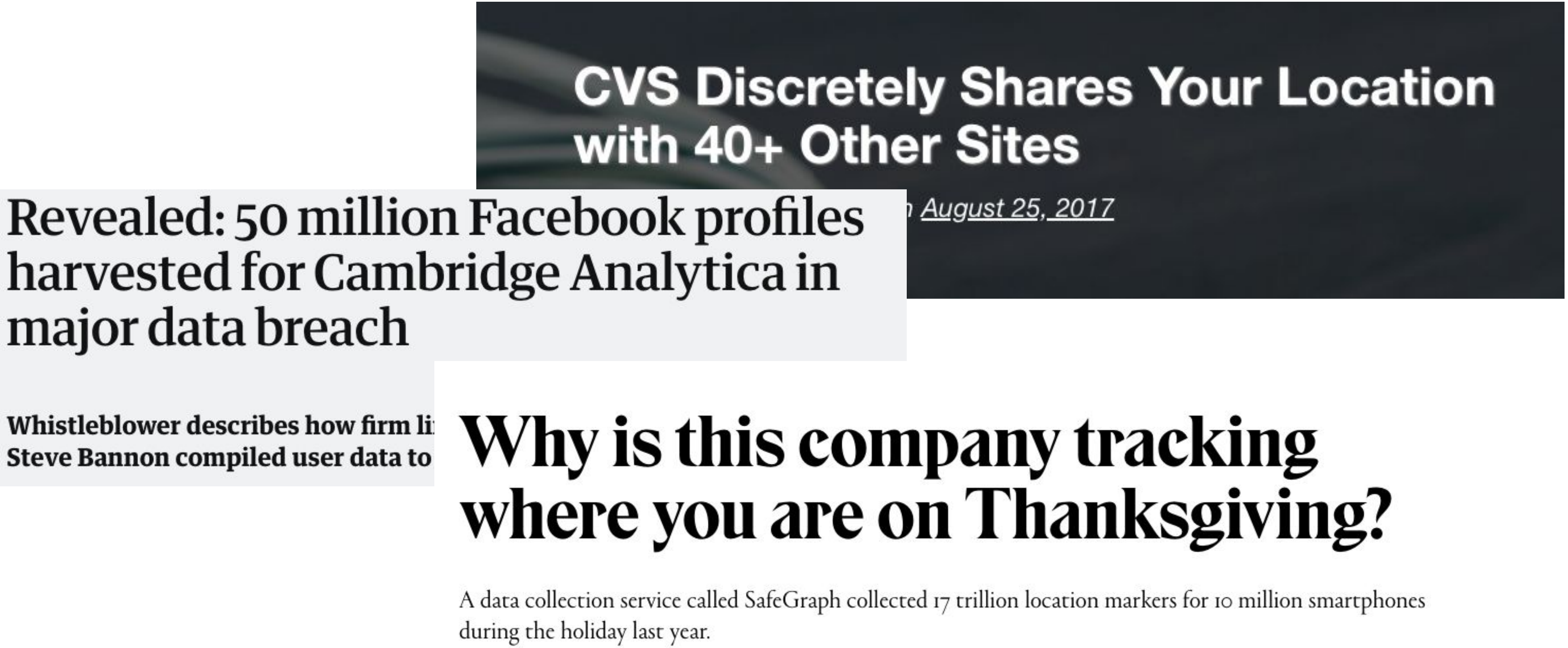
A Global Study of the Mobile Tracking Ecosystem (NDSS18)

Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, Phillipa Gill

Presenter: Xueqing Liu

Mobile Tracking



CVS Discretely Shares Your Location with 40+ Other Sites

n August 25, 2017

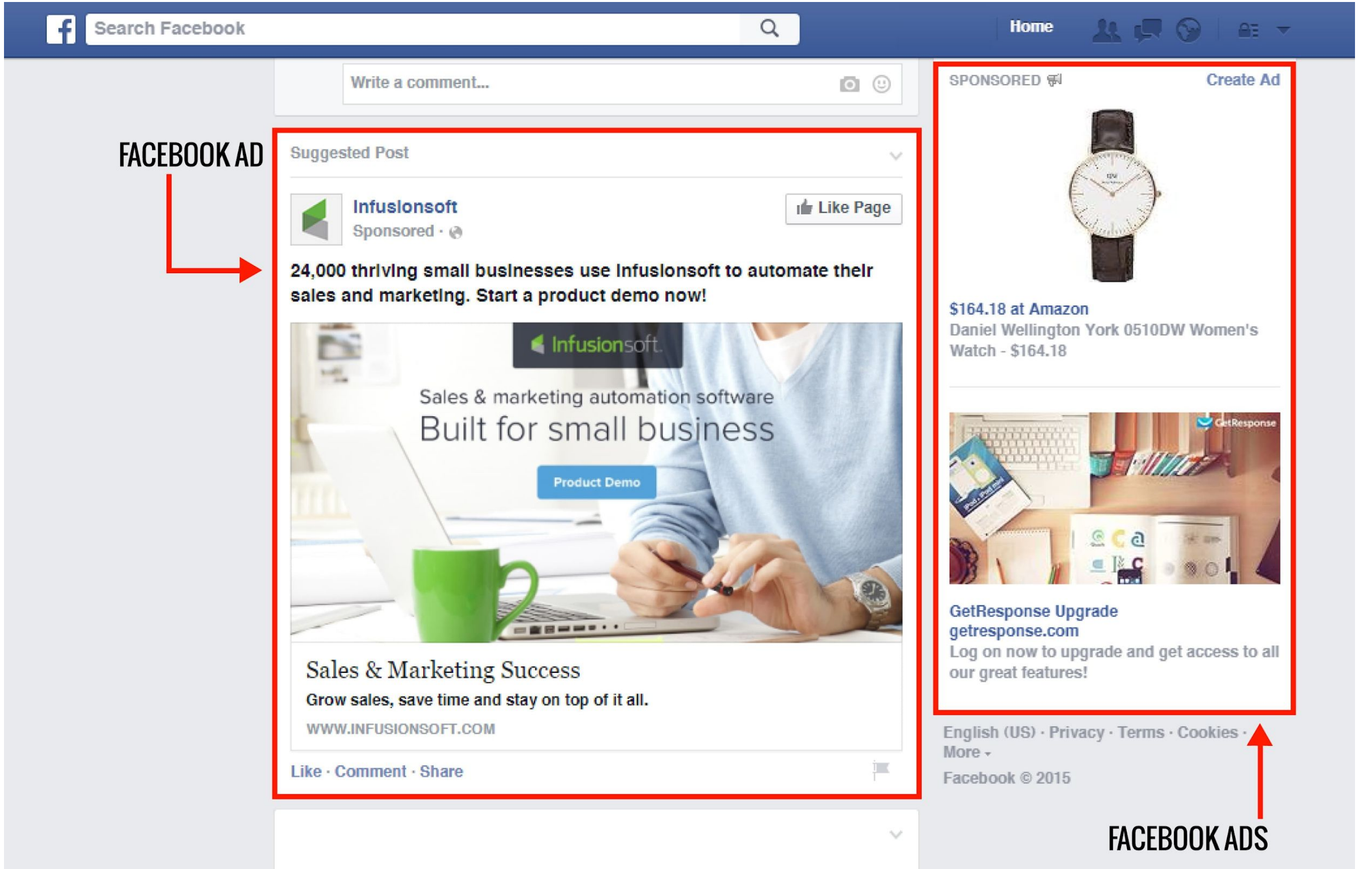
Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

Whistleblower describes how firm li
Steve Bannon compiled user data to

Why is this company tracking where you are on Thanksgiving?

