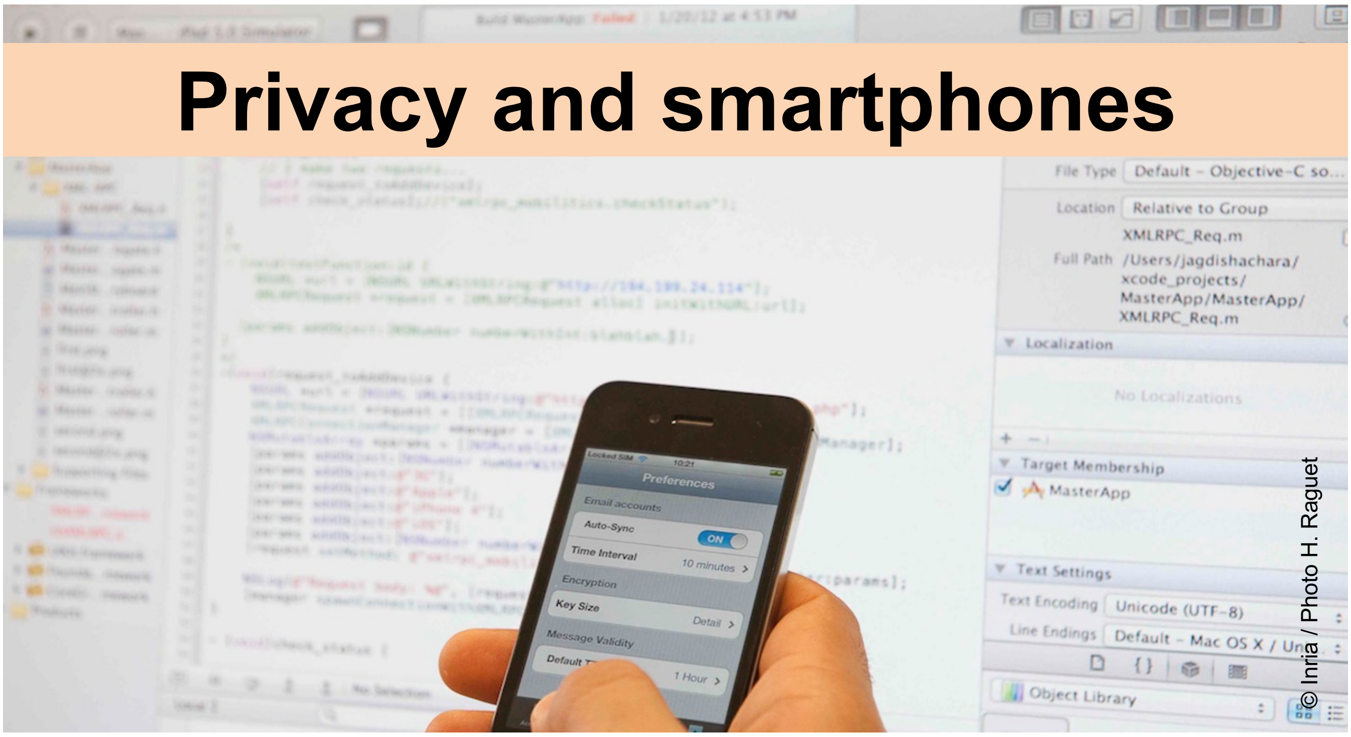
Privacy and smartphones



File Type Default - Objective-C so...

Location Relative to Group

XMLRPC_Req.m

Full Path /Users/jagdishachara/xcode_projects/MasterApp/MasterApp/XMLRPC_Req.m

Text Encoding Unicode (UTF-8)

Line Endings Default - Mac OS X / Uni

© Inria / Photo H. Raguet

Vincent Roca, Inria PRIVATICS, vincent.roca@inria.fr

Marster 2 Cybersecurity – Grenoble, December 5th, 2016

Inria Grenoble Rhône-Alpes Privatics team

- understanding and formalizing privacy
- building privacy preserving systems

Outline

1. introduction
   - **two examples**
   - **“personal information” and the French/EU law**
2. smartphones and personal information eco-system
   - **why are we here?**
   - **let’s come back to smartphones**
   - **who does what, who earns what?**
   - **free in exchange of targeted advertising: where's the problem?**
3. the Mobilitics project
4. a few ideas and results from Mobilitics
   - **the OS manufacturer approaches to control PI**
   - **the case of the "ACCESS_WIFI_STATE" Android permission**
   - **applications: a rush towards stable identifiers**
   - **the RATP application, 2013 version**
   - **tracking in the physical world with the smartphone Wifi interface**
5. conclusions

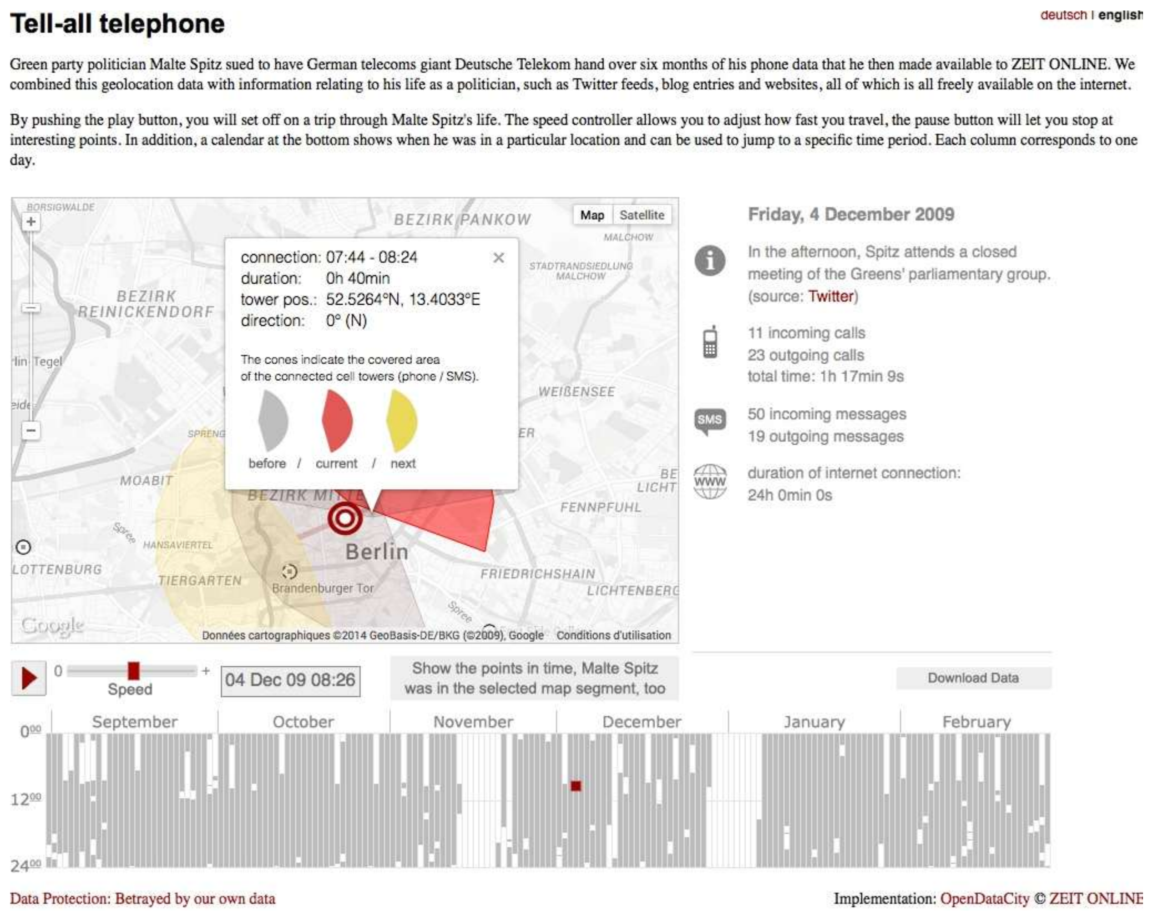
Introduction

- Two examples to start with…

Example 1 : geolocation data of a telecom operator (2009)

- Malte Spitz (German Green Party) asked his telecom operator to access his data
  - ❍enriched with publicly available data (e.g., twitter)
  - ❍a dedicated application has been designed to navigate in the history
    - http://www.zeit.de/datenschutz/malte-spitz-data-retention/

Example 1… (cont')



deutsch I english

Tell-all telephone

Green party politician Malte Spitz sued to have German telecoms giant Deutsche Telekom hand over six months of his phone data that he then made available to ZEIT ONLINE. We combined this geolocation data with information relating to his life as a politician, such as Twitter feeds, blog entries and websites, all of which is all freely available on the internet.

By pushing the play button, you will set off on a trip through Malte Spitz's life. The speed controller allows you to adjust how fast you travel, the pause button will let you stop at interesting points. In addition, a calendar at the bottom shows when he was in a particular location and can be used to jump to a specific time period. Each column corresponds to one day.

connection: 07:44 - 08:24
duration: 0h 40min
tower pos.: 52.5264°N, 13.4033°E
direction: 0° (N)

The cones indicate the covered area of the connected cell towers (phone / SMS).

before / current / next

Friday, 4 December 2009

In the afternoon, Spitz attends a closed meeting of the Greens' parliamentary group. (source: Twitter)

11 incoming calls
23 outgoing calls
total time: 1h 17min 9s

50 incoming messages
19 outgoing messages

duration of internet connection:
24h 0min 0s

Speed

04 Dec 09 08:26

Show the points in time, Malte Spitz was in the selected map segment, too

Download Data

Data Protection: Betrayed by our own data

Implementation: OpenDataCity © ZEIT ONLINE

Example 1… (cont')

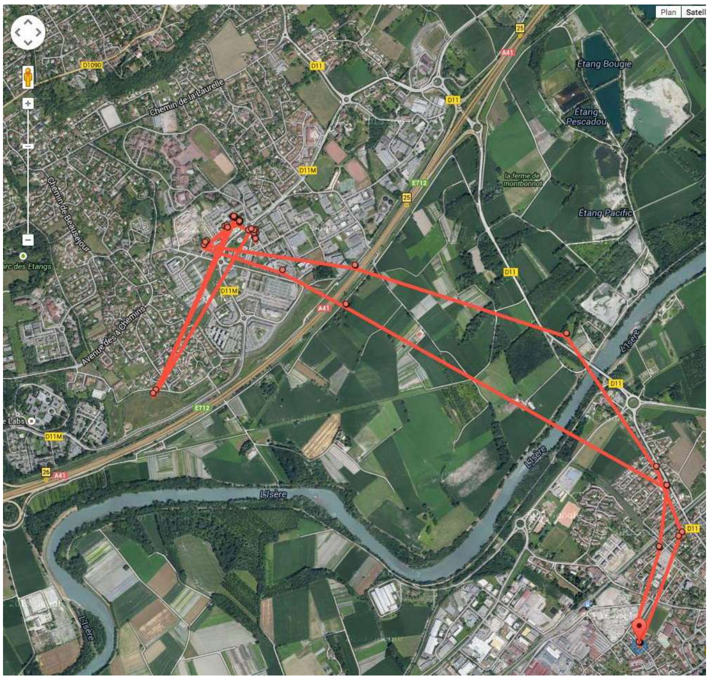
- okay, but a legal framework exists that protects the citizens ☺
  - the telecom operator has legal obligations
  - data exists but is only available under specific conditions, after an official request of the authorities

Example 2 : geolocation made in Google

- geolocation collected by my Android smartphone for Google services
  - ❍available
    - NB: login with the gmail account used for the smartphone https://maps.google.com/locationhistory/
  - ❍it's worth having a look at it!

