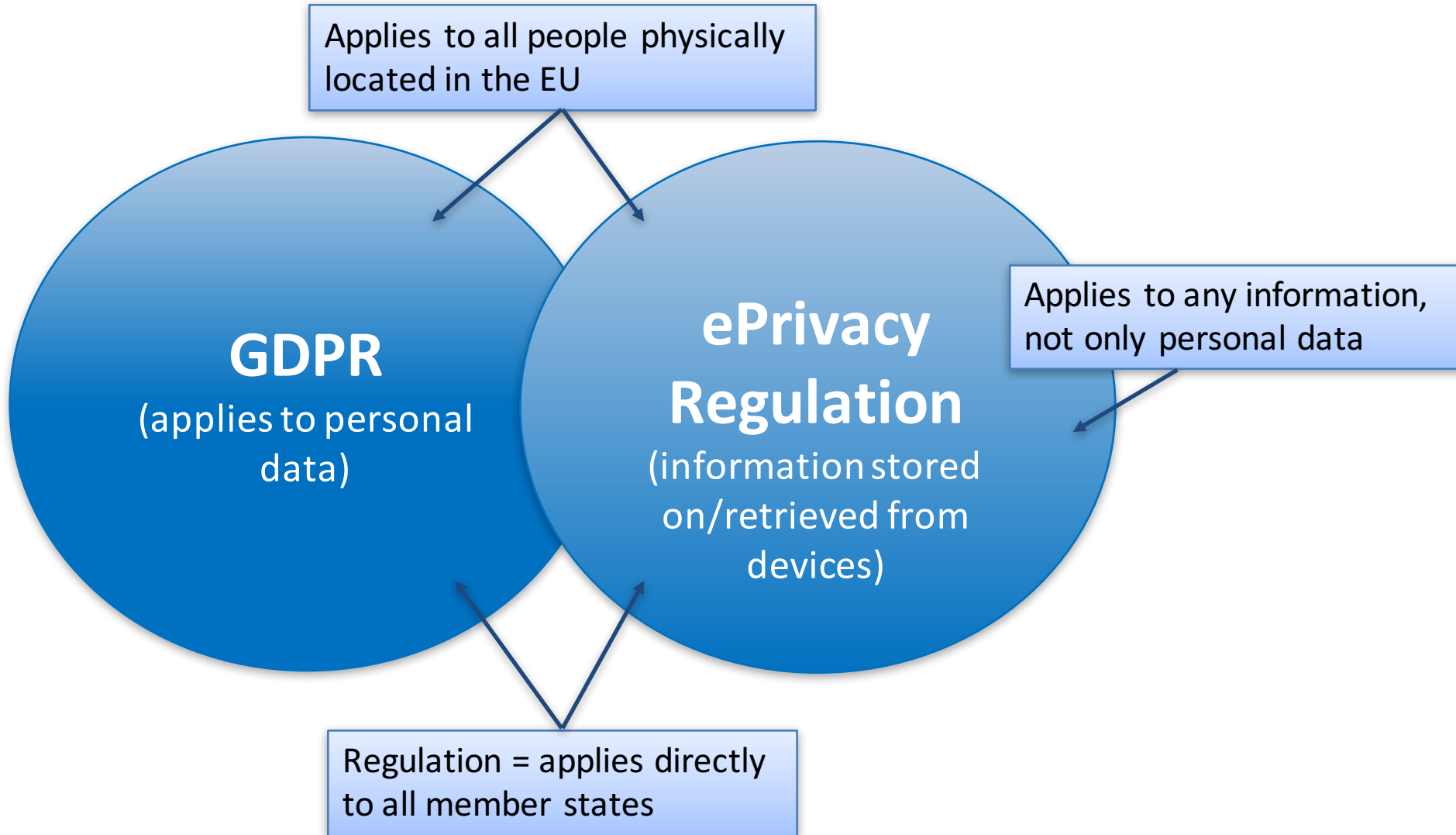
free websites



USERS



EU Data Protection Regulations



Why Web Tracking is important?

- Collection of our data without our knowledge
 - on sensitive websites
 - collection of our browsing patterns, preferences, tastes, even mood...



- **Usage of our data!**
 - targeted advertisement
 - **manipulation**



Don't browser extensions solve it?