A data collection service called SafeGraph collected 17 trillion location markers for 10 million smartphones during the holiday last year.

Mobile Tracking



Search Facebook

Home

Write a comment...

FACEBOOK AD

Suggested Post

Infusionsoft

Sponsored ·

Like Page

24,000 thriving small businesses use Infusionsoft to automate their sales and marketing. Start a product demo now!

Sales & Marketing Success

Grow sales, save time and stay on top of it all.

WWW.INFUSIONSOFT.COM

Like · Comment · Share

SPONSORED

Create Ad

$164.18 at Amazon

Daniel Wellington York 0510DW Women's Watch - $164.18

GetResponse Upgrade

getresponse.com

Log on now to upgrade and get access to all our great features!

English (US) · Privacy · Terms · Cookies · More

Facebook © 2015

FACEBOOK ADS

How Are Users Tracked by **Third-Party Services**?

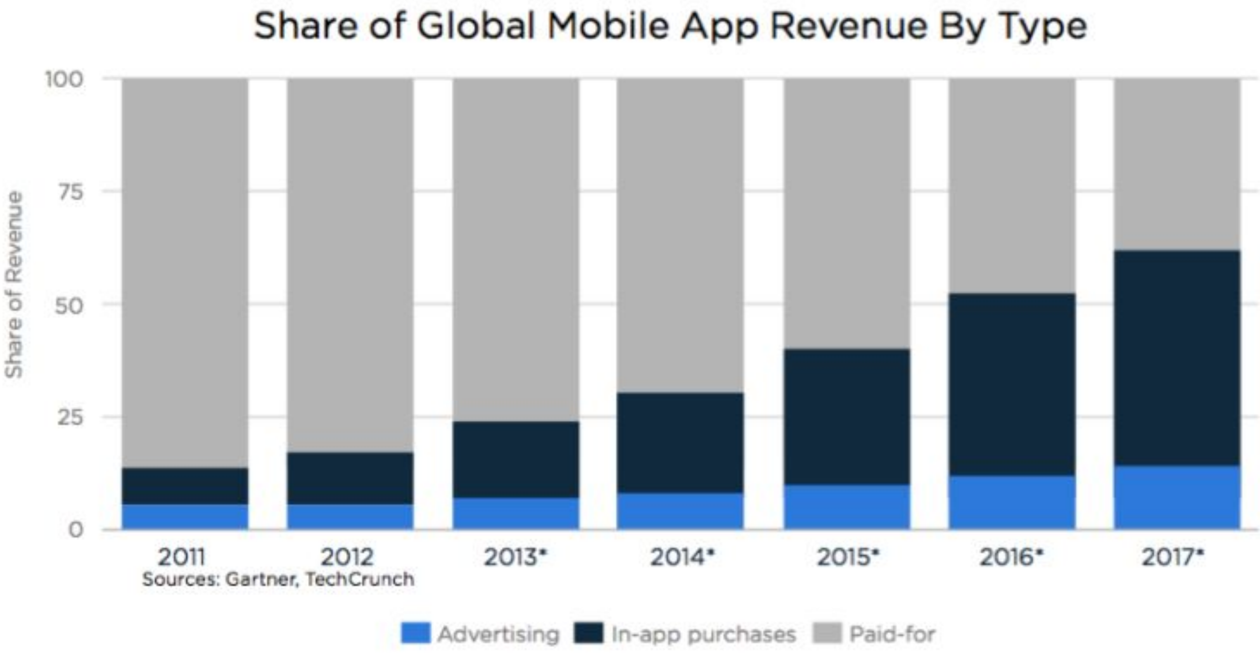
AdMob by Google
Crashlytics
amazon
MOBILE ANALYTICS

Advertising and Tracking Services (ATS)

Advertising and Tracking Services - capable (ATS-c)

Monetization with Advertising

- 94% free apps



Sources: Gartner, TechCrunch

33 billion

Violation of Least-Privileged Principle

	Permission 1	Permission 2	Permission 3
A			
A			

Opacity to user:

- Which 3rd party services

ed

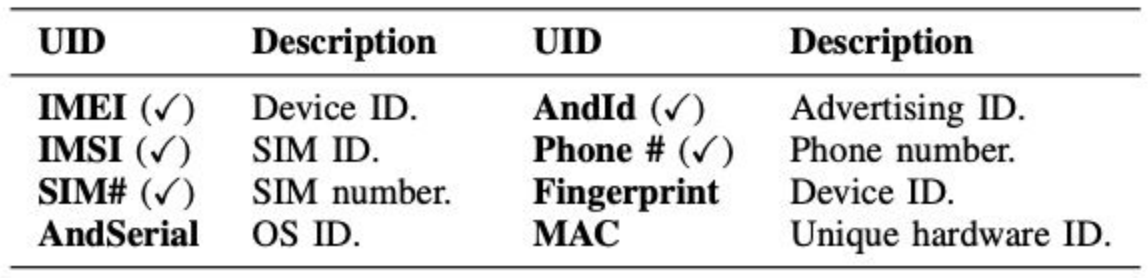
Bring Transparency to the Ecosystem!

Part 1: Data Collection through Crowdsourcing

- Leverage Android VPN permission
- Route packages to local device

- Send summarized and anonymized data

- Intercept traffic via TLS proxy with user consent



UID	Description	UID	Description
IMEI (✓)	Device ID.	**AndId** (✓)	Advertising ID.
IMSI (✓)	SIM ID.	**Phone #** (✓)	Phone number.
SIM# (✓)	SIM number.	**Fingerprint**	Device ID.
AndSerial	OS ID.	**MAC**	Unique hardware ID.

,384 users from

0+ countries

,599 apps

,533 domains

Correlate Information Flow with Contextual Info:

- Identity
- Location
- Contact list, SMS, call logs

Identify PII in payload

Ethical Consideration

- IRB approved
    - Not involving human subject, analyzing software, not users
- Informed consent on interception
- Allow to disable interception at any time
- Summarized and anonymized

Discussion

- Is there any ethical problem with their approach?