NB: Google recently changed this page to hide details!
Only a summary is provided. Far less frightening

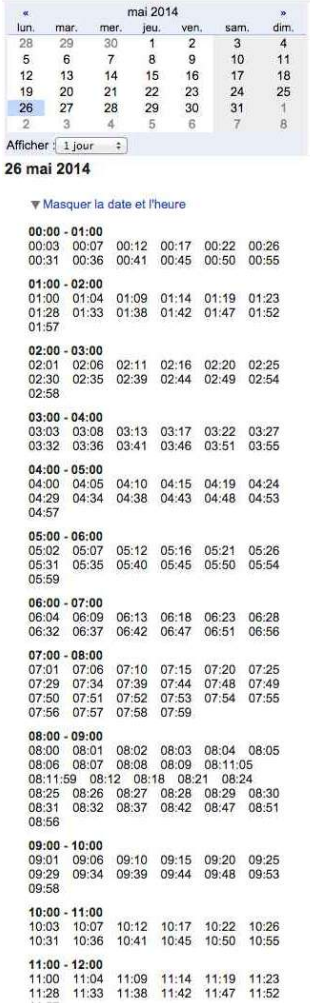
Is it reasonable?



- Google knows where I work, where I live, what I'm doing during the day, how I move…
  - ❍**you too now ;-)**

Is it reasonable… (cont.’)

- … with an incredible accuracy
  - here is the full list of geolocation points in Google database
    - a record every 5min during the whole night
    - … and every minute during the day if I’m moving!



«			mai 2014			»
lun.	mar.	mer.	jeu.	ven.	sam.	dim.
28	29	30	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	1
2	3	4	5	6	7	8

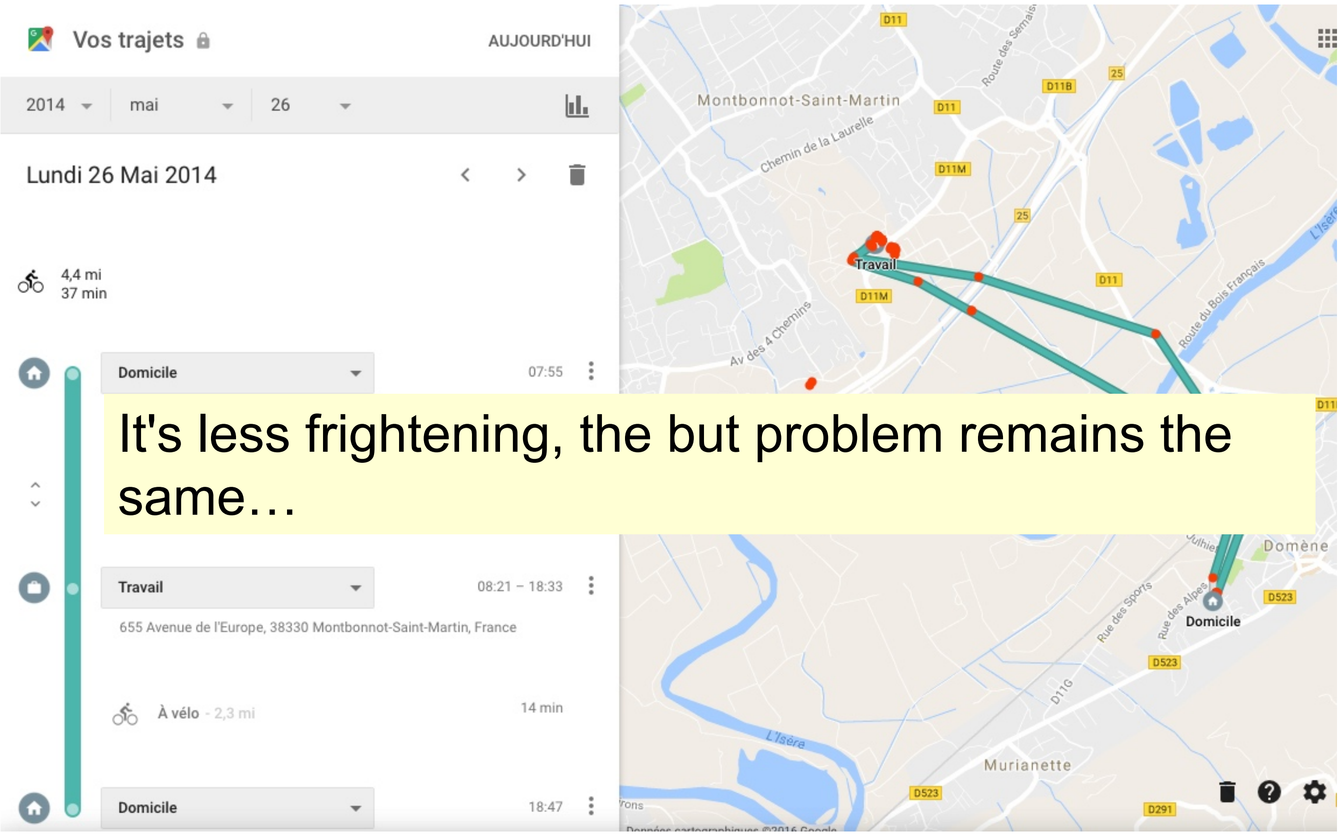
Afficher : 1 jour

26 mai 2014

▼ Masquer la date et l'heure

```
00:00 - 01:00
00:03   00:07   00:12   00:17   00:22   00:26
00:31   00:36   00:41   00:45   00:50   00:55

01:00 - 02:00
01:00   01:04   01:09   01:14   01:19   01:23
01:28   01:33   01:38   01:42   01:47   01:52
01:57

02:00 - 03:00
02:01   02:06   02:11   02:16   02:20   02:25
02:30   02:35   02:39   02:44   02:49   02:54
02:58

03:00 - 04:00
03:03   03:08   03:13   03:17   03:22   03:27
03:32   03:36   03:41   03:46   03:51   03:55

04:00 - 05:00
04:00   04:05   04:10   04:15   04:19   04:24
04:29   04:34   04:38   04:43   04:48   04:53
04:57

05:00 - 06:00
05:02   05:07   05:12   05:16   05:21   05:26
05:31   05:35   05:40   05:45   05:50   05:54
05:59

06:00 - 07:00
06:04   06:09   06:13   06:18   06:23   06:28
06:32   06:37   06:42   06:47   06:51   06:56

07:00 - 08:00
07:01   07:06   07:10   07:15   07:20   07:25
07:29   07:34   07:39   07:44   07:48   07:49
07:50   07:51   07:52   07:53   07:54   07:55
07:56   07:57   07:58   07:59

08:00 - 09:00
08:00   08:01   08:02   08:03   08:04   08:05
08:06   08:07   08:08   08:09   08:11:05
08:11:59   08:12   08:18   08:21   08:24
08:25   08:26   08:27   08:28   08:29   08:30
08:31   08:32   08:37   08:42   08:47   08:51
08:56

09:00 - 10:00
09:01   09:06   09:10   09:15   09:20   09:25
09:29   09:34   09:39   09:44   09:48   09:53
09:58

10:00 - 11:00
10:03   10:07   10:12   10:17   10:22   10:26
10:31   10:36   10:41   10:45   10:50   10:55

11:00 - 12:00
11:00   11:04   11:09   11:14   11:19   11:23
11:28   11:33   11:38   11:42   11:47   11:52
```

BTW, Google simplified the page design!



Vos trajets

AUJOURD'HUI

2014 ▾ mai ▾ 26 ▾

Lundi 26 Mai 2014

4,4 mi
37 min

Domicile ▾ 07:55

Travail ▾ 08:21 – 18:33

655 Avenue de l'Europe, 38330 Montbonnot-Saint-Martin, France

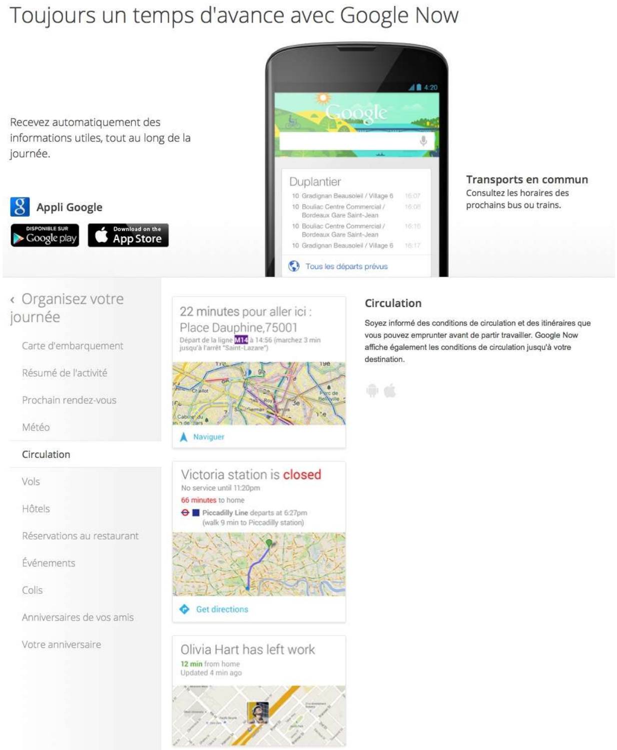
À vélo - 2,3 mi 14 min

Domicile ▾ 18:47

It's less frightening, the but problem remains the same…

Is it reasonable… (cont.')

- why is it so?
  - I've enabled Google Now : http://www.google.com/landing/now/



Toujours un temps d'avance avec Google Now

Recevez automatiquement des informations utiles, tout au long de la journée.

Appli Google

Duplantier

10	Gradignan Beausoleil / Village 6	16:07
10	Bouliac Centre Commercial / Bordeaux Gare Saint-Jean	16:08
10	Bouliac Centre Commercial / Bordeaux Gare Saint-Jean	16:16
10	Gradignan Beausoleil / Village 6	16:17

Tous les départs prévus

Transports en commun
Consultez les horaires des prochains bus ou trains.