AdBlockPlus: blocks scripts/requests **only from known advertising companies**

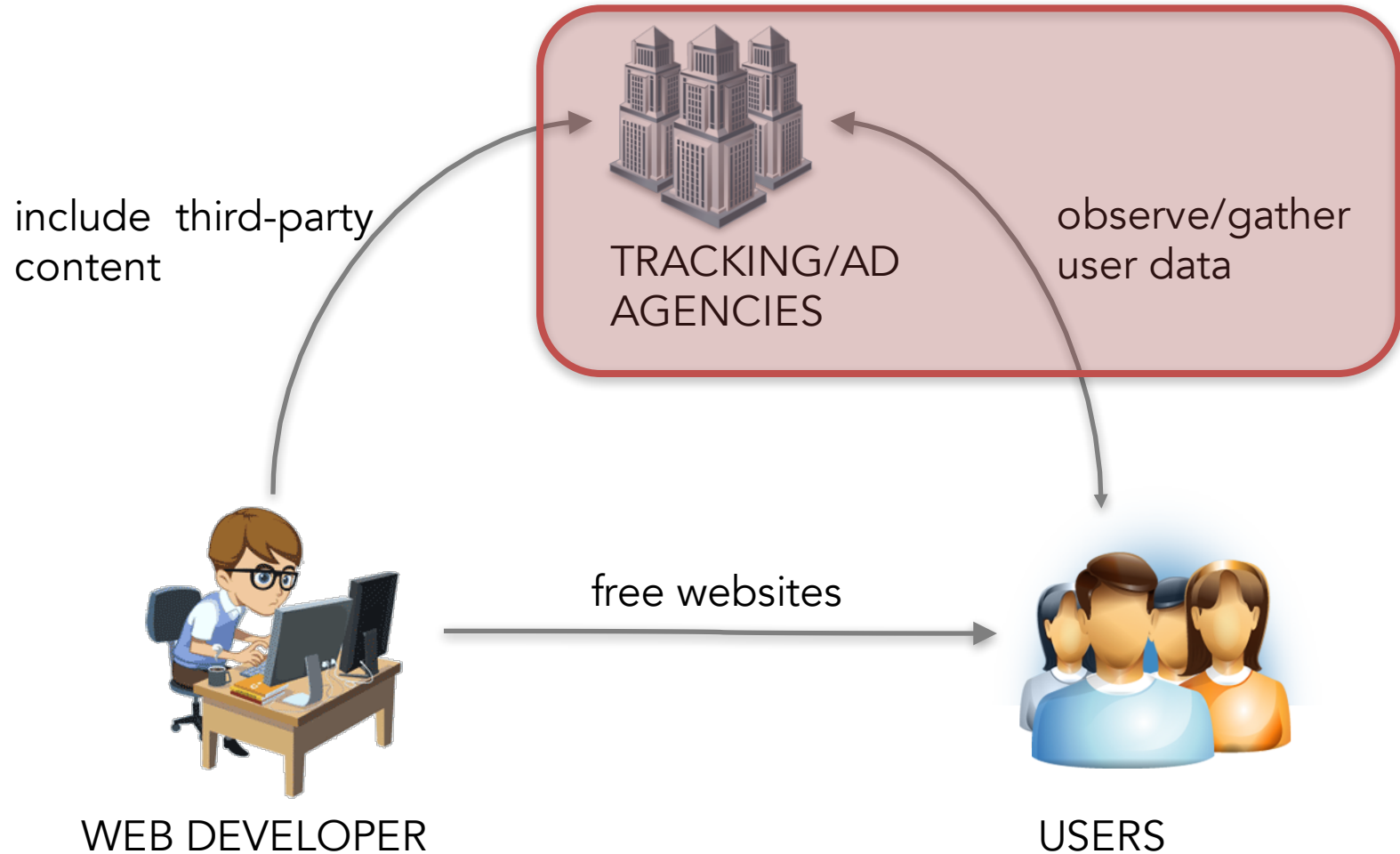


Ghostery: blocks scripts/requests **only from known tracking companies**

- They don't protect from tracking
 - by other companies
 - by the main website
 - from cookie synchronization

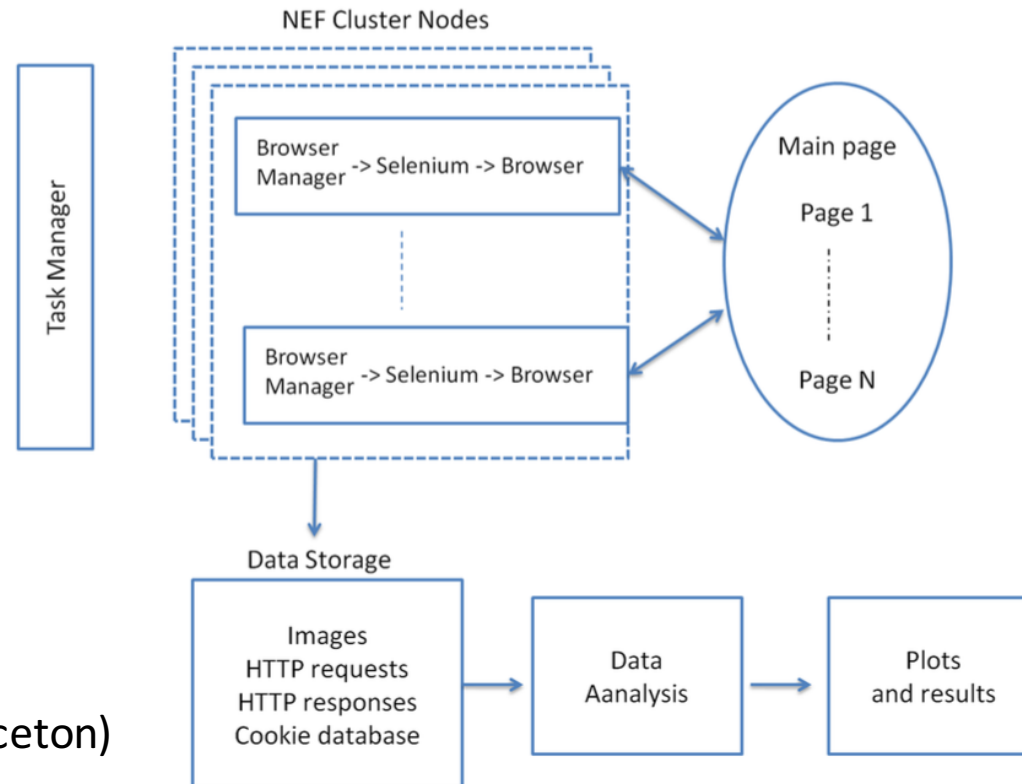
Research Challenges

Detection and Measurement



Distributed crawling

- Crawling infrastructure to **simulate user visits to websites**
- First distributed crawling with Nef cluster (Inria Sophia)
- 20 browsers per node

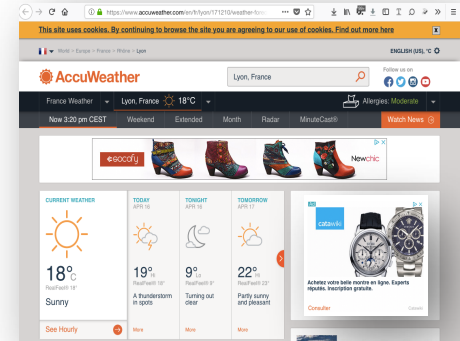


Based on OpenWPM (Princeton)

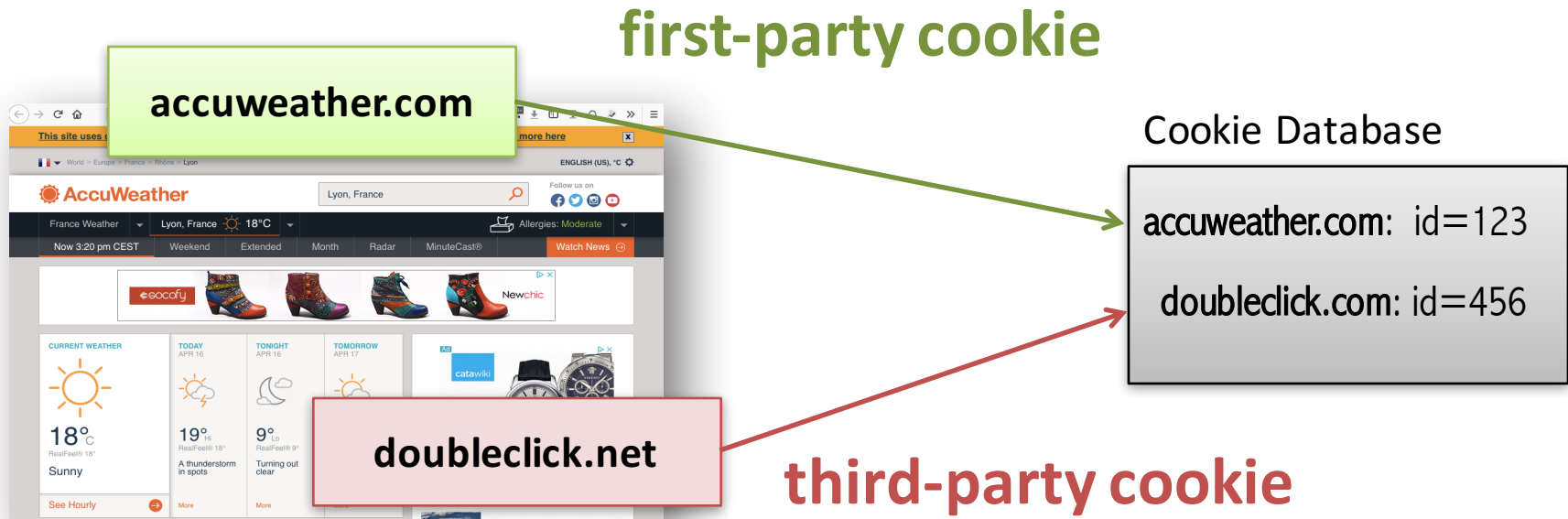
Nataliia Bielova (INDES)