Comparison with Similar Studies

Lumen:

- Capture user data *locally* on device
- Correlate contextual information (e.g., process ID) with flows

Static Analysis:

- False positive
- Scalability
- Obfuscation

ReCon:

- Sending all device traffic to a proxy server
- Intercept at the server side

Dynamic analysis:

- Not real user engagement
- Low coverage with automated UI execution tool

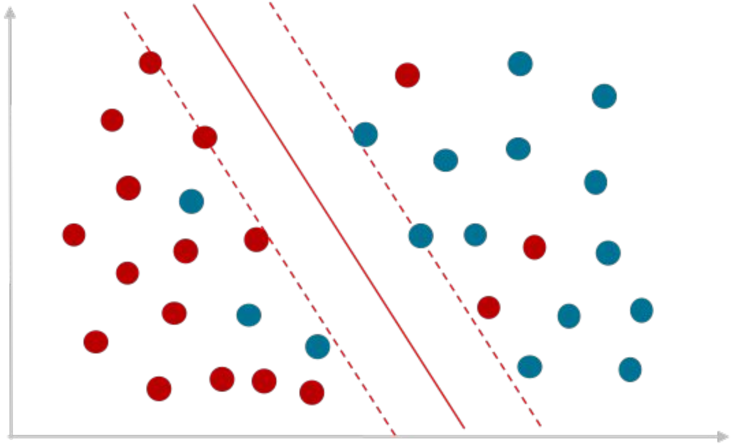
Higher precision

“Won’t Somebody Think of the Children?” Examining COPPA Compliance at Scale

ReCon: Revealing and Controlling PII Leaks in Mobile Network Systems

Discussion

- What do you think of ReCon vs. this paper? Precision?



Part 2: Classification on Third-party Domains

Classifying the Destination Domain

Baseline: leveraging publicly available services/list

- e.g., EasyList, OpenDNS domain tagger

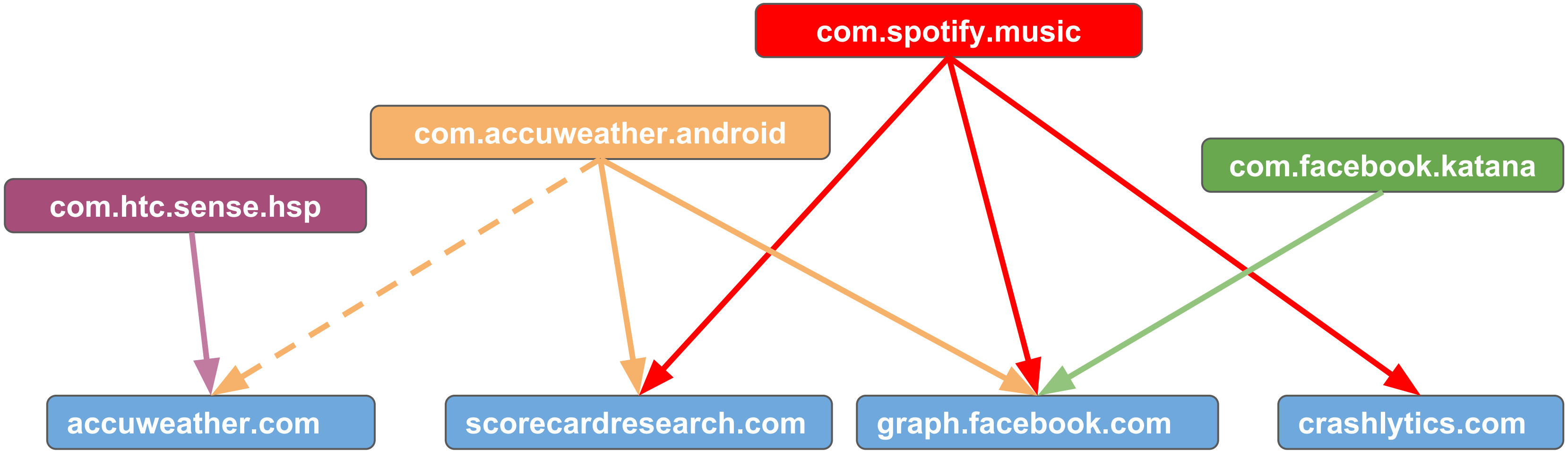
http://googleadsservices.com -> "Advertising"

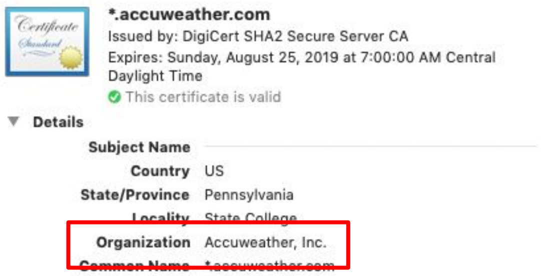
Deficiency: low coverage

Their three-step approach

- Identifying third-party domain by comparing TLS certificate
- Identifying ATS domains with machine learning
- Identifying ATS-c from the rest which UID is sent to

First Step: Identifying Third-Party Domains

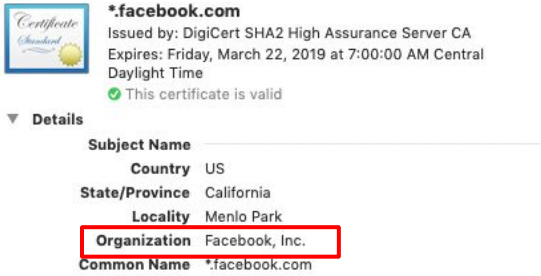




***.accuweather.com**
Issued by: DigiCert SHA2 Secure Server CA
Expires: Sunday, August 25, 2019 at 7:00:00 AM Central Daylight Time
This certificate is valid

▼ **Details**

Subject Name	
Country	US
State/Province	Pennsylvania
Locality	State College
Organization	Accuweather, Inc.
Common Name	*.accuweather.com