‹ Organisez votre journée

Carte d'embarquement
Résumé de l'activité
Prochain rendez-vous
Météo
Circulation
Vols
Hôtels
Réservations au restaurant
Événements
Colis
Anniversaires de vos amis
Votre anniversaire

22 minutes pour aller ici : Place Dauphine,75001
Départ de la ligne M14 à 14:56 (marchez 3 min jusqu'à l'arrêt "Saint-Lazare")

Naviguer

Victoria station is closed
No service until 11:20pm
66 minutes to home
Piccadilly Line departs at 6:27pm (walk 9 min to Piccadilly station)

Get directions

Olivia Hart has left work
12 min from home
Updated 4 min ago

Circulation

Soyez informé des conditions de circulation et des itinéraires que vous pouvez emprunter avant de partir travailler. Google Now affiche également les conditions de circulation jusqu'à votre destination.

Is it reasonable… (cont.')

- of course…
  - Google Now can be disabled (OFF by default)
  - I can reset geolocation data on Google web site

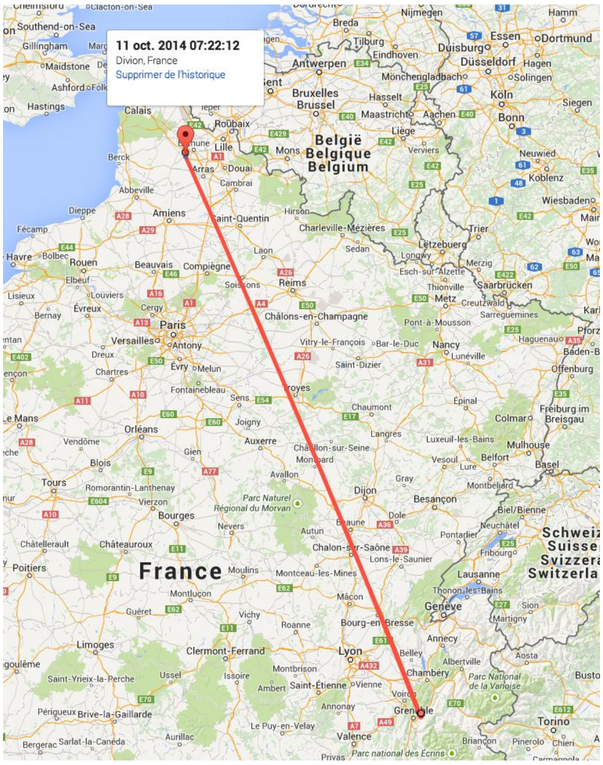
- but…
  - isn't it **disproportionate** with respect to the service provided?
    - **there's a general principle: "collect the minimum needed to provide a given service"**
    - **does the service require to keep all the records in the database for long periods?**

  - there are also geolocation **errors**…

Is it reasonable… (cont.')

- at Grenoble at 7:20, in the north 2 minutes later

❍here the mistake is obvious but sometimes it's credible!



11 oct. 2014 07:22:12
Divion, France
Supprimer de l'historique

Introduction

- **"Personal Information" (PI) and the French/EU law**

Loi informatique et liberté (1978)

identity is not required as long as a path to an identity can be found

"Constitue une donnée à caractère personnel **toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement**, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de **considérer l'ensemble des moyens** en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement **ou toute autre personne**."

no limit on the technical means

no limit: anybody in the world

Loi informatique et liberté (1978)… (cont.')

- the **nature** of the information does not matter…
  - ❍**can be anything (e.g., temperature in a home)**
- …if there is a link to a person, it's a Personal Info (PI)
- this link can be **direct**…
  - ❍**e.g., we record temperature + name**
- or **indirect**
  - ❍**e.g., we record temperature + EDF client ID**
- a person is considered identifiable if the **data controller** has the information to identify him
  - ❍**e.g., EDF collects your home temperature + EDF client ID**
- or **anybody else** in the world
  - ❍**e.g., EDF collects your home temperature + IP address of the sensor. Here the ISP can link the IP to the ADSL user**

Loi informatique et liberté (1978)… (cont.')

- French and EU definition of PI is very broad
  - ❍**in US the linkability to a person is restricted only to the data controller (i.e., database owner)**
  - ❍**MAJOR DIFFERENCE!**

- NB: a common term, PII (Personally Identifiable Information)

Loi informatique et liberté (1978)… (cont.')

- Question 1: what about the following claim?

  "we don't collect your name, age or address, only non personal information"

  - ❍**wrong if linkability to a person remains possible**

- Question 2: is an IP address a PI?
  - ❍**yes in France and in EU**
  - ❍**no in the US, apart from the ISP**

Loi informatique et liberté (1978)… (cont.')

- **sensitive information** CANNOT be collected/processed

> « Il est **interdit** de collecter ou de traiter des données à caractère personnel qui font apparaître, **directement ou indirectement**, les **origines raciales ou ethniques**, les **opinions politiques, philosophiques ou religieuses** ou **l'appartenance syndicale** des personnes, ou qui sont relatives à la **santé** ou à la **vie sexuelle** de celles-ci. »

- it's clear, non ambiguous: it's prohibited

- in practice it's pretty complex because of inference
  - **if Google knows I'm at a church every Sunday morning (thanks to geolocation) he knows something whose collection is prohibited**

Loi informatique et liberté (1978)… (cont.')

- many obligations to the data controller

« 1° Les données sont collectées et traitées de manière loyale et licite ;

fair collection

2° Elles sont collectées pour des finalités déterminées, explicites et légitimes et ne sont pas traitées ultérieurement de manière incompatible avec ces finalités. […];

well defined goal

3° Elles sont adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont collectées et de leurs traitements ultérieurs ;

collect the bare minimum

4° Elles sont exactes, complètes et, si nécessaire, mises à jour ;[…]

accuracy

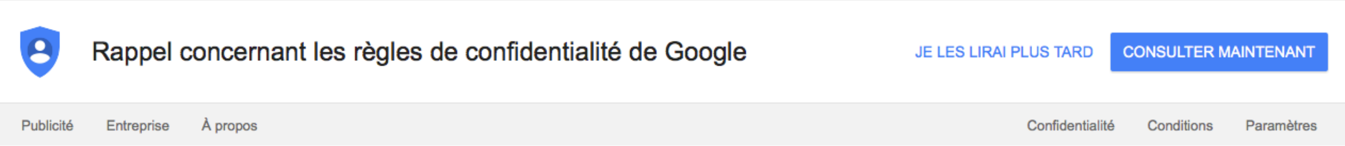
5° Elles sont conservées sous une forme permettant l'identification des personnes concernées pendant une durée qui n'excède pas la durée nécessaire aux finalités […]. »

limited duration

Ways to escape the PI rules

the data collector can do a lot if…

- solution 1: they get the **free and informed consent of the user**
  - **“consentement libre et éclairé”**
  - **explains why Google urges the user to read their confidentiality rules**



Rappel concernant les règles de confidentialité de Google

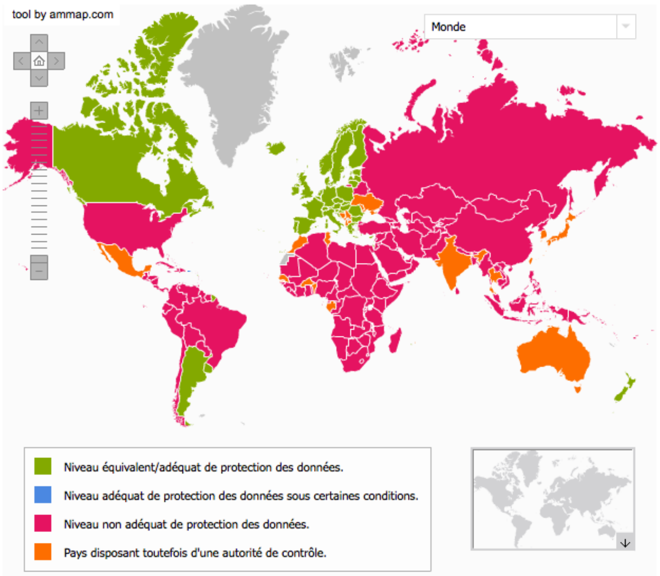
  - **is it sufficient?**
    - no if the user is not free to use the service (no alternative)
    - no if the privacy rules are not compliant with French / EU law (ex. Facebook)

Ways to escape the PI rules… (cont.’)

- solution 2: data is **anonymized**
  - **if linkability to a person is impossible it is no longer PI**
  - **but secure anonymization can be pretty hard to achieve**
    - because of inference attacks with side information
  - **and not necessarily sufficient**
    - if a group of people is known to have a certain property, and if I’m known to belong to this group, even if my individual record cannot be identified in the database, one knows I have this property too

PI transmission beyond EU

- personal info cannot be sent beyond EU borders
    - there are exceptions for countries whose data protection law is compliant with that of EU


tool by ammap.com
Monde
🟩 Niveau équivalent/adéquat de protection des données.
🟦 Niveau adéquat de protection des données sous certaines conditions.
🟥 Niveau non adéquat de protection des données.
🟧 Pays disposant toutefois d'une autorité de contrôle.

-
    - there are exception for companies who signed a specific contract

PI transmission beyond EU… (cont.')

- close-up on **US companies**
    - US is not recognized as trustworthy W.R.T. PI protection
    - the "Safe Harbor" program was used to authorize PI collection till Oct. 2015
        - **EUJC judgment (Max Schrems) concluded the US law does not guaranty the security of EU citizens PI**
    - no rule today and PI collection is therefore prohibited…
        - **… but negotiations are on the way to establish new legal foundations**
    - in the meantime high pressure of US companies to get the "free and informed" user consent

Outline

1. introduction
   - ❍ two examples
   - ❍ “personal information” and the French/EU law
2. smartphones and personal information eco-system
   - ❍ why are we here?
   - ❍ let’s come back to smartphones
   - ❍ who does what, who earns what?
   - ❍ free in exchange of targeted advertising: where's the problem?
3. the Mobilitics project
4. a few ideas and results from Mobilitics
   - ❍ the OS manufacturer approaches to control PI
   - ❍ the case of the "ACCESS_WIFI_STATE" Android permission
   - ❍ applications: a rush towards stable identifiers
   - ❍ the RATP application, 2013 version
   - ❍ tracking in the physical world with the smartphone Wifi interface
5. conclusions

The smartphones and personal information eco-system

● Why are we here?