Measuring Web Tracking

- 100,000 domains crawled
- 845,575 pages crawled
- Goal: **analyze data collection and data sharing practices** between third-party companies

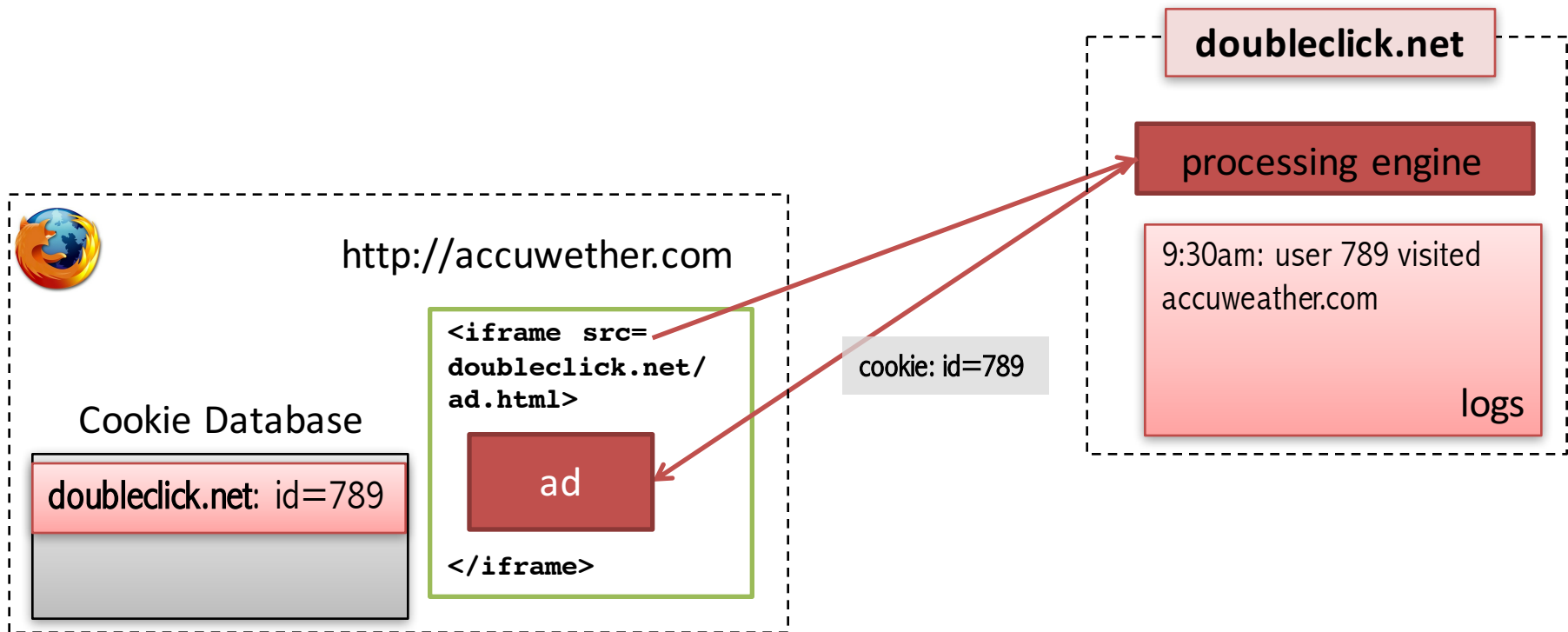


Cookies: first- and third-party



Cross-site Tracking

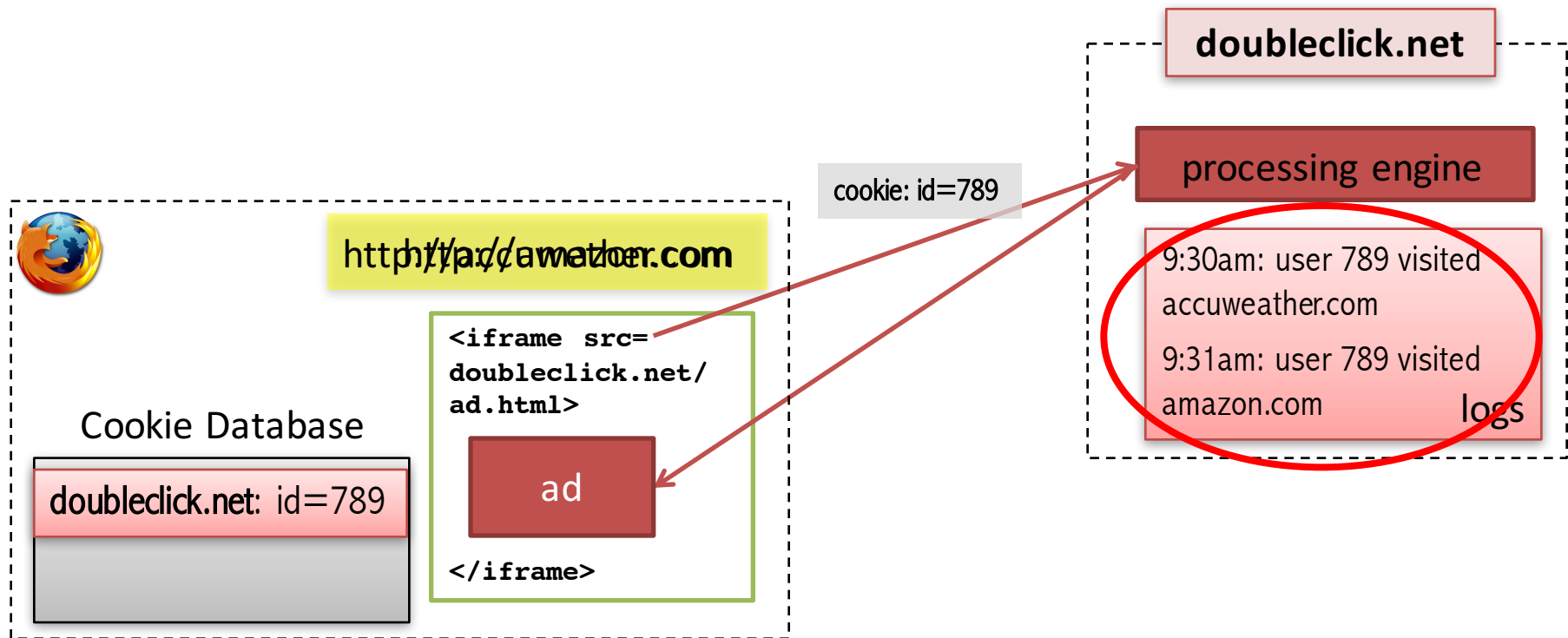
Third-party cookies are used to track users **across sites** and collect their browsing history.