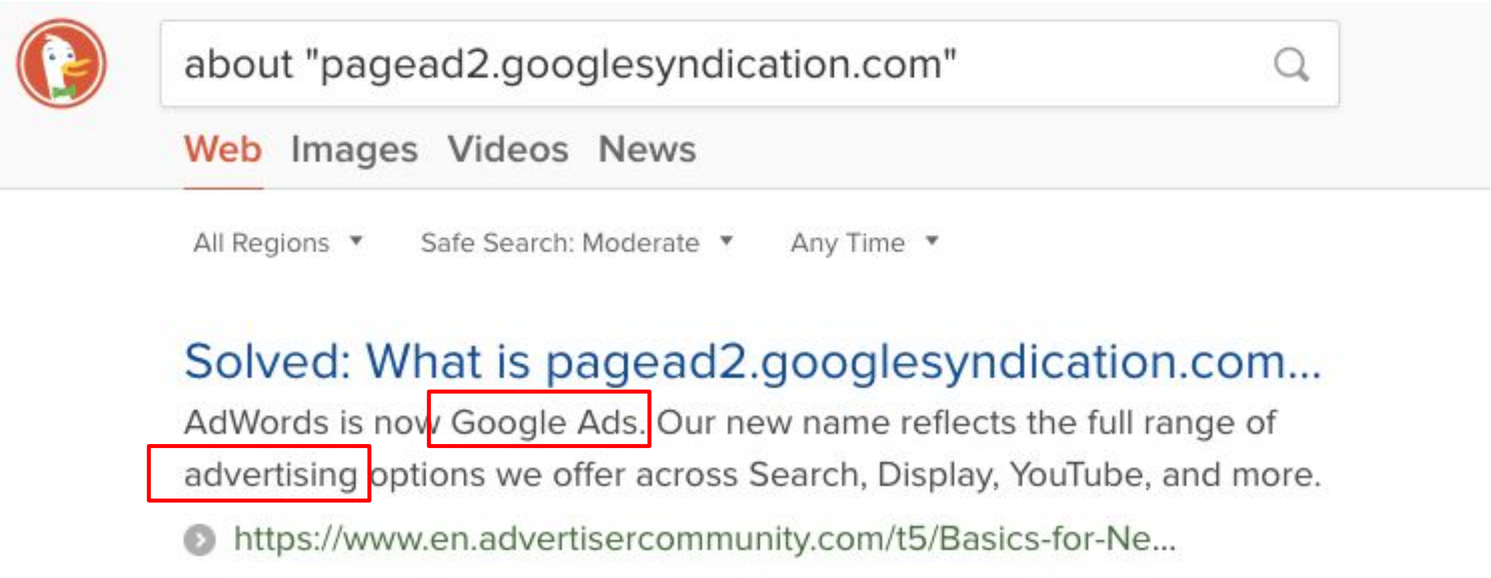
***.facebook.com**
Issued by: DigiCert SHA2 High Assurance Server CA
Expires: Friday, March 22, 2019 at 7:00:00 AM Central Daylight Time
This certificate is valid

▼ **Details**

Subject Name	
Country	US
State/Province	California
Locality	Menlo Park
Organization	Facebook, Inc.
Common Name	*.facebook.com

Second Step: Classifying ATS-domains

- Train an SVM classifier:



about "pagead2.googlesyndication.com"

Solved: What is pagead2.googlesyndication.com...

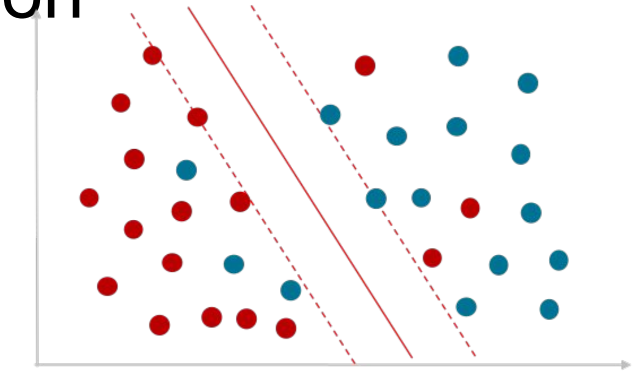
AdWords is now Google Ads. Our new name reflects the full range of advertising options we offer across Search, Display, YouTube, and more.

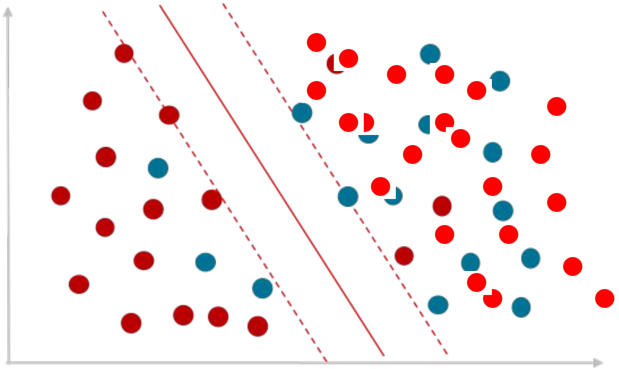
https://www.en.advertisercommunity.com/t5/Basics-for-Ne...

- Ac
- Evaluation: 200 predicted ATS, 100 predicted non-ATS
- 4% false positive, 10% false negative

Discussion

- Identifying ATS-domains:
    - Data noise -> low precision?
        - Topic ATS: “ads”, “analytics”, “services”
        - Topic non-ATS: anything

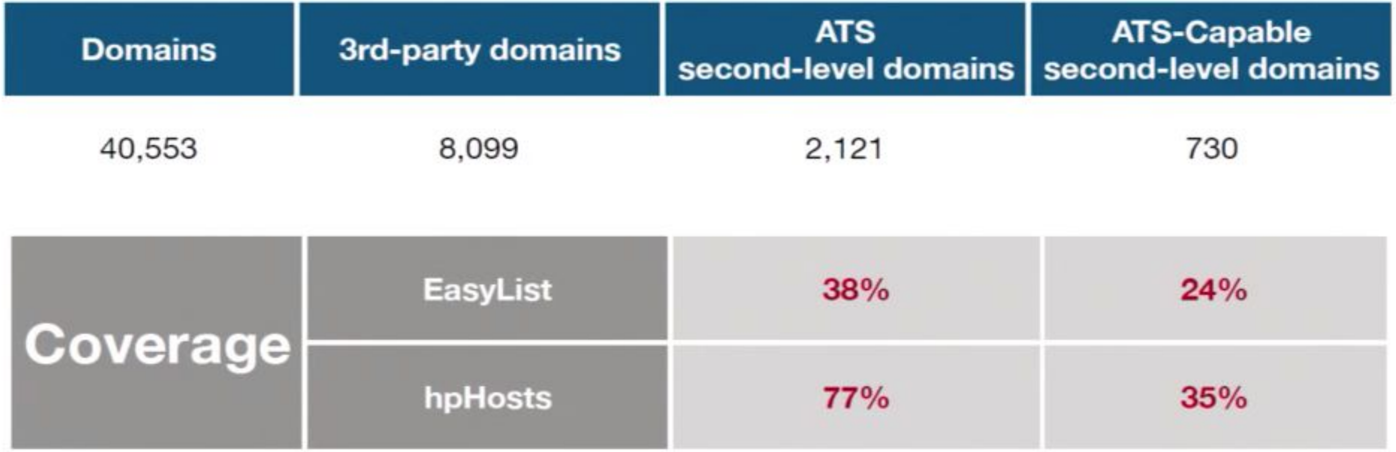




Third Step: Identifying ATS-c Domain

- Classify a domain as ATS-c if:
  - It is not ATS
  - Some user identifiers are sent to the domain

Evaluation on Coverage



Domains	3rd-party domains	ATS second-level domains	ATS-Capable second-level domains
40,553	8,099	2,121	730
Coverage	EasyList	38%	24%
	hpHosts	77%	35%