A massive worldwide surveillance

- we leave traces that are **systematically** recorded whenever we use Internet and our smartphone
    - **on the "visible" web**
    - **on the "invisible" web**
  - for **economic** or **security** reasons

Surveillance on the "visible" web

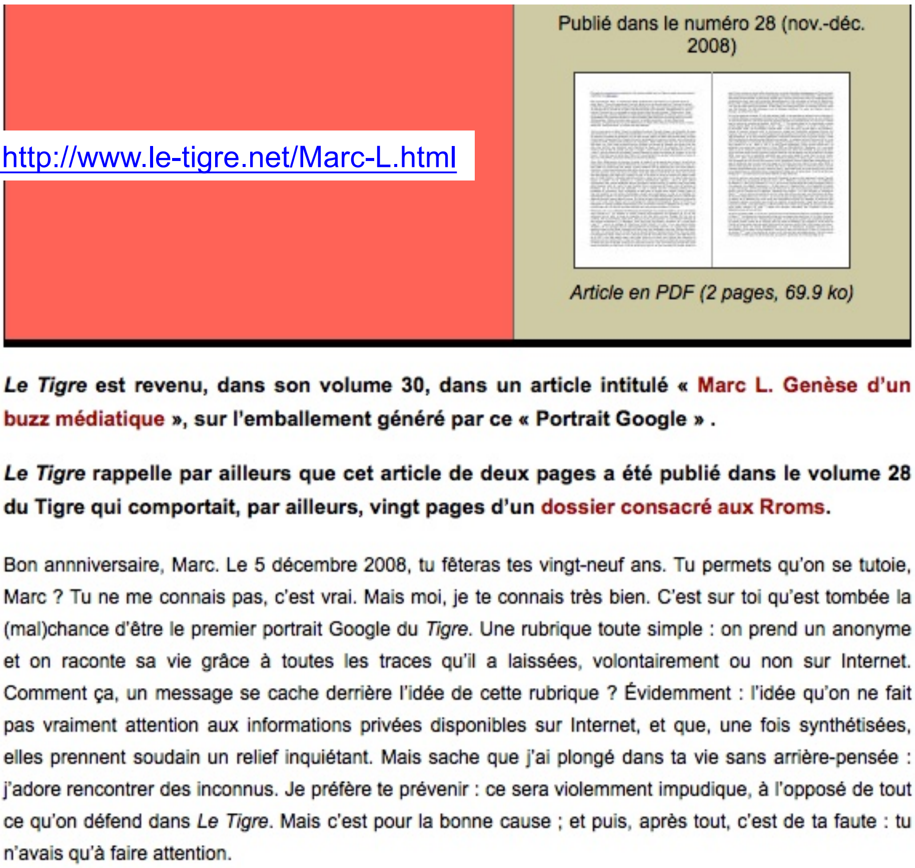
- Foursquare knows **where you are**
- Flickr knows **what you are watching**
- Facebook knows **what you're doing**
- Linkedin knows **where and with whom you're working**
- Twitter knows **what you're saying**
- Amazon knows **what you're buying**
- Google knows **what you're thinking**
- and much more…

If we cross all information, it's becoming terrifying…

*Courtesy of F. Bancihlon

Surveillance on the “visible” web… (cont’)

From http://www.le-tigre.net/Marc-L.html



Publié dans le numéro 28 (nov.-déc. 2008)

Article en PDF (2 pages, 69.9 ko)

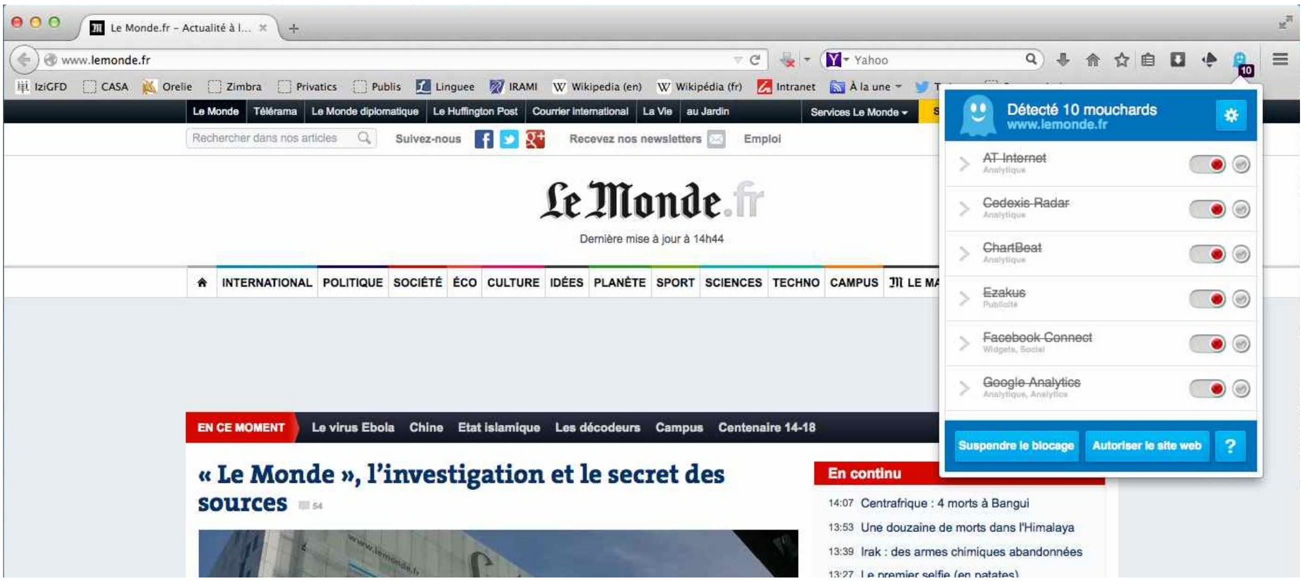
Le Tigre **est revenu, dans son volume 30, dans un article intitulé « Marc L. Genèse d’un buzz médiatique », sur l’emballement généré par ce « Portrait Google » .**

Le Tigre **rappelle par ailleurs que cet article de deux pages a été publié dans le volume 28 du Tigre qui comportait, par ailleurs, vingt pages d’un dossier consacré aux Rroms.**

Bon annniversaire, Marc. Le 5 décembre 2008, tu fêteras tes vingt-neuf ans. Tu permets qu’on se tutoie, Marc ? Tu ne me connais pas, c’est vrai. Mais moi, je te connais très bien. C’est sur toi qu’est tombée la (mal)chance d’être le premier portrait Google du *Tigre*. Une rubrique toute simple : on prend un anonyme et on raconte sa vie grâce à toutes les traces qu’il a laissées, volontairement ou non sur Internet. Comment ça, un message se cache derrière l’idée de cette rubrique ? Évidemment : l’idée qu’on ne fait pas vraiment attention aux informations privées disponibles sur Internet, et que, une fois synthétisées, elles prennent soudain un relief inquiétant. Mais sache que j’ai plongé dans ta vie sans arrière-pensée : j’adore rencontrer des inconnus. Je préfère te prévenir : ce sera violemment impudique, à l’opposé de tout ce qu’on défend dans *Le Tigre*. Mais c’est pour la bonne cause ; et puis, après tout, c’est de ta faute : tu n’avais qu’à faire attention.

Surveillance on the “invisible” web

- thanks to cookies, pixels, “I like” buttons, etc. of web sites
  - one can easily **track** and **profile** all users



Détecté 10 mouchards
www.lemonde.fr

~~AT Internet~~ Analytique
~~Cedexis Radar~~ Analytique
~~ChartBeat~~ Analytique
~~Ezakus~~ Publicité
~~Facebook Connect~~ Widgets, Social
~~Google Analytics~~ Analytique, Analytics

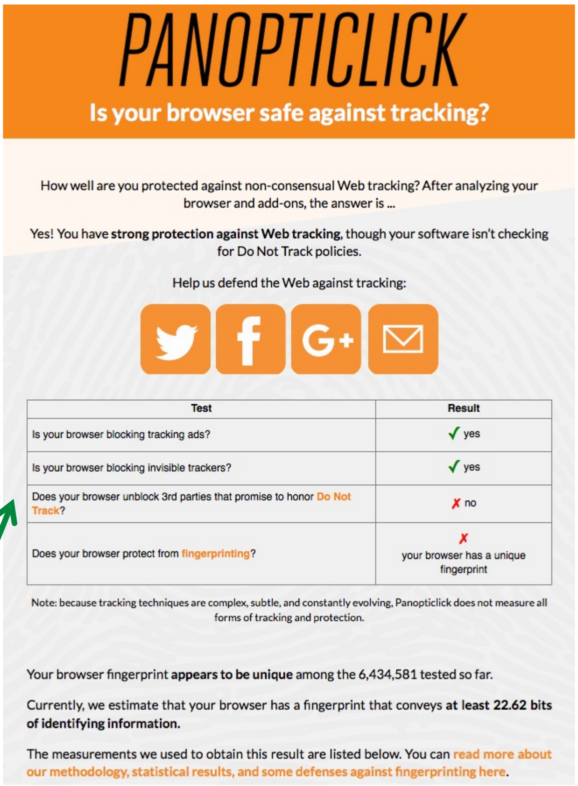
« Le Monde », l’investigation et le secret des sources

14:07 Centrafrique : 4 morts à Bangui
13:53 Une douzaine de morts dans l'Himalaya
13:39 Irak : des armes chimiques abandonnées
13:27 Le premier selfie (en patates)

Surveillance on "invisible" web… (cont')

- even if you don't provide your ID, anyway your browser is unique in the world and can be tracked
  - Panopticlick
    - fingerprinting based on config, version, OS, screen resolution, etc.
  - add blockers do help but are not 100% efficient

I'm using Adblock, Ghostery and Privacy badger!



PANOPTICLICK

Is your browser safe against tracking?

How well are you protected against non-consensual Web tracking? After analyzing your browser and add-ons, the answer is ...

Yes! You have **strong protection against Web tracking**, though your software isn't checking for Do Not Track policies.

Help us defend the Web against tracking:

Test	Result
Is your browser blocking tracking ads?	✓ yes
Is your browser blocking invisible trackers?	✓ yes
Does your browser unblock 3rd parties that promise to honor Do Not Track?	✗ no
Does your browser protect from fingerprinting?	✗ your browser has a unique fingerprint

Note: because tracking techniques are complex, subtle, and constantly evolving, Panopticlick does not measure all forms of tracking and protection.

Your browser fingerprint **appears to be unique** among the 6,434,581 tested so far.

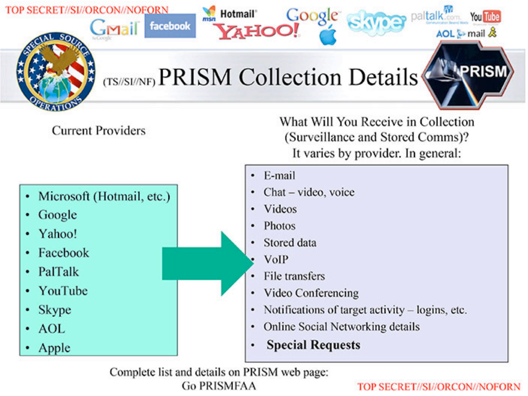
Currently, we estimate that your browser has a fingerprint that conveys **at least 22.62 bits of identifying information.**

The measurements we used to obtain this result are listed below. You can read more about our methodology, statistical results, and some defenses against fingerprinting here.