*Hypothetical relations

Cross-site Tracking

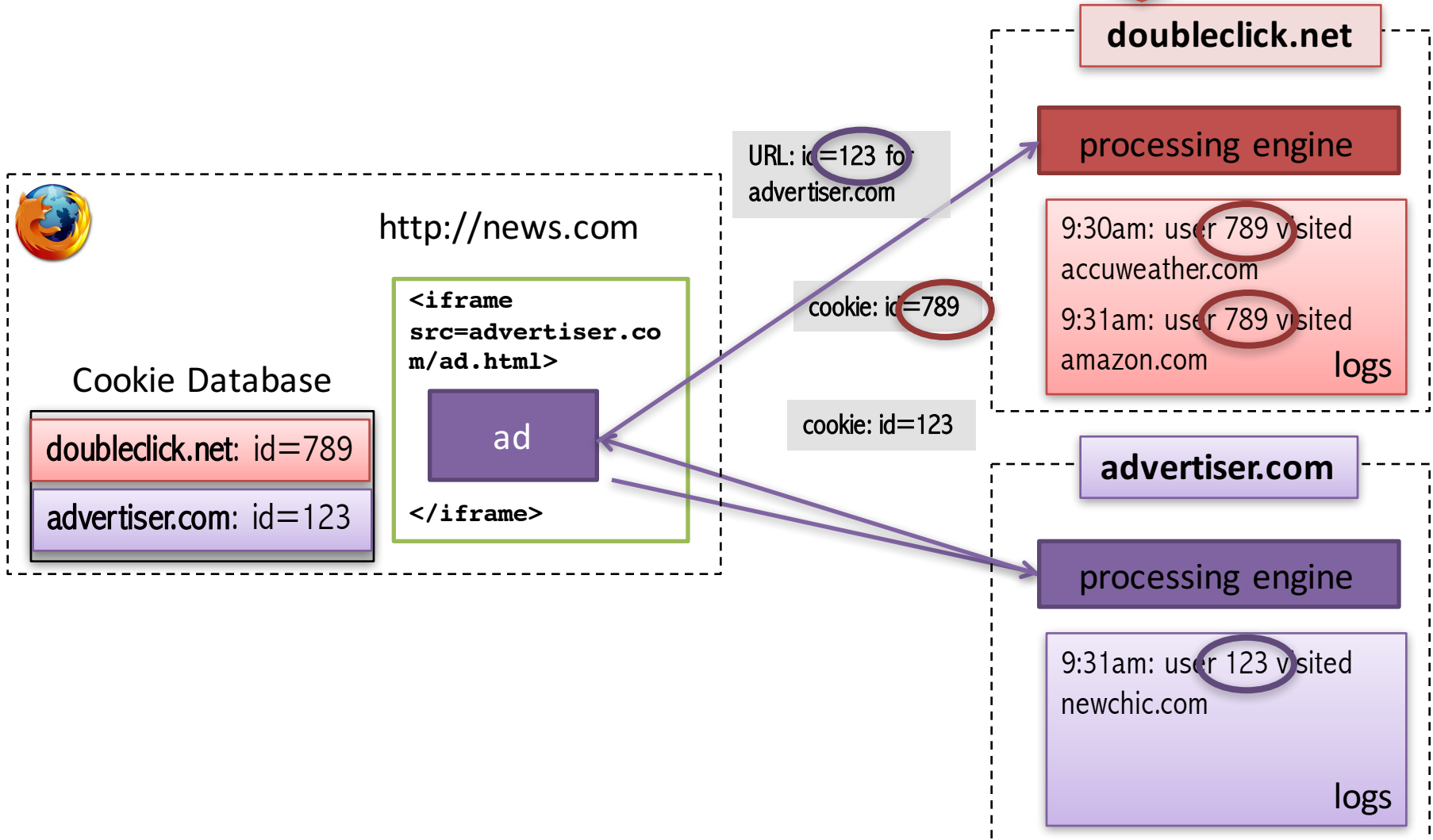
Third-party cookies are used to track users **across sites** and collect their browsing history.



*Hypothetical relations

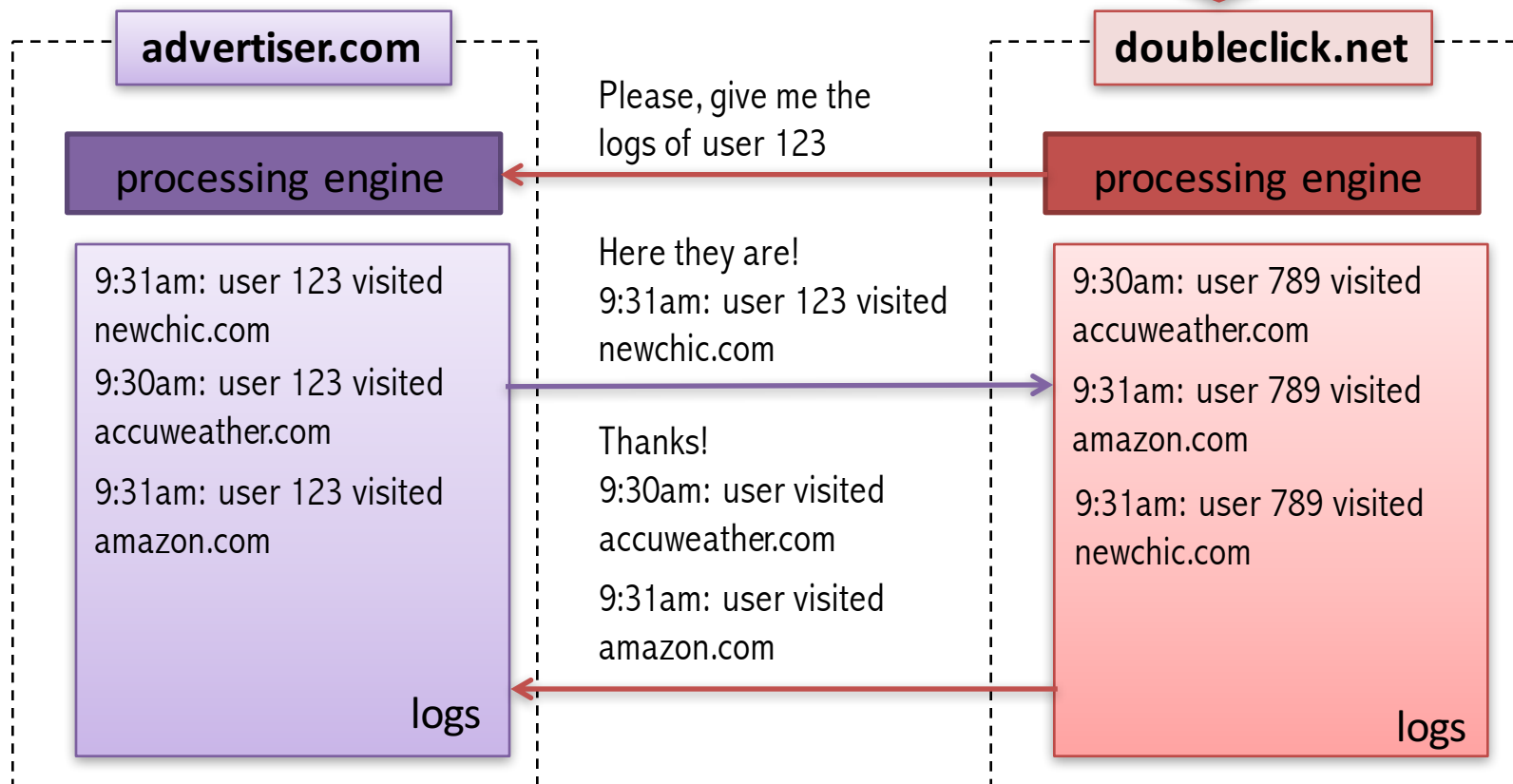
Cookie synchronization

User 789 has identifier 123 for advertiser.com!