233 domains not covered by any list/service

Part 3: Basic Analysis on ATS data

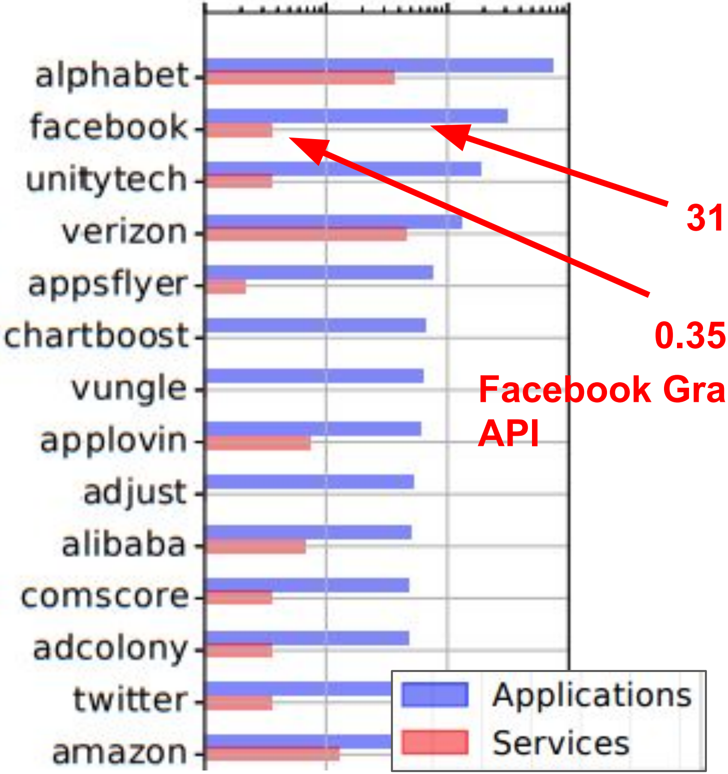
UID harvesting

- 3rd party domain = 20% of all domains
- But they are responsible for 40% of UID harvesting
- Only 14.4% of all ATSes harvest UID from the device => other tracking e.g., HTTP headers, cookies
- Most commonly harvested data is Android ID
- Android ID should not be associated with any other PII in 34% cases

Which Companies Own the Most ATSes?

- Map domains to parent company:
  - D&B Hoovers, Crunchbase

#1 Alphabet
Crashlytics
doubleclick by Google
Google AdWords
Google APIs Client Libraries
#4 verizon
FLURRY
Aol.
YAHOO!



31%

0.35%

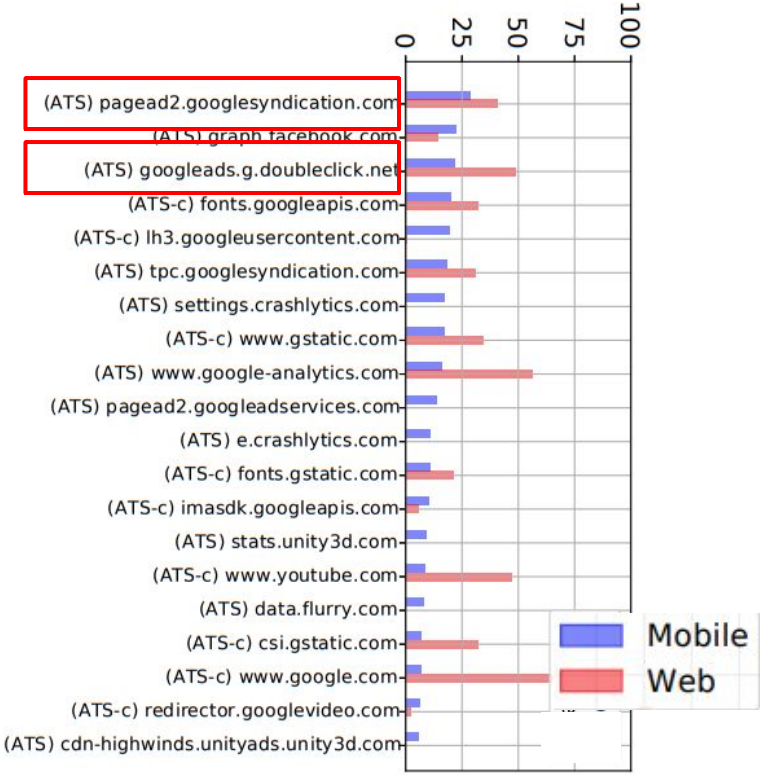
Facebook Graph API

Does Paid Apps Free You from Being Tracked?

- 82% apps connects to at least 1 ATS
- 29% apps connects to at least 5 ATSs
- Free apps: 2 ATSs, 1 ATS-c
- Paid apps: 1 ATS, 1 ATS-c
- Apps with In-app Purchase: 3 ATSs, 2 ATS-c

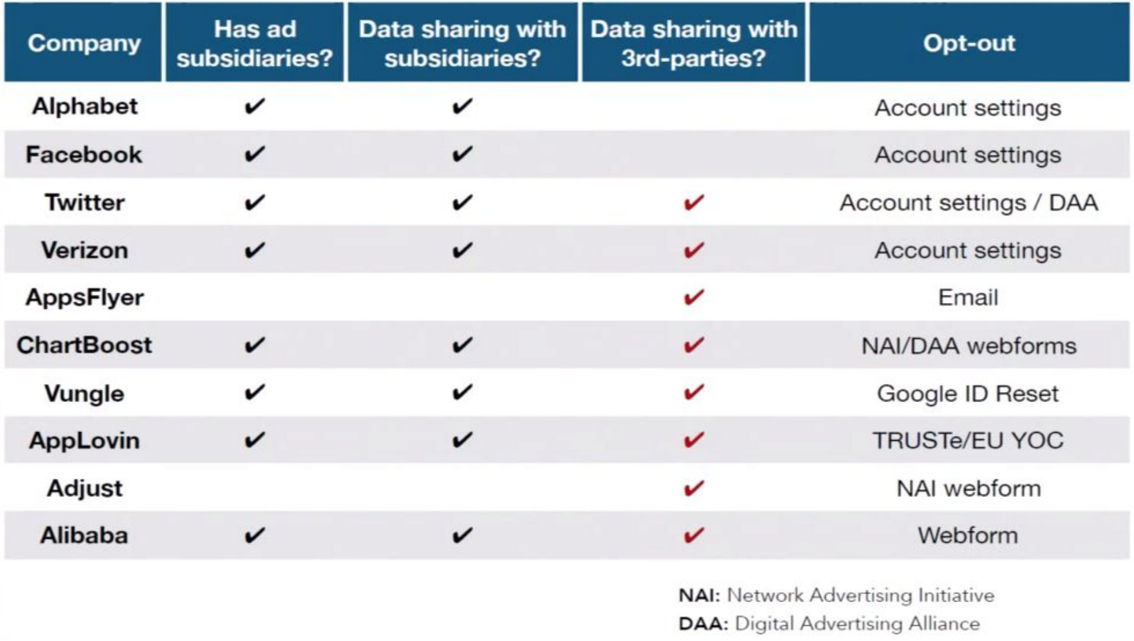
Who Tracks You on Both Mobile and Web?