ELECTRONIC FRONTIER FOUNDATION
DEFENDING YOUR RIGHTS IN THE DIGITAL WORLD

This situation can easily lead to abuses

- NSA…
  - the core issue is not to track well identified targets, but
    - **to track citizens throughout the world**
    - **to compromise the security of our tools**



TOP SECRET//SI//ORCON//NOFORN

(TS//SI//NF) PRISM Collection Details

Current Providers

- Microsoft (Hotmail, etc.)
- Google
- Yahoo!
- Facebook
- PalTalk
- YouTube
- Skype
- AOL
- Apple

What Will You Receive in Collection (Surveillance and Stored Comms)?
It varies by provider. In general:

- E-mail
- Chat – video, voice
- Videos
- Photos
- Stored data
- VoIP
- File transfers
- Video Conferencing
- Notifications of target activity – logins, etc.
- Online Social Networking details
- **Special Requests**

Complete list and details on PRISM web page:
Go PRISMFAA

TOP SECRET//SI//ORCON//NOFORN

- and NSA is not the only agency that does it

This situation can lead to abuses… (cont’)

- … that may follow you till the rest of your life
  - linkability between pieces of information
    - metro card, debit card, cellphone data, etc.
  - taken individually, every piece of information is probably accurate, but not necessarily their link

"Metadata aggregated over a person's life tells a story about you. The story is made of facts, but that's not necessarily true. Now if a person has a perception that you've done something, it will follow you during the rest of your life."

from Jacob Appelbaum,

"Citizenfour" (offset 16'03-18'12)

What is Tor & How Does It Work?

Interview With Jacob Appelbaum

THE CARTOON BANK

“On the Internet, nobody knows you’re a dog.”

In 1993…

© NewYorker 1993

In 2015…

© NewYorker 2015

“Remember when, on the Internet, nobody knew who you were?”

The smartphones and personal information eco-system

- **let’s come back to smartphones…**

Smartphones have a key responsibility

- our everyday “companions”…
  - useful, always connected, easy to customize
- but they also

concentrate

personal information

when we use them: phone calls, SMS, web, applications, etc.

generate

personal information

GPS, NFC, WiFi, camera, fingerprint sensor, heart rate sensor, etc.

A key responsibility… (cont.')

- they know a lot on our cyber-activities
  - applications generate many **opportunities** to leak personal information
    - **it justifies that web site you visit invite you to download and use their own App…**

"notre mouchard de poche préféré ?"

What is the subject of this talk?

- a smartphone is composed of
  - an application processor
  - an operating system (OS)
    - **Android (Google), iOS (Apple), Windows Phone, FirefoxOS ✝, Tizen, Cyanogen OS, etc.**
  - applications ("Apps")

our subject (Android/ iOS)

-
  - a full system (processor + OS) for baseband communications
    - **hidden, no open spec, closed industry**

very complex to study

http://events.ccc.de/congress/2011/Fahrplan/attachments/2022_11-ccc-qcombbdbg.pdf

The smartphones and personal information eco-system

- **Who does what, who earns what?**

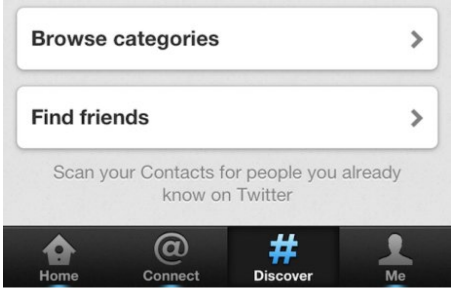
A complex eco-system

- **complex** because several actors are involved
  - ❍ **« first party » :** **owns the App**
    ⇒ those we see
  - ❍ **« third party » :** **Advertising and Analytics (A&A)**
    ⇒ those we never see
  - ❍**the third party has clients (e.g., advertising companies)**
  - ❍**certain actors play multiple roles (e.g., Google and Facebook)**

- it's impossible to trust everybody
  - ❍two examples…

Example 1: information leaks "by error"

- Twitter (Feb. 2012):
  - "La fonctionnalité de recherche d'amis de […] Twitter permet au service en ligne de télécharger sur ses serveurs les carnets d'adresses et la liste de contacts des utilisateurs. Une fois téléchargées sur ses serveurs, ces données sont conservées 18 mois."


Browse categories

Find friends

Scan your Contacts for people you already know on Twitter

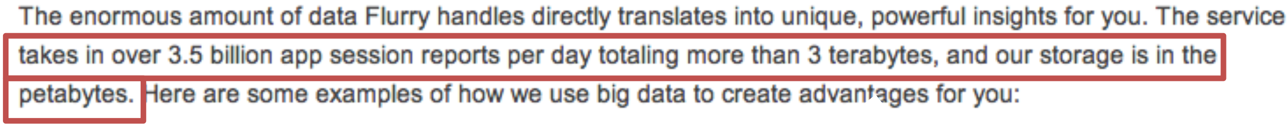
http://www.zdnet.fr/actualites/twitter-copie-et-conserve-18-mois-sans-consentement-les-carnets-d-adresses-des-utilisateurs-39768632.htm

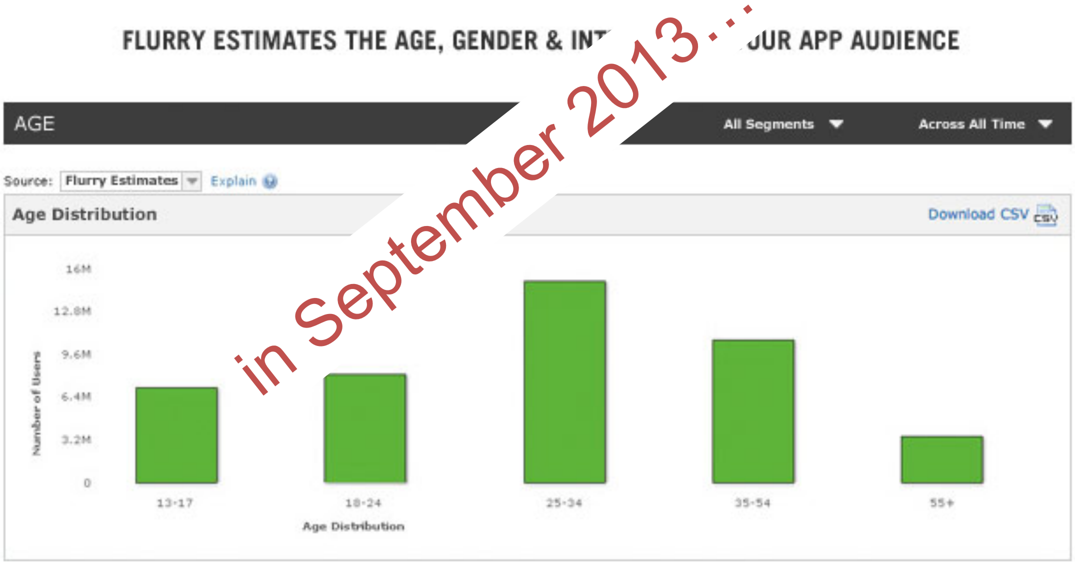
  - similar scandals happened with LinkedIn et Path en 2012!

- those are strategic errors
  - immediately fixed in a new version of the App
  - reputation is essential for those companies and risks are huge

Example 2: massive, organized collection

- Flurry (from Yahoo)
  - http://www.flurry.com


The enormous amount of data Flurry handles directly translates into unique, powerful insights for you. The service takes in over 3.5 billion app session reports per day totaling more than 3 terabytes, and our storage is in the petabytes. Here are some examples of how we use big data to create advantages for you:



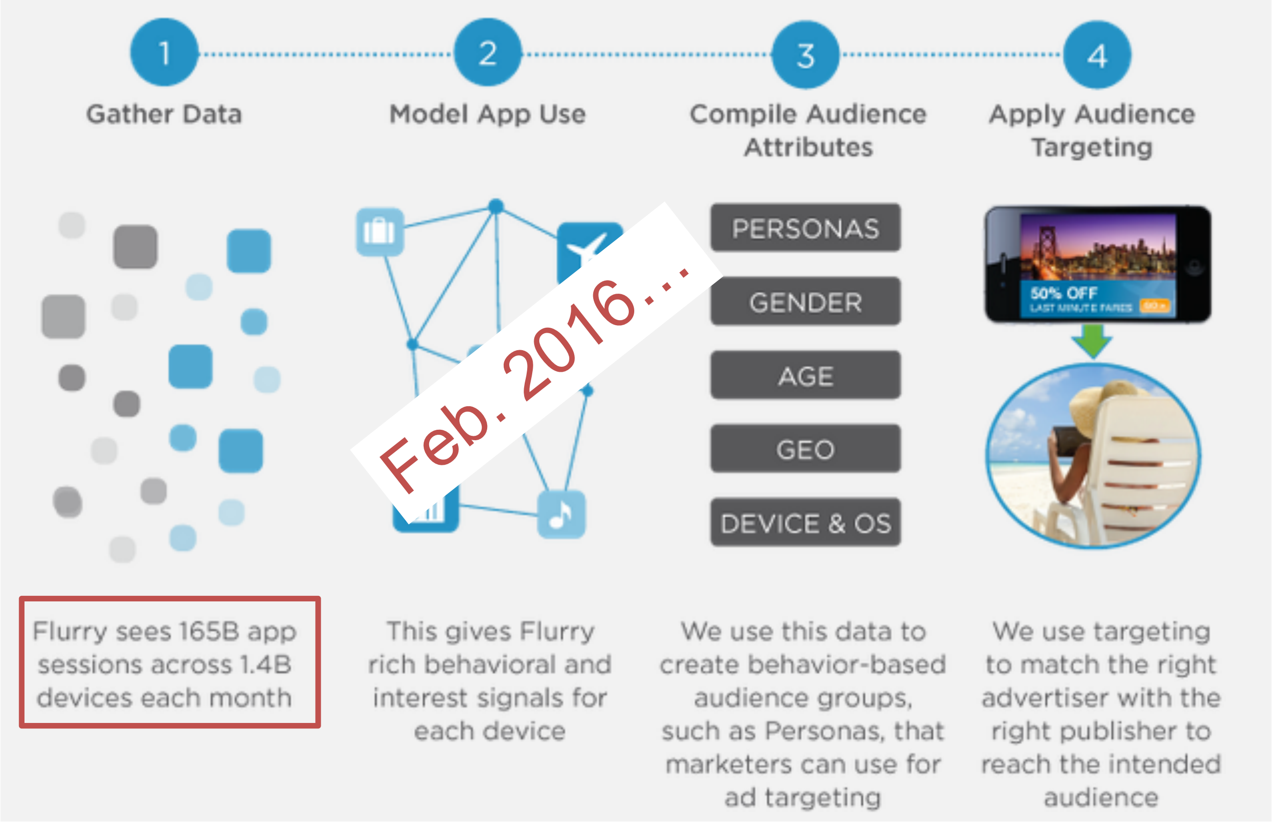
in September 2013…

Example 2: massive collection… (cont.')

- what for?
  - in order to **track users**
    - does the same user come back? What Apps does he use? With what frequency? When?
  - in order to **profile users**
    - is he a middle-age man? Does he like sport, technology? Does he read news, etc.
- final goal is to
  - sell **targeted advertising** on the smartphone
    - high click ratio because ad is targeted
  - but this database can easily be user for other purposes…
    - massive surveillance

Example 2: massive collection… (cont.')