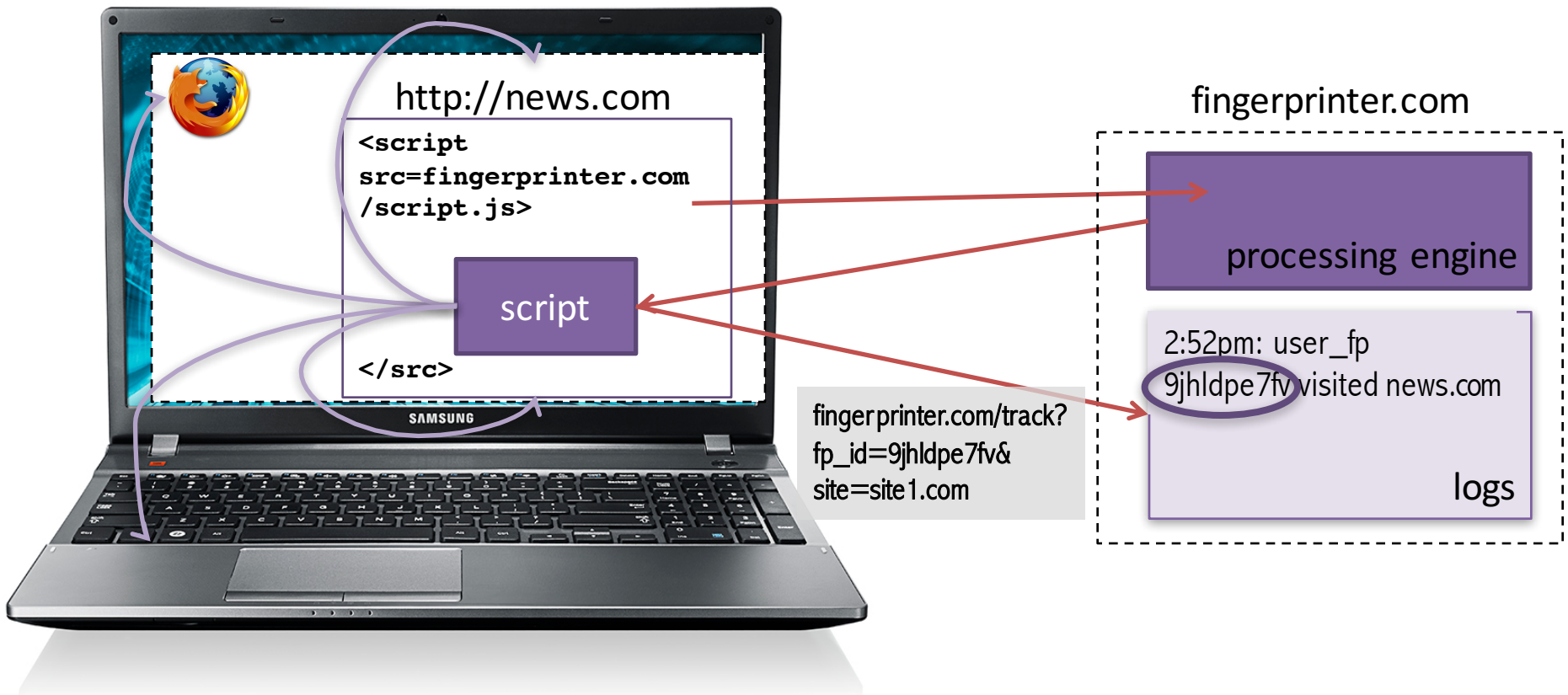
Cookie synchronization

User 789 has identifier 123 for advertiser.com!



Browser fingerprinting

Browser and operating system properties are used to track repeated visits to a site.



Example of a browser fingerprint

Attribute	Value
User agent	Mozilla/5.0 (X11; Fedora; Linux x86_64; rv:55.0) Gecko/20100101 Firefox/55.0
HTTP headers	text/html, application/xhtml+xml, application/xml;q=0.9,*/*;q=0.8 gzip, deflate, br en-US,en;q=0.5
Plugins	Plugin 0: QuickTime Plug-in 7.6.6; libtotem-narrow-space-plugin.so; Plugin 1: Shockwave Flash; Shockwave Flash 26.0 r0; libflashplayer.so.
Fonts	Century Schoolbook, Source Sans Pro Light, DejaVu Sans Mono, Bitstream Vera Serif, URW Palladio L, Bitstream Vera Sans Mono, Bitstream Vera Sans, ...
Platform	Linux x86_64
Screen resolution	1920x1080x24
Timezone	-480 (UTC+8)
OS	Linux 3.14.3-200.fc20.x86 32-bit
WebGL vendor	NVIDIA Corporation
WebGL renderer	GeForce GTX 650 Ti/PCIe/SSE2
Canvas	Cwm fjordbank glyphs vext quiz, ☺ Cwm fjordbank glyphs vext quiz, ☺



Maverick
Ocean Front Villas
Mandarin Sea
Regency
Sassafras & Ginger
Dollhouse
Athletics Dept.



Behavioral fingerprinting

Are you what you install to your computer?



Boda et al.: User Tracking on the Web via Cross-Browser Fingerprinting (2011)

[Google.com](https://www.google.com) [Youtube.com](https://www.youtube.com)
[Facebook.com](https://www.facebook.com) [Baidu.com](https://www.baidu.com)
[Yahoo.com](https://www.yahoo.com) [Wikipedia.org](https://www.wikipedia.org)
[Google.co.in](https://www.google.co.in) [Qq.com](https://www.qq.com) [Sohu.com](https://www.sohu.com)
[Google.co.jp](https://www.google.co.jp) [Taobao.com](https://www.taobao.com)
[Tmall.com](https://www.tmall.com) [Live.com](https://www.live.com) [Amazon.com](https://www.amazon.com)
[Vk.com](https://www.vk.com) [Twitter.com](https://www.twitter.com)
[Instagram.com](https://www.instagram.com) [360.cn](https://www.360.cn)

Are you what sites you have visited?

Su et al.: De-anonymizing Web Browsing Data with Social Networks (2017)

Nataliia Bielova (INDES)

based on slides of Gábor György Gulyás

22

Behavioral fingerprinting – wait, what?!