- Collect website tracking statistics from Alexa Top 1,000
- Both mobile and web:
  - pagead2.googlesyndication.com
  - Googleads,g,doubleclick.net
- Web >> mobile:
  - www.youtube.com
  - www.google.com



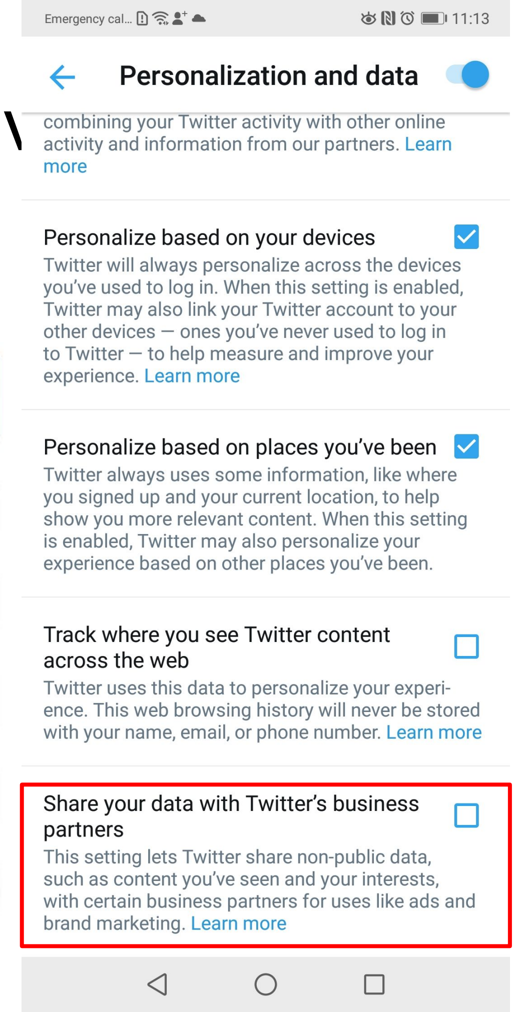
Where Did The Data Go Ev

- Privacy policy statement about data sharing



Company	Has ad subsidiaries?	Data sharing with subsidiaries?	Data sharing with 3rd-parties?	Opt-out
Alphabet	✓	✓		Account settings
Facebook	✓	✓		Account settings
Twitter	✓	✓	✓	Account settings / DAA
Verizon	✓	✓	✓	Account settings
AppsFlyer			✓	Email
ChartBoost	✓	✓	✓	NAI/DAA webforms
Vungle	✓	✓	✓	Google ID Reset
AppLovin	✓	✓	✓	TRUSTe/EU YOC
Adjust			✓	NAI webform
Alibaba	✓	✓	✓	Webform

NAI: Network Advertising Initiative
DAA: Digital Advertising Alliance



Personalization and data

combining your Twitter activity with other online activity and information from our partners. Learn more

Personalize based on your devices ☑
Twitter will always personalize across the devices you've used to log in. When this setting is enabled, Twitter may also link your Twitter account to your other devices — ones you've never used to log in to Twitter — to help measure and improve your experience. Learn more

Personalize based on places you've been ☑
Twitter always uses some information, like where you signed up and your current location, to help show you more relevant content. When this setting is enabled, Twitter may also personalize your experience based on other places you've been.

Track where you see Twitter content across the web □
Twitter uses this data to personalize your experience. This web browsing history will never be stored with your name, email, or phone number. Learn more

Share your data with Twitter's business partners □
This setting lets Twitter share non-public data, such as content you've seen and your interests, with certain business partners for uses like ads and brand marketing. Learn more

Part 4: Analysis regarding Regulation Compliance

General Data Protection Regulation

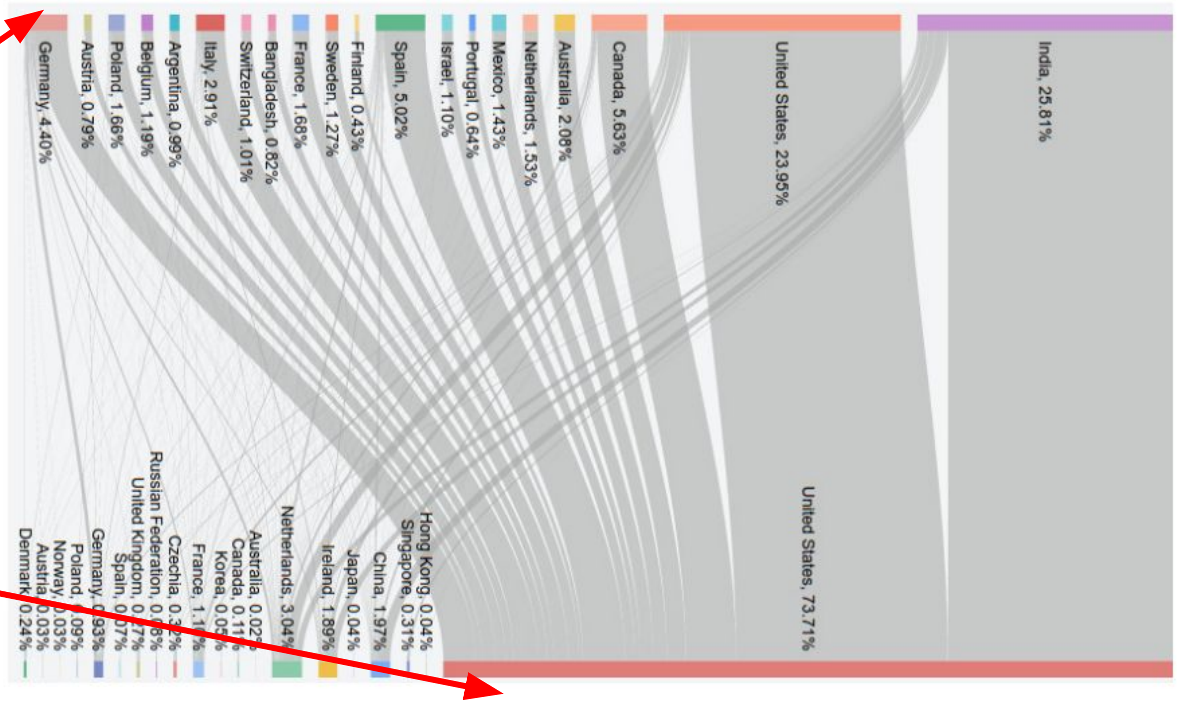
- European Union data protection law
- Protection of the data belonging to European users (EU) and European Economic Area (EEA)
- In effect since May 25, 2018
- “Data protection by design and by default” (Article 25)

GDPR Content Related to Mobile Security

- Explicit consent:
  - Must explicitly request user consent for accessing data (opt-in)
  - Explain the purpose with plain words
- Right to access/erasure:
  - Data processor must provide a copy of accessed user data
  - User can opt-out and require to erase the data at any time
- Transfer data outside Europe:
  - Strictly prohibited