- Ex. Flurry



1 Gather Data

2 Model App Use

3 Compile Audience Attributes

4 Apply Audience Targeting

PERSONAS
GENDER
AGE
GEO
DEVICE & OS

Feb. 2016…

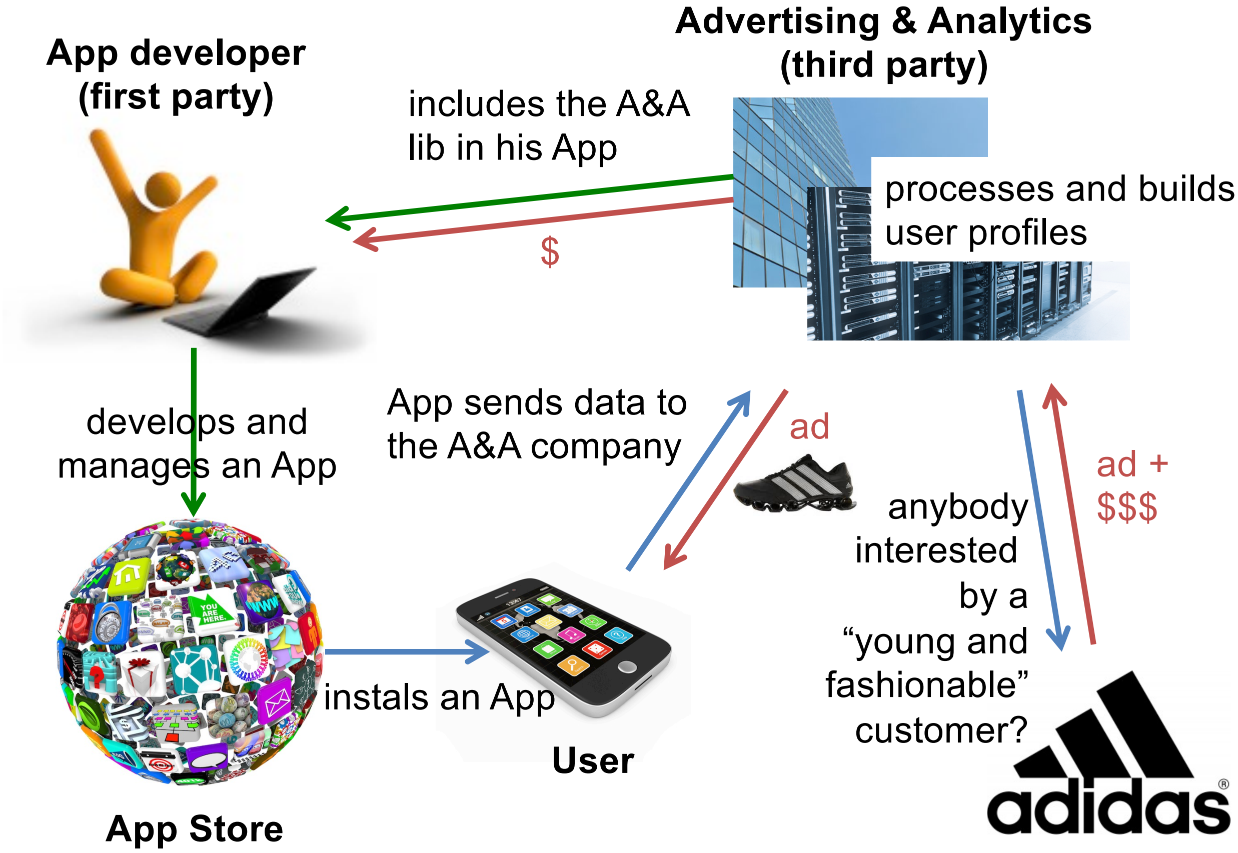
Flurry sees 165B app sessions across 1.4B devices each month

This gives Flurry rich behavioral and interest signals for each device

We use this data to create behavior-based audience groups, such as Personas, that marketers can use for ad targeting

We use targeting to match the right advertiser with the right publisher to reach the intended audience

The actors and their relationships



App developer
(first party)

Advertising & Analytics
(third party)

includes the A&A
lib in his App

$

processes and builds
user profiles

develops and
manages an App

App sends data to
the A&A company

ad

ad +
$$$

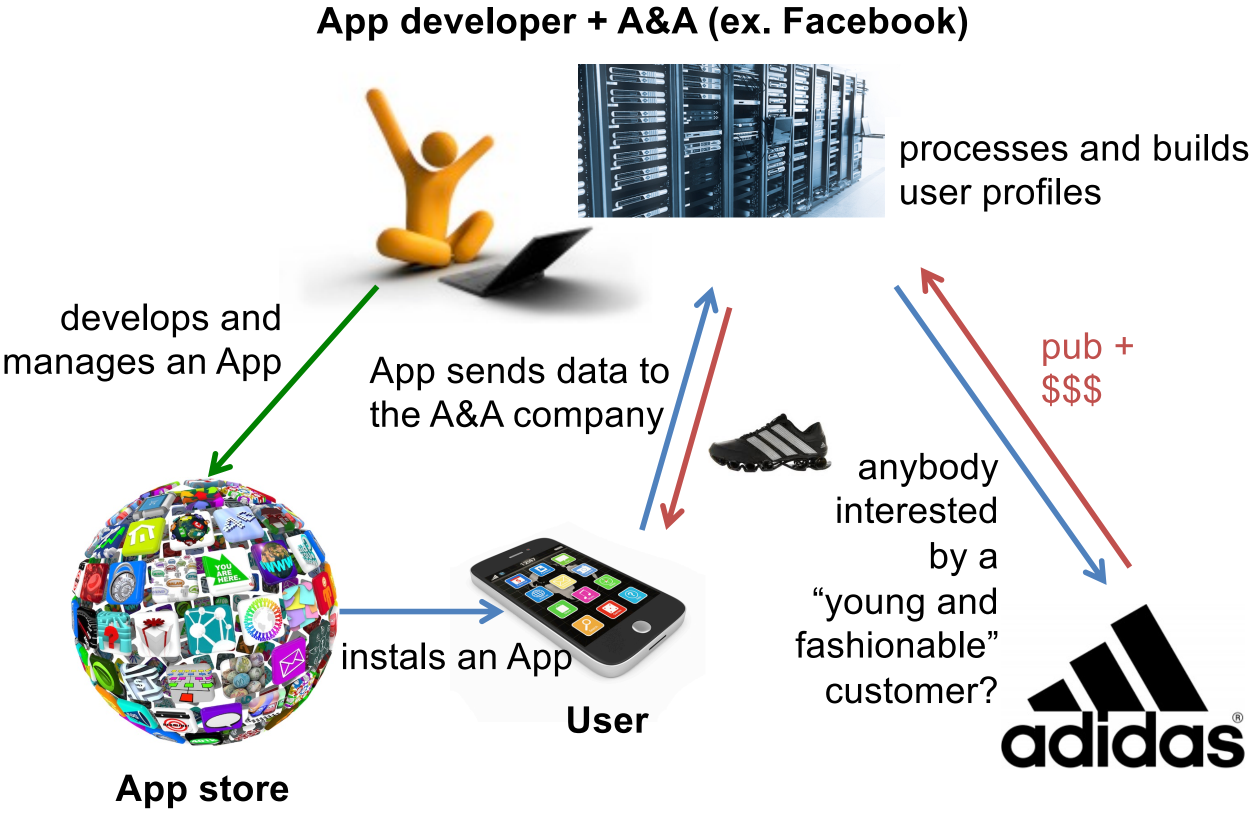
anybody
interested
by a
“young and
fashionable”
customer?

instals an App

User

App Store

The actors… a variant



App developer + A&A (ex. Facebook)

processes and builds
user profiles

develops and
manages an App

App sends data to
the A&A company

pub +
$$$

anybody
interested
by a
“young and
fashionable”
customer?

instals an App

User

App store

About mobile advertising

- many companies

AdMob by Google

millennial media

iAd

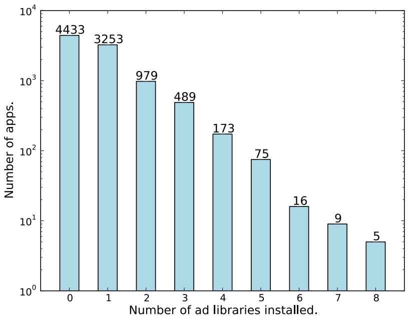
FLURRY

criteo

  - **>8 B$** of revenues for mobile advertising in 2013 for Google

About mobile advertising… (cont.')

- a few facts for Android (2011 data)
  - 77% of 50 free Apps are supported by ad. [1]
  - 35% of free Apps use at least two A&A libraries [2]
    - **in the hope to earn more?**



  - A&A libs require additional authorizations
    - **a free App asks 2-3 additional authorizations WRT paying version of the App [1]**

- **[1] "Don't kill my ads! Balancing Privacy in an Ad-Supported Mobile Application Market", HotMobile 2012.**
- **[2] "AdSplit: Separating smartphone advertising from applications", Usenix Security 2012.**

The smartphones and personal information eco-system

- **where is the problem?**

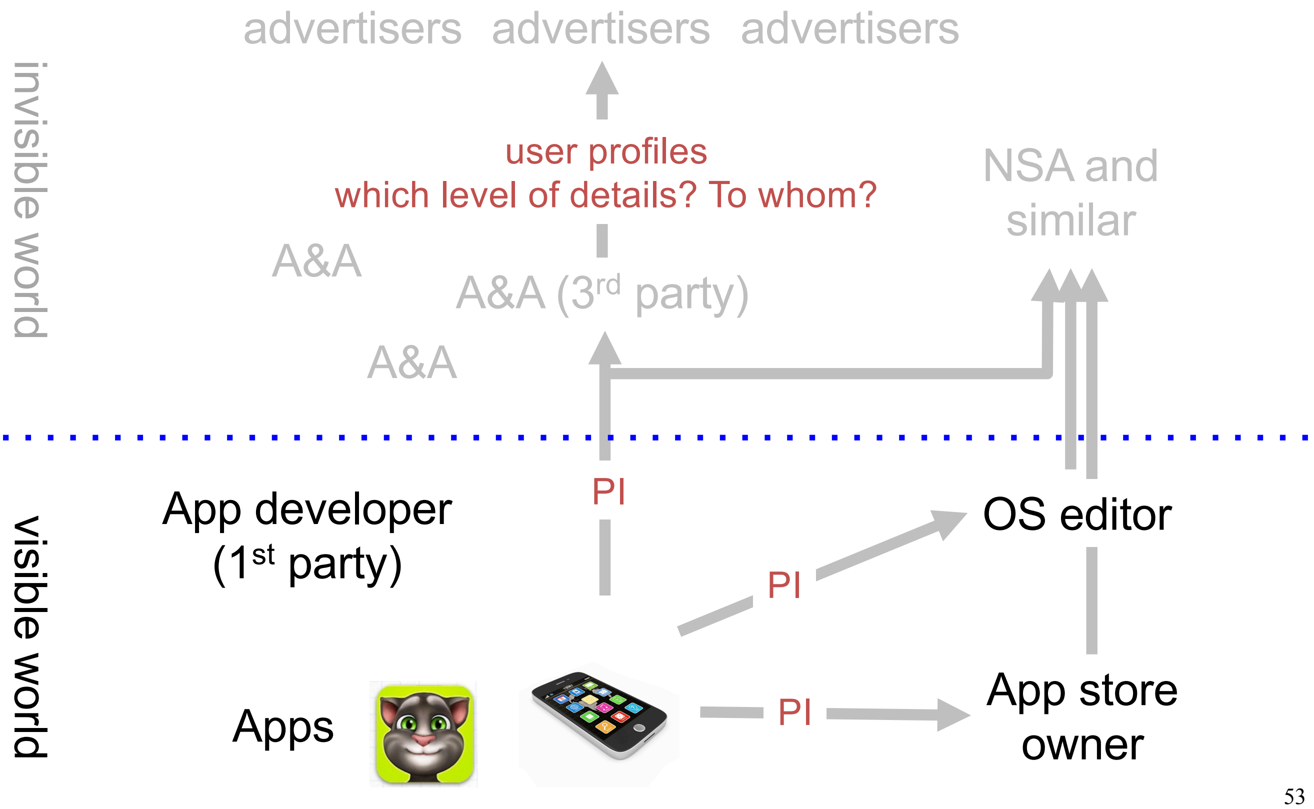
Where is the problem?