- GDPR Art. 4. (14) on **biometric data**:

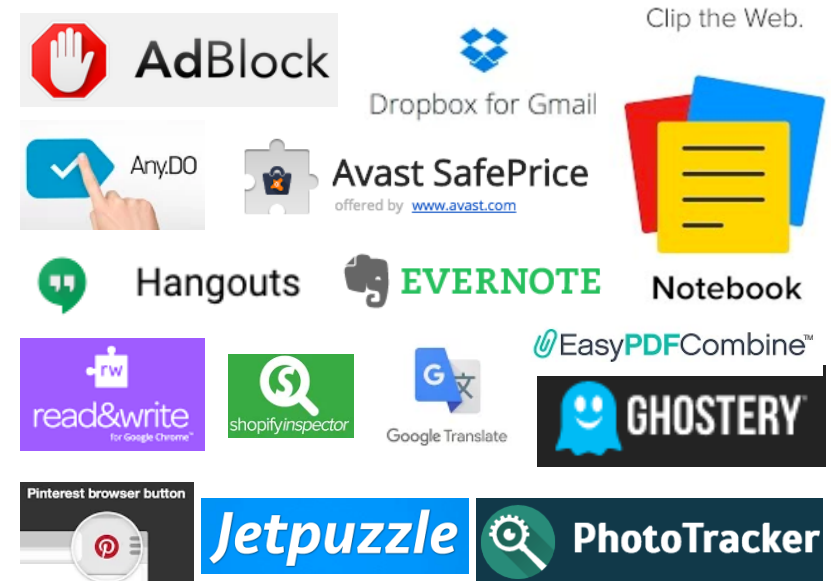
„means personal data resulting from specific technical processing relating to [...] behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, [...]”

- Should have a **huge impact** on many data use cases:
 - fingerprinting,
 - uses of location data,
 - user preferences/likes,
 - ...



Extension detection

- Detecting web accessible resources of extensions
- ~16 000 extensions detected (~28% of all Chrome extensions)
- No false-positives (disabled extensions not reported)



chrome-extension://mlomiejdfkolichcflejclcbmpeanii/app/images/apps_pages/tracker.png

Extension ID
(Ghostery)

Local filepath

Detecting web logins

- Methods:
 - Redirection URL hijacking
 - Misusing CSP report
 - 58 popular websites were covered



<https://inria.fr/login?return=logo.png>



Logged in
(silent & unchecked
redirection to image)

Not logged in
(login page)





Browser Extension and Login-Leak Experiment

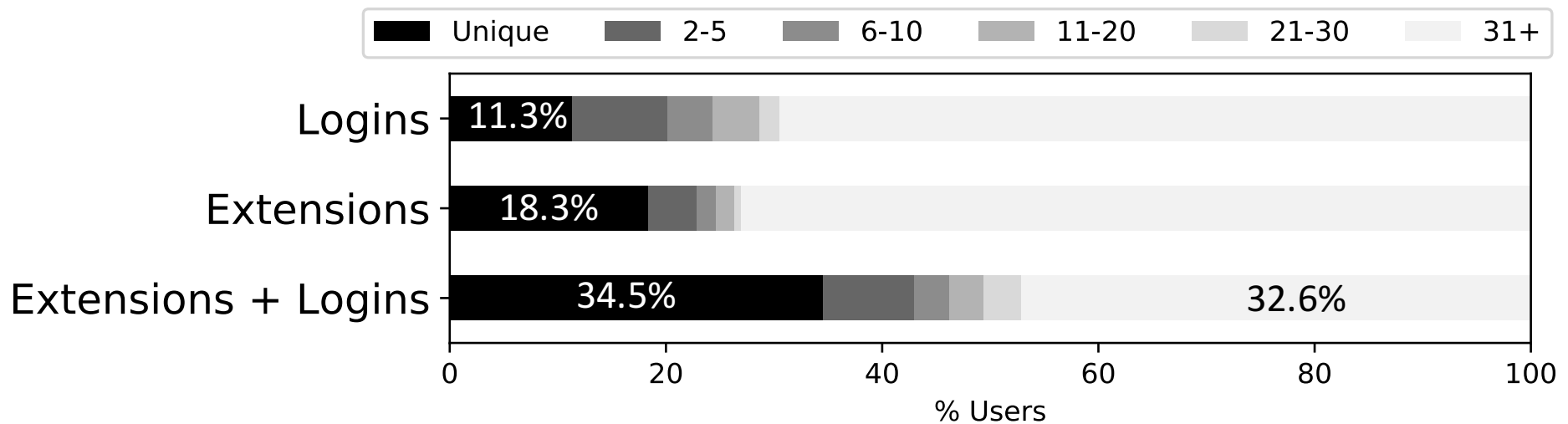
Are you identifiable?

Yes, you are identifiable, as there are no other users who looks like you among the 22308 users we tested so far:



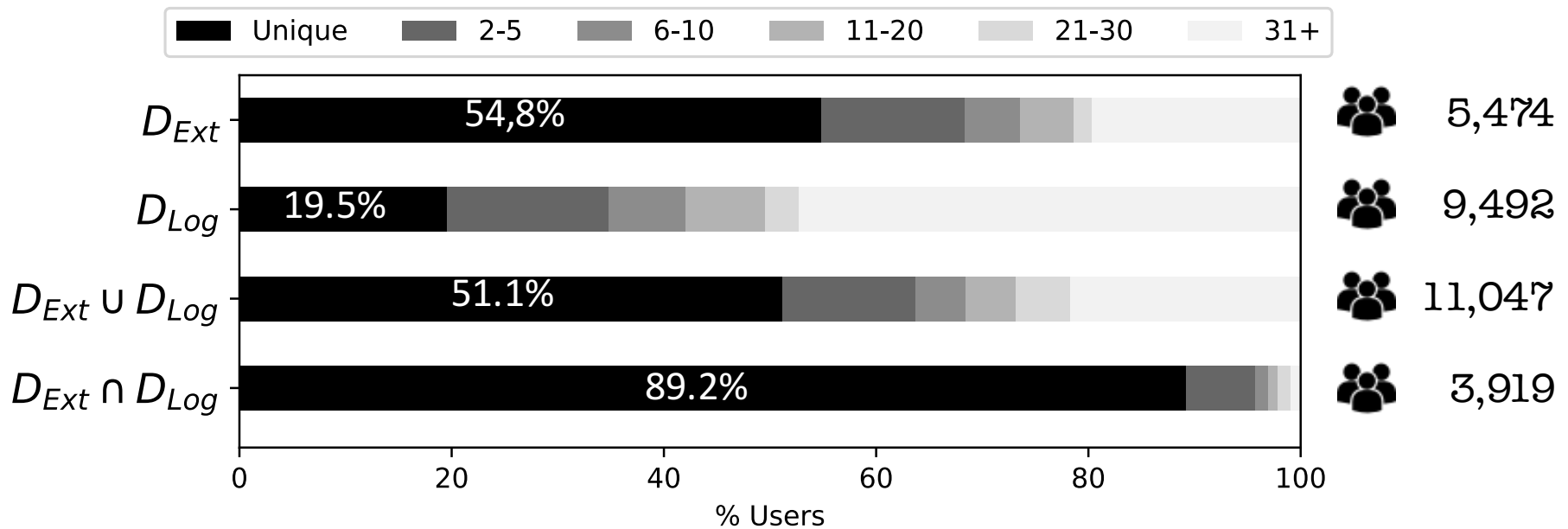
<https://extensions.inrialpes.fr>

How unique you are? (full dataset)



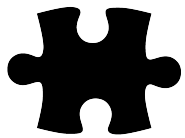
 16,393 users

How unique you are? (sub-datasets)

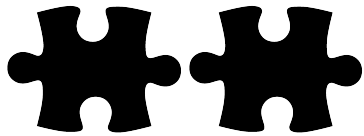


G.G. Gulyás, D. F. Some, **N. Bielova** and C. Castelluccia. To Extend or not to Extend: on the Uniqueness of Browser Extensions and Web Logins. *WPES@ACM CCS 2018*.