A Geographical View of Data Flow

Germany Italy Spain Canada USA India



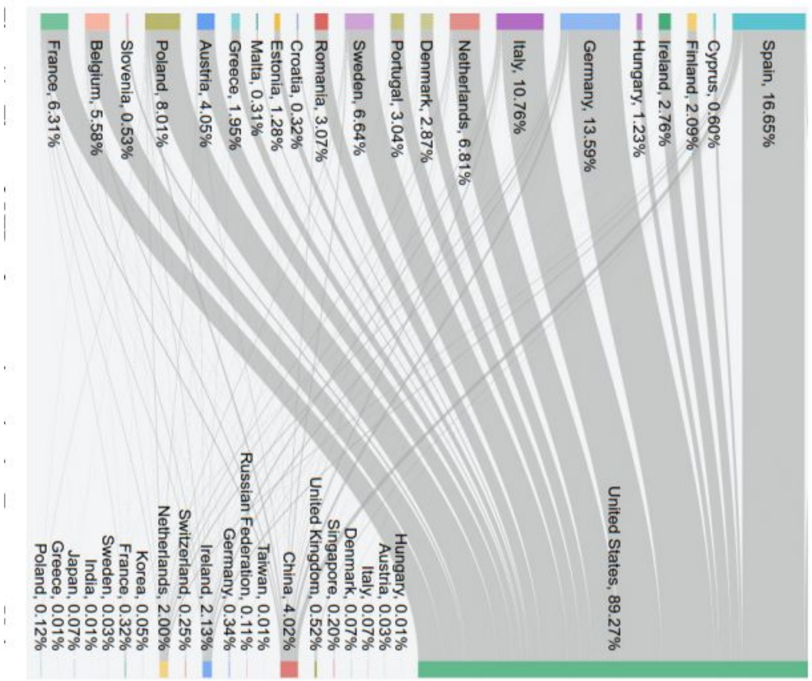
USA

Location of Lumen user

ATS-related IP address

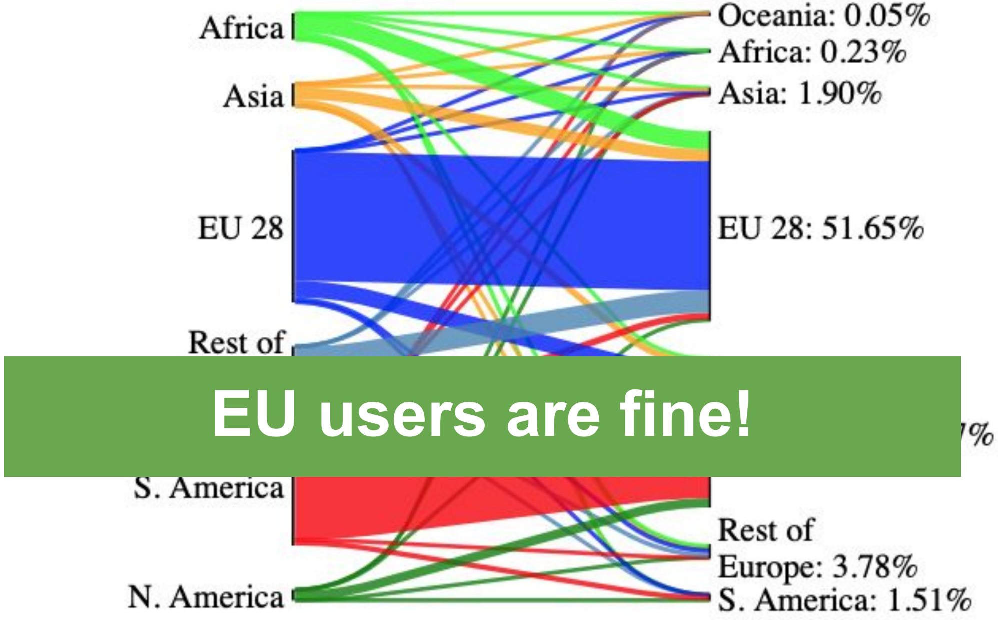
Cross-Continent Flow

Netherland Germany Spain



A Different Measurement Result

- Browser information flow
- “Inaccurate geolocation on IP“
  - Physical location of Google server -> Mountain View
  - Use Improved IP mapping
  - RIPE IPmap

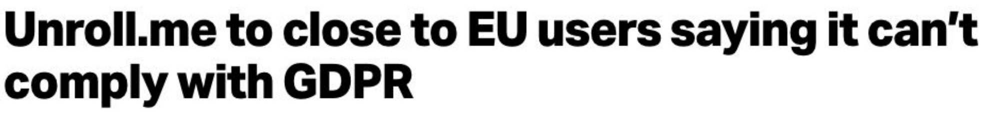


EU users are fine!

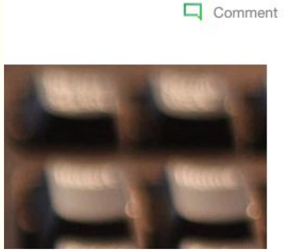
Tracing Cross Border Web Tracking, IMC 2018

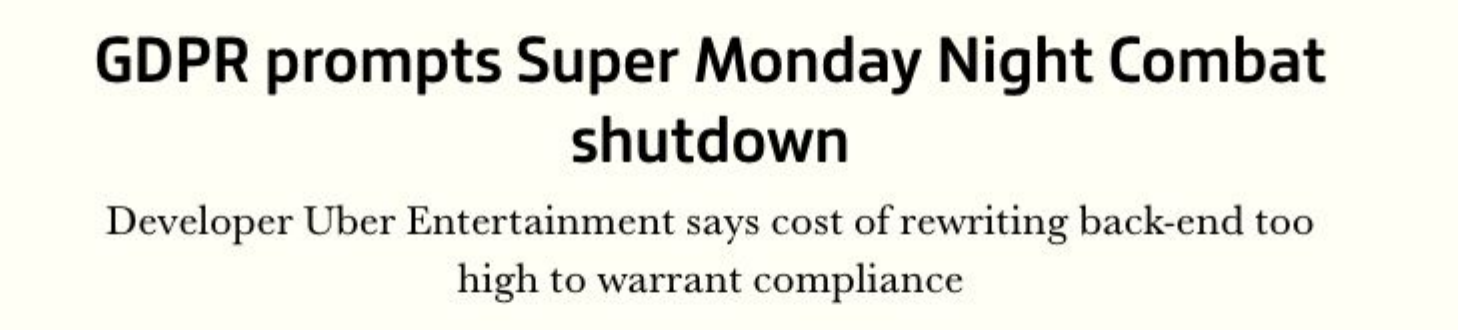
GDPR Reception

How much are busin



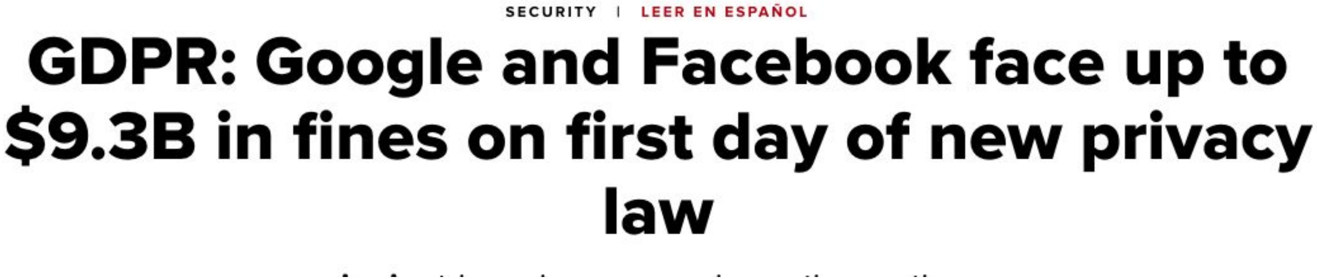
Unroll.me to close to EU users saying it can't comply with GDPR