- just another business model?

> « Les données personnelles sont le nouveau pétrole de l'internet et la nouvelle monnaie du monde numérique. »
>
> ***M. Kuneva, Commissaire europ. à la consommation, 2009***

- maybe the price to pay for free Apps/services, but…

1- The ecosystem is so complex we cannot trust all actors



invisible world

advertisers advertisers advertisers

user profiles
which level of details? To whom?

NSA and similar

A&A

A&A (3rd party)

A&A

visible world

App developer (1st party)

PI

OS editor

PI

Apps

PI

App store owner

2- There are unreasonable practices

- a collect of our PI that is:

MASSIVE

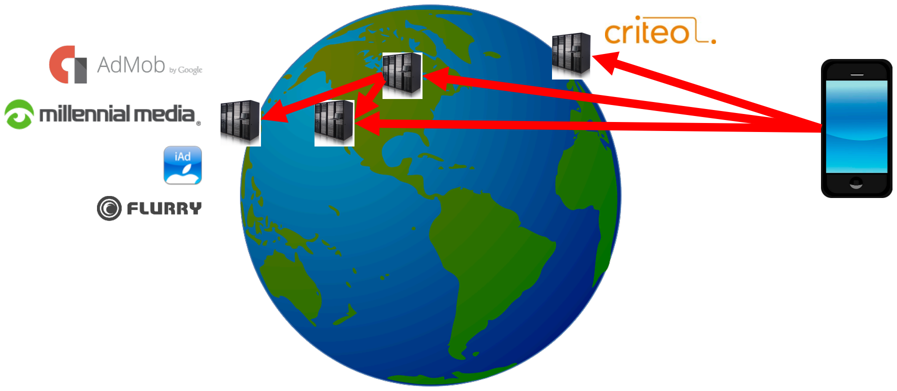
disproportionate

unnoticed

- It's not in line with FR/EU law

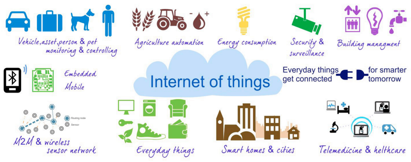
3- Uncontrolled collection of our PI

- data is immediately **exfiltrated** beyond EU in order to be stored, processed or exchanged in unknown conditions, **without any control**
    - ❍**FR and EU laws apply difficultly in those countries**
- under FR law, a user must be able to access, correct and withdraw his PI which is not always the case here!



And it's just the beginning…

- PI collection will be more and more intrusive:
    - ❍generalization of smartphone payment
    - ❍wearable connected devices
    - ❍home connected appliances
        - ❍**e.g., intelligent thermometer**
    - ❍"quantified self" trend
    - ❍connected cars
    - ❍IoT



Outline

1. introduction
   - ❍two examples
   - ❍“personal information” and the French/EU law
2. smartphones and personal information eco-system
   - ❍why are we here?
   - ❍let’s come back to smartphones
   - ❍who does what, who earns what?
   - ❍free in exchange of targeted advertising: where's the problem?
3. the Mobilitics project
4. a few ideas and results from Mobilitics
   - ❍the OS manufacturer approaches to control PI
   - ❍the case of the "ACCESS_WIFI_STATE" Android permission
   - ❍applications: a rush towards stable identifiers
   - ❍the RATP application, 2013 version
   - ❍tracking in the physical world with the smartphone Wifi interface
5. conclusions

The Mobilitics Inria-CNIL project

Inria informatiques mathématiques CNIL

- Jan.-2012 – Dec. 2014
- focuses on Android et iOS
  - ❍because they dominate
- analyze personal information leaks in **Apps** and **OS services**

The Mobilitics Inria-CNIL project… (2)

- **compare** the two ecosystems
  - what are the PI access possibilities?
  - how can a user control the situation?

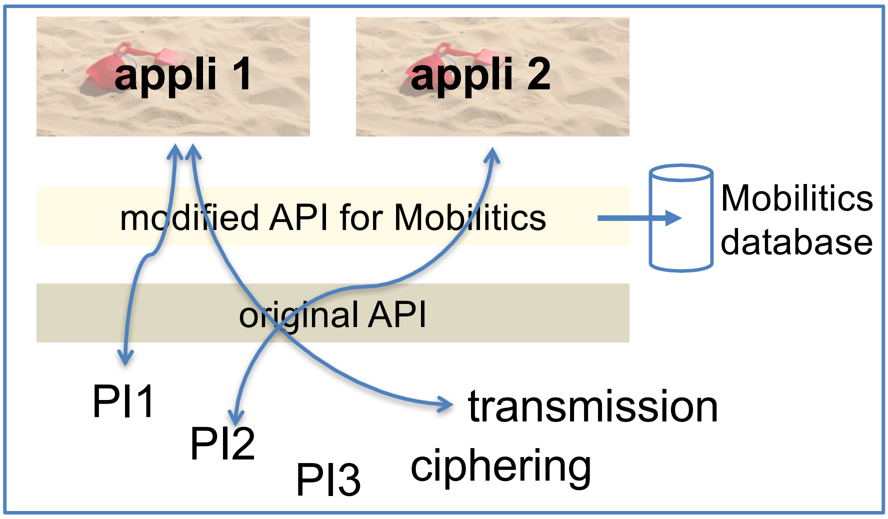
- **highlight** practices

"tracking the trackers"

- reputation is a powerful lever to convince stakeholders to change their behavior if need be
- it's complementary to the legal actions
- provide raw data and facts

The Mobilitics Inria-CNIL project… (3)

- Mobilitics, this is:
  - ❍instrumented versions of iOS and Android



  - ❍à postériori analysis tools
  - ❍in-lab experiments…
  - ❍and "in vivo" experiments with volunteers

Outline

1. introduction
   - ❍two examples
   - ❍“personal information” and the French/EU law
2. smartphones and personal information eco-system
   - ❍why are we here?
   - ❍let’s come back to smartphones
   - ❍who does what, who earns what?
   - ❍free in exchange of targeted advertising: where's the problem?
3. the Mobilitics project
4. a few ideas and results from Mobilitics
   - ❍the OS manufacturer approaches to control PI
   - ❍the case of the "ACCESS_WIFI_STATE" Android permission
   - ❍applications: a rush towards stable identifiers
   - ❍the RATP application, 2013 version
   - ❍tracking in the physical world with the smartphone Wifi interface
5. conclusions

A few ideas and results from Mobilitics

1. the OS manufacturer approaches to control PI
2. the case of the "ACCESS_WIFI_STATE" Android permission
3. applications: a rush towards stable identifiers
4. the RATP application, 2013 version
5. tracking in the physical world with the smartphone Wifi interface

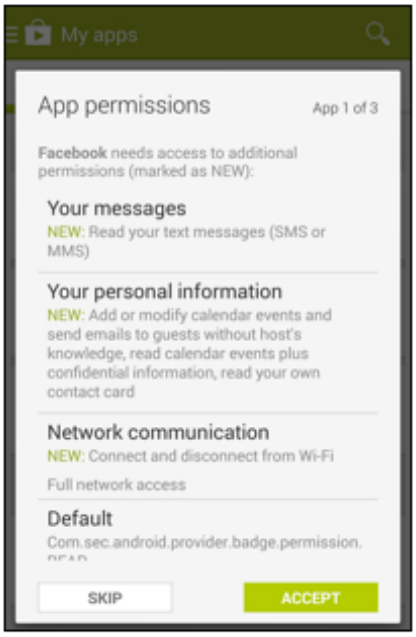
Complementary approaches

- several approaches
  - **market centric:** the market owner checks the App before accepting it
  - **user centric:** ask for the user consent…
    - **… upon installing the App**
    - **… or dynamically, when using the App**

About installation based authorizations

Google/Android

- an App having specific requirements asks for user consent at installation time
  - **responsibility** is transferred to the user
  - very basic approach



App permissions

App 1 of 3

Facebook needs access to additional permissions (marked as NEW):

Your messages
NEW: Read your text messages (SMS or MMS)

Your personal information
NEW: Add or modify calendar events and send emails to guests without host's knowledge, read calendar events plus confidential information, read your own contact card

Network communication
NEW: Connect and disconnect from Wi-Fi
Full network access

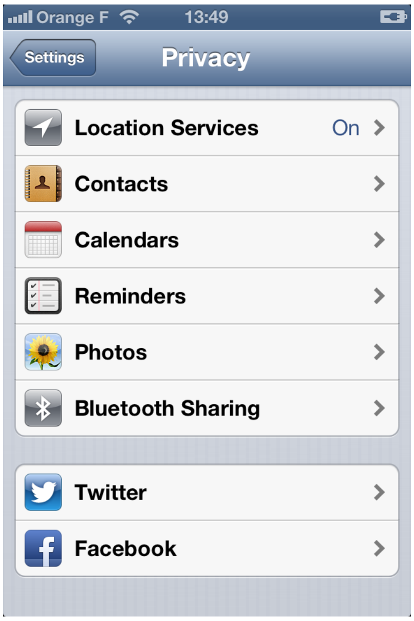
Default
Com.sec.android.provider.badge.permission.READ

About dynamic authorizations

Essentially Apple/iOS

(also quickly introduced in Android 4.3, then removed)

- a dedicated control panel enables users to authorize or ban access to PI of each App
  - ❍ **responsibility** is transferred to the user but this latter can change its mind at any time
  - ❍ here since iOS 6… and progressively improved



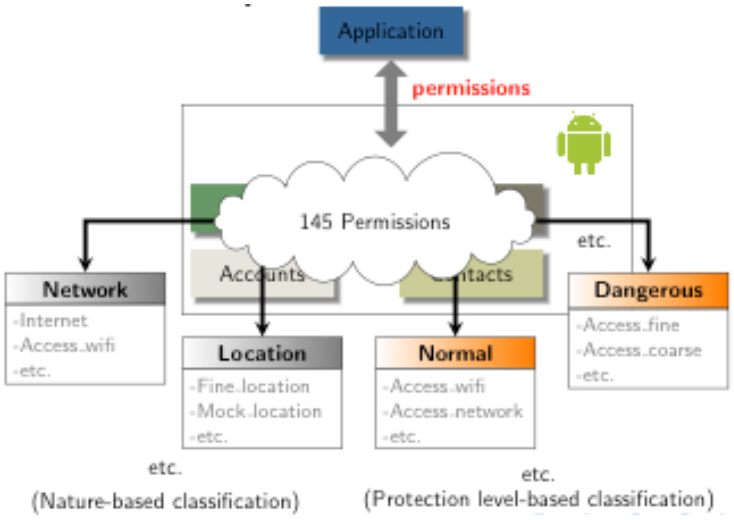
Orange F 13:49

Settings | Privacy

- Location Services On >
- Contacts >
- Calendars >
- Reminders >
- Photos >
- Bluetooth Sharing >
- Twitter >
- Facebook >

A complex authorization system…

- 145 different types of authorizations



Application

permissions

145 Permissions

Accounts

Contacts

etc.