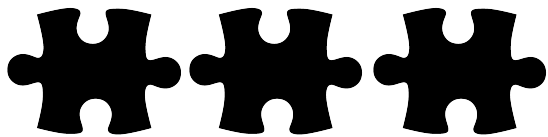
Am I unique if I use a few extensions?



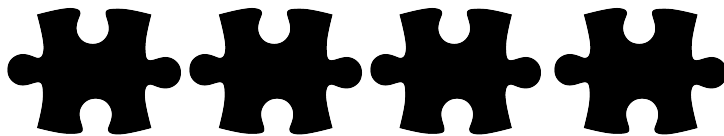
54.86% unique



76.25% unique



92.22% unique

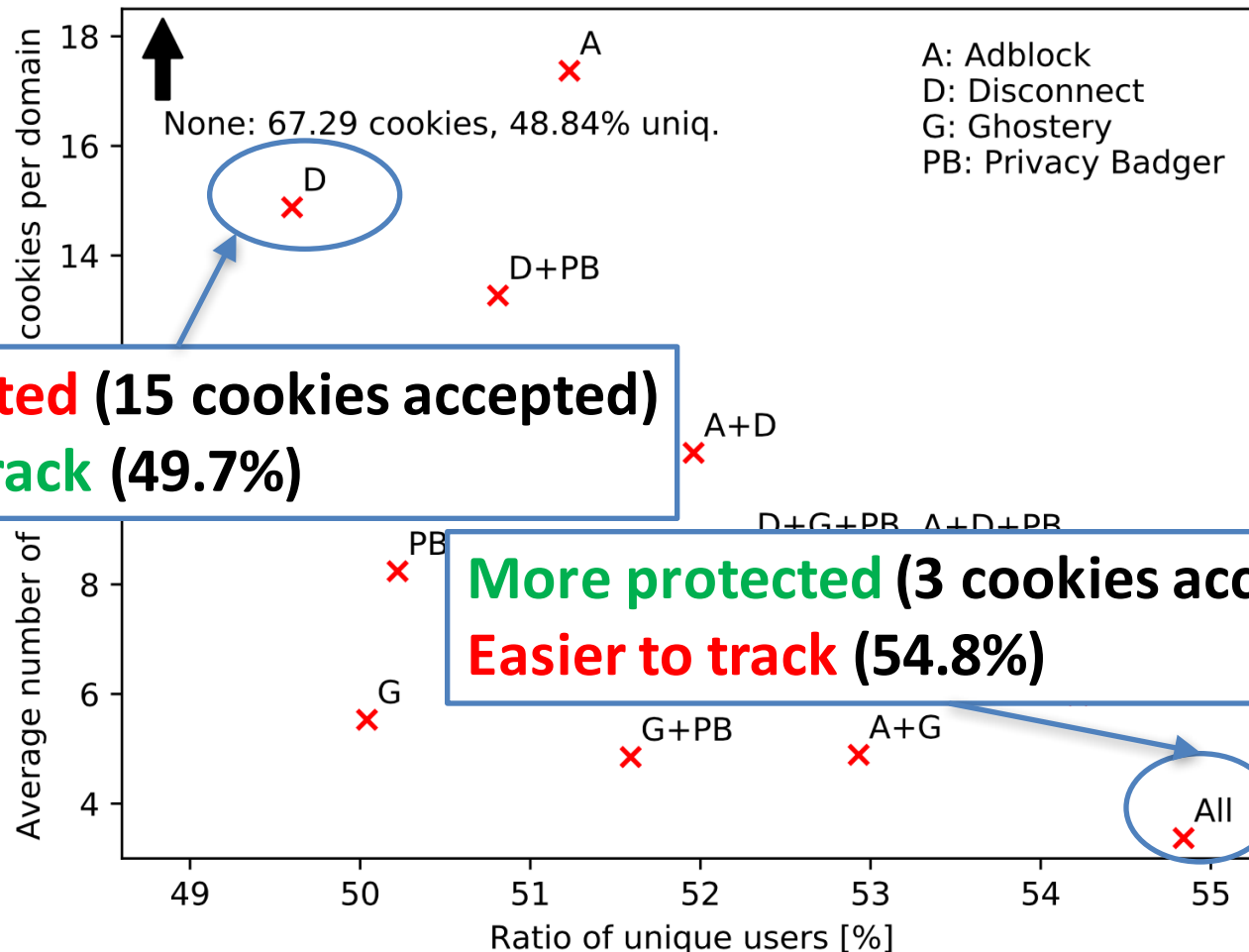


95.85% unique

The dilemma of privacy extensions

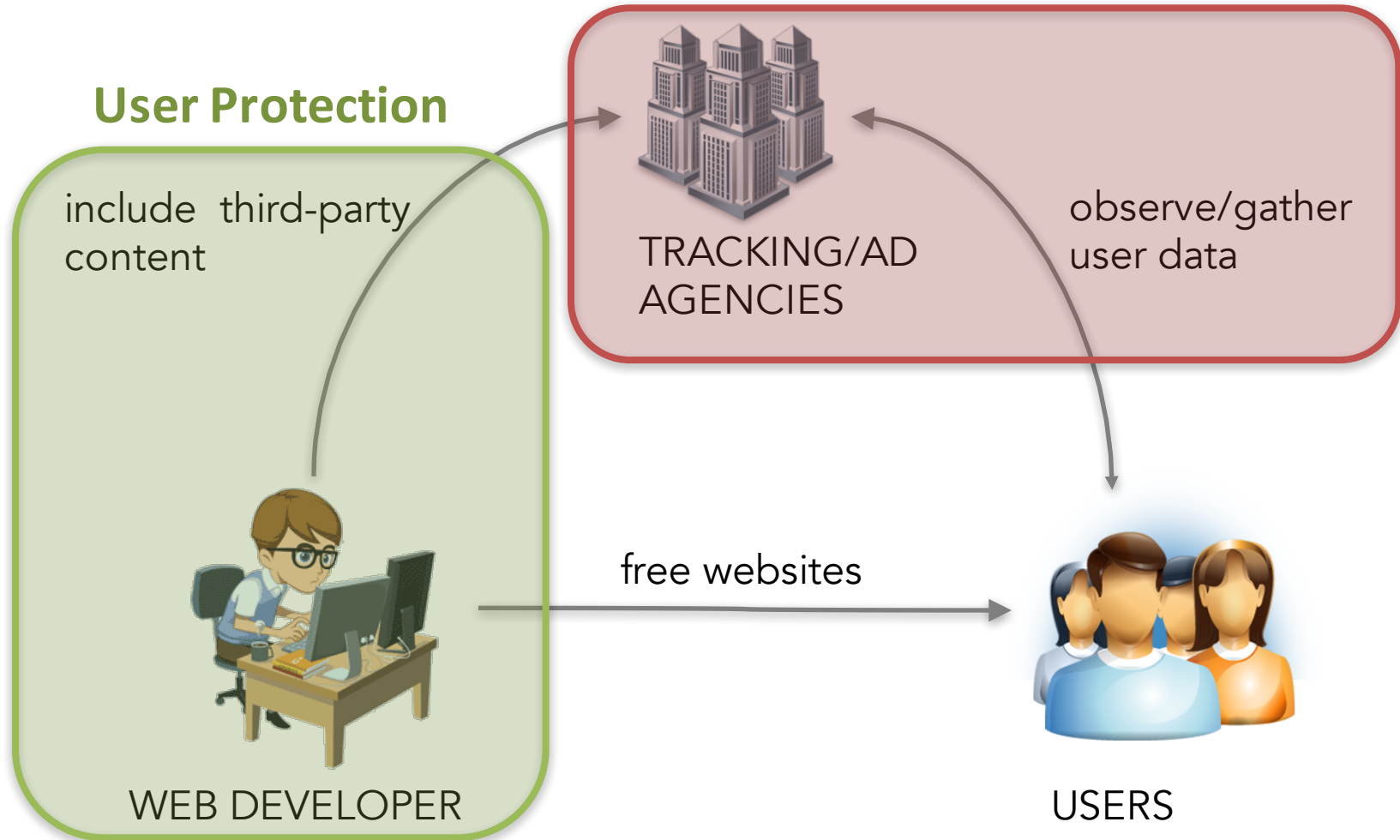
- Privacy extensions **block some trackers**
- Privacy extensions **make a user more unique**
- What is the trade-off between **privacy gain** (some trackers are blocked) and **privacy loss** (user is more unique)?

Uniqueness of users vs. number of accepted third-party cookies

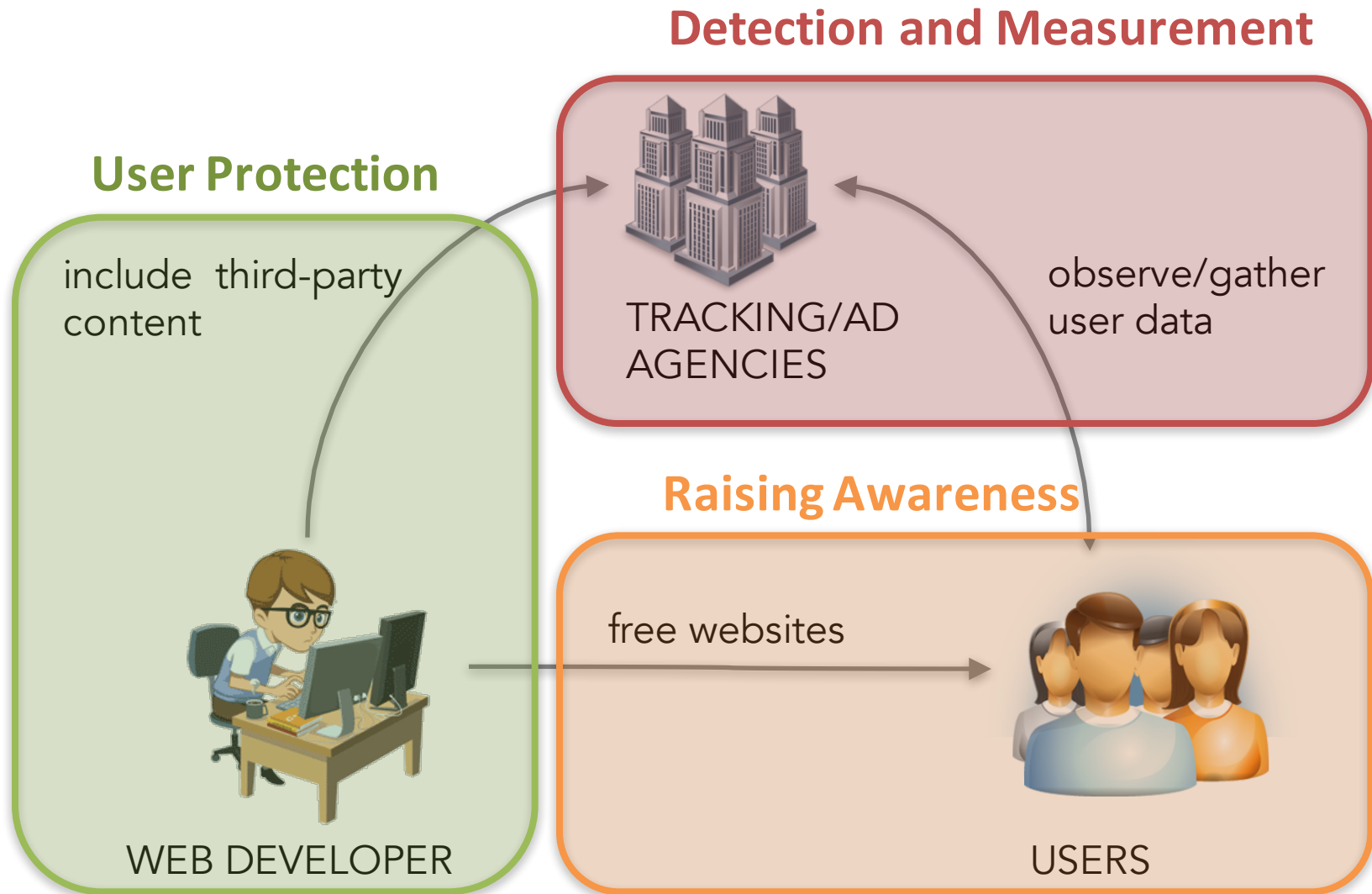


Research Challenges

Detection and Measurement



Research Challenges



<https://www.fun-mooc.fr/courses/course-v1:inria+41015+session02/about>

Home • All courses • Protection de la vie privée dans le monde numérique



Protection de la vie privée dans le monde numérique

Thematics

Droit et juridique

Numérique, technologie

Registration is open!
Classes start: Nov 5, 2018
Classes end: Jan 7, 2019

Inria



Once you are registered, videos are available even after the end of the course!



Cédric Lauradoux, Vincent Roca, **Nataliia Bielova** avec l'aide de l'équipe [Inria Learning Lab](#). MOOC *Protection de la vie privée dans le monde numérique*.

Conclusions

tracker.com: id=789

- **Detect and measure Web Tracking** & analyse when Web tracking becomes illegal under **GDPR and ePrivacy**
- **Evaluate new fingerprinting methods** via browser extensions
- **Help Web developers** to transform their Web applications into privacy-compliant applications
- **Raise awareness** about tracking and protection
- **Next:** EU Data Protection laws in the context of IoT