GDPR prompts Super Monday Night Combat shutdown

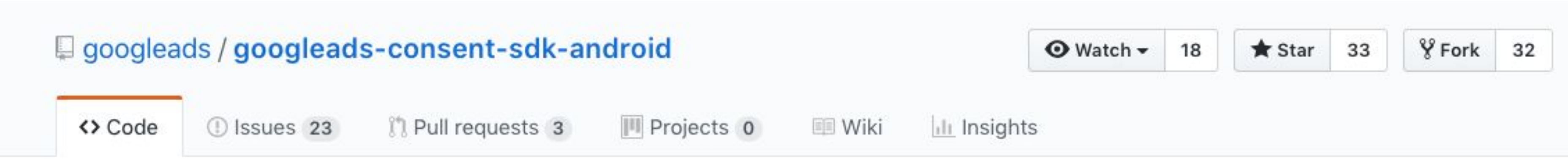
Developer Uber Entertainment says cost of rewriting back-end too high to warrant compliance



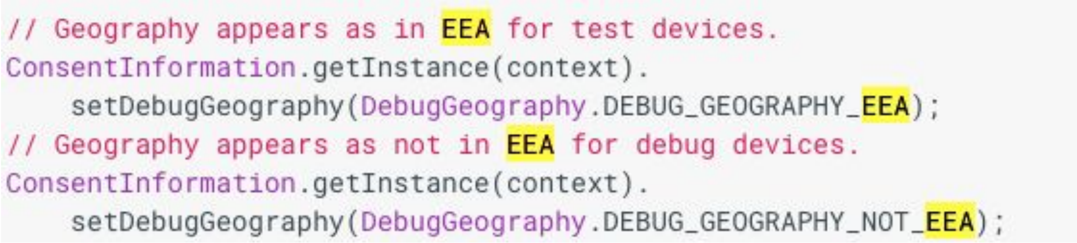
SECURITY | LEER EN ESPAÑOL

GDPR: Google and Facebook face up to $9.3B in fines on first day of new privacy law

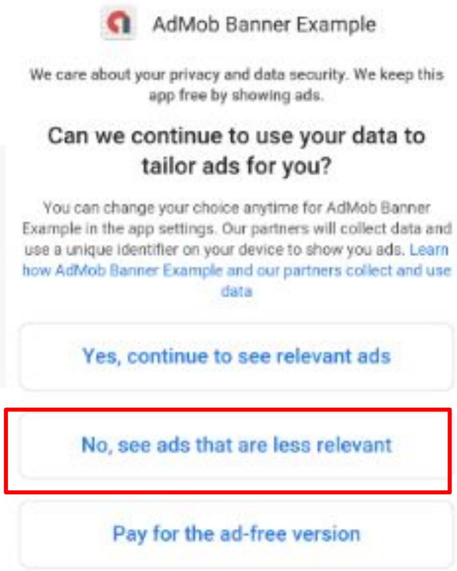
Google Ads Consent SDK



Consent SDK https://developers.google.com/admob/a...



```
// Geography appears as in EEA for test devices.
ConsentInformation.getInstance(context).
    setDebugGeography(DebugGeography.DEBUG_GEOGRAPHY_EEA);
// Geography appears as not in EEA for debug devices.
ConsentInformation.getInstance(context).
    setDebugGeography(DebugGeography.DEBUG_GEOGRAPHY_NOT_EEA);
```



AdMob Banner Example

We care about your privacy and data security. We keep this app free by showing ads.

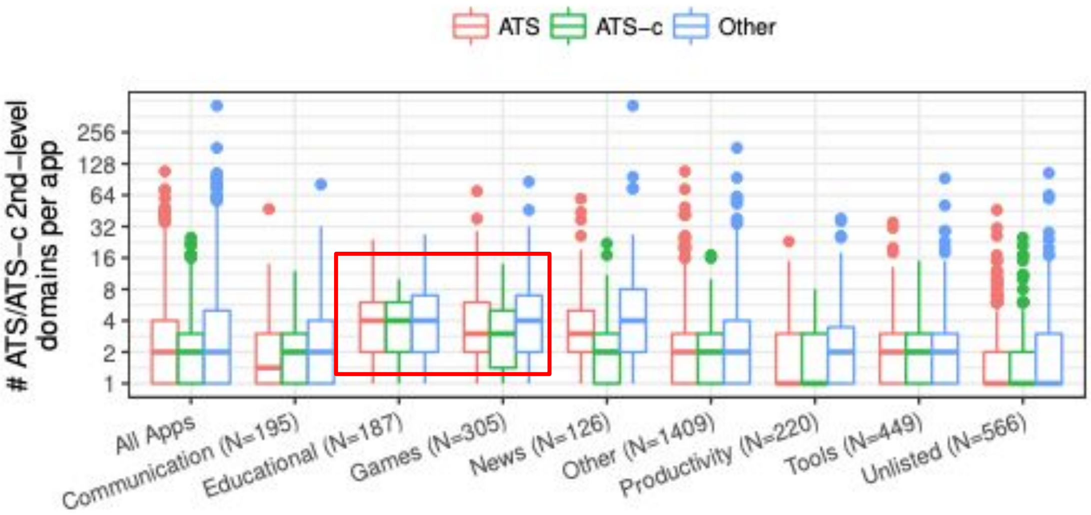
Can we continue to use your data to tailor ads for you?

You can change your choice anytime for AdMob Banner Example in the app settings. Our partners will collect data and use a unique identifier on your device to show you ads. Learn how AdMob Banner Example and our partners collect and use data

Yes, continue to see relevant ads
No, see ads that are less relevant
Pay for the ad-free version

Compliance to COPPA

- 88% Game & educational apps are under 13
- Do not use less ATS/ATS-c



Insights on Regulation Compliance

- Due to the opacity of ATS, it is difficult to uncover how organizations collect, store and share the data
- The clarity of GDPR needs further improvement
  - How consent must be obtained? Install-time permission OK?
  - How exact to withdraw the consent? Uninstall enough?
- User has no control of who has access to their data

Future Work

- What is the impact of GDPR on ATS tracking?
- Do apps behave the same after opt-out?

Takeaway

- ATS tracking are pervasive
- Big companies are the biggest data brokers
- You can get somewhat less tracking by paying for it
- Difficult to strictly enforce GDPR on ATS
- Would not judge individual compliance

Questions?