Network
-Internet
-Access_wifi
-etc.

Location
-Fine_location
-Mock_location
-etc.

Normal
-Access_wifi
-Access_network
-etc.

Dangerous
-Access_fine
-Access_coarse
-etc.

etc.
(Nature-based classification)

etc.
(Protection level-based classification)

- users won't necessarily understand the **implications**
  - ❍ **example : ACCESS_WIFI_STATE**
    - many PI can be inferred without the user being aware of it

…that is also extremely limited

- accept or go elsewhere
  - we're not living in a binary world!
- no **behavioral** control of the App
  - authorizing an App to access my location and Internet for a punctual service does not mean I authorize this App to access my geolocation every minute and to send it to foreign servers
- no control on the **composition** of authorizations
  - authorizing an App to access my contacts and Internet does not mean I authorize this App to SEND my contacts to remote servers

What about Apple?

- much better, but not yet sufficient
- no **behavioral** control of the App
  - idem
  - authorizing access to a PI does not mean I authorize any access and processing modality for this PI

A few ideas and results from Mobilitics

1. the OS manufacturer approaches to control PI
2. the case of the "ACCESS_WIFI_STATE" Android permission
3. applications: a rush towards stable identifiers
4. the RATP application, 2013 version
5. tracking in the physical world with the smartphone Wifi interface

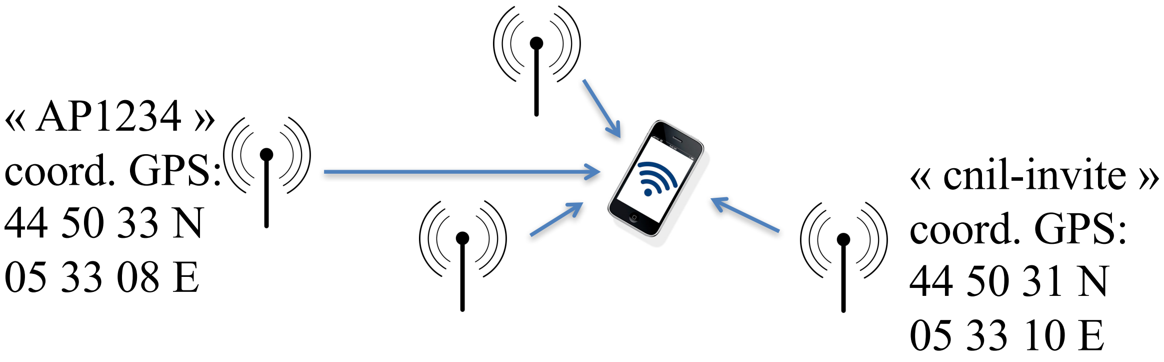
ACCESS_WIFI_STATE: an Android authorization with unexpected implications

- imagine an App, that without asking the user explicit authorization…

- … can **track** the user thanks to a stable identifier
  - it's the Wifi MAC **address**
    - **e.g.** `68:a8:6d:28:ce:1f`
  - guaranteed to be **unique** in the world
  - **impossible** to re-initialize

ACCESS_WIFI_STATE… (2)

- imagine an App, that without asking the user explicit authorization…
- … knows your **location**
  - by listening **Wifi networks in range**, then thanks to a broad database giving the geolocation of all AP can locate the smartphone by triangulation
  - in urban environments, can be **very accurate**



« AP1234 »
coord. GPS:
44 50 33 N
05 33 08 E

« cnil-invite »
coord. GPS:
44 50 31 N
05 33 10 E

ACCESS_WIFI_STATE… (3)

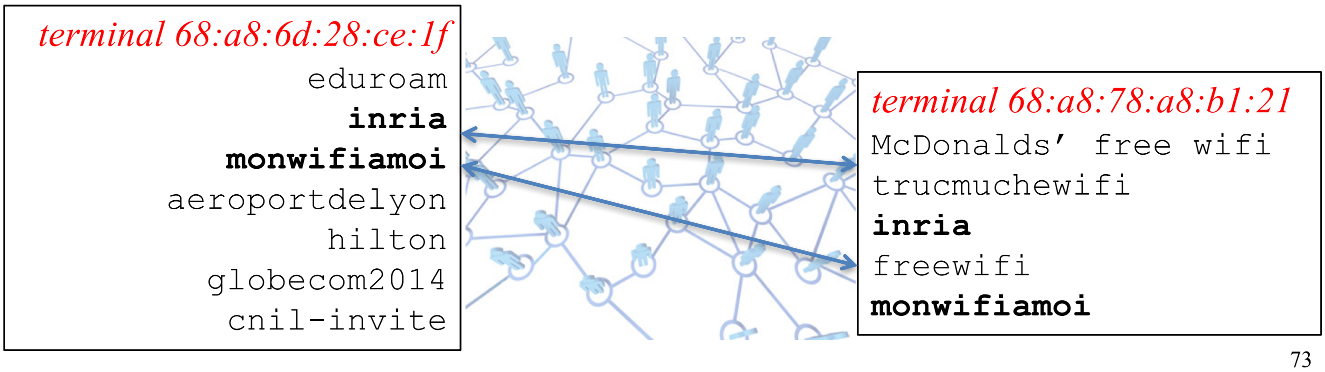
- imagine an App, that without asking the user explicit authorization…
- … knows a part of your **travels** and your **profile**
  - via the list of Wifi AP to which you connected, which is automatically registered in your smartphone

terminal 68:a8:6d:28:ce:1f

```
eduroam
Inria
monwifiamoi
aeroportdelyon
hilton
globecom2014
cnil-invite
```

ACCESS_WIFI_STATE… (4)

- imagine an App, that without asking the user explicit authorization…
- … can infer **social links** between users
  - by calculating the distance between their Wifi connection list, after creating a large dedicated database



terminal 68:a8:6d:28:ce:1f

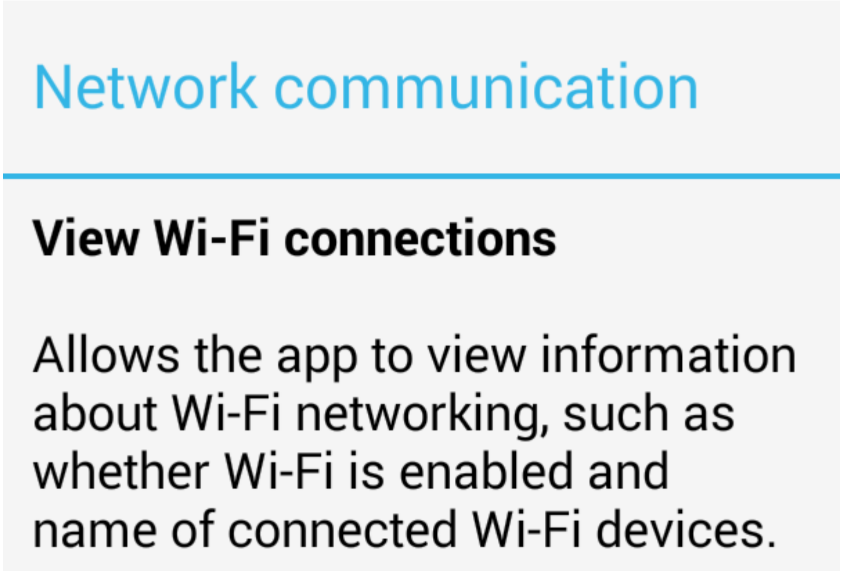
```
eduroam
inria
monwifiamoi
aeroportdelyon
hilton
globecom2014
cnil-invite
```

terminal 68:a8:78:a8:b1:21

```
McDonalds' free wifi
trucmuchewifi
inria
freewifi
monwifiamoi
```

ACCESS_WIFI_STATE… (5)

- it is sufficient to ask the **ACCESS_WIFI_STATE** and **INTERNET** authorization at installation time…
  - no user can imagine this is possible
  - and the authorization descriptions gives no clue!



Network communication

View Wi-Fi connections

Allows the app to view information about Wi-Fi networking, such as whether Wi-Fi is enabled and name of connected Wi-Fi devices.

ACCESS_WIFI_STATE: is it in use?

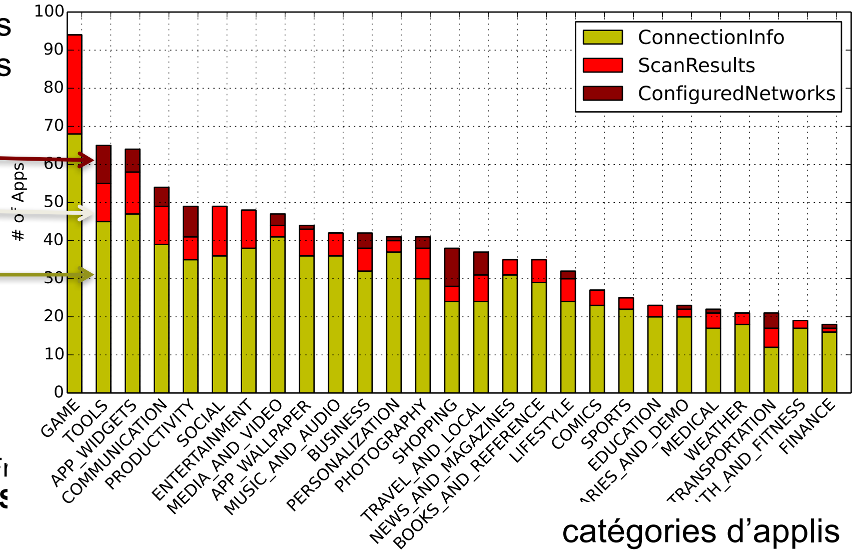
- **Yes…** Within the 2700 most popular Apps, 41% ask both permissions and many of them use it

nombre d'applis
concernées

wifi connection list

in rage Wifi AP list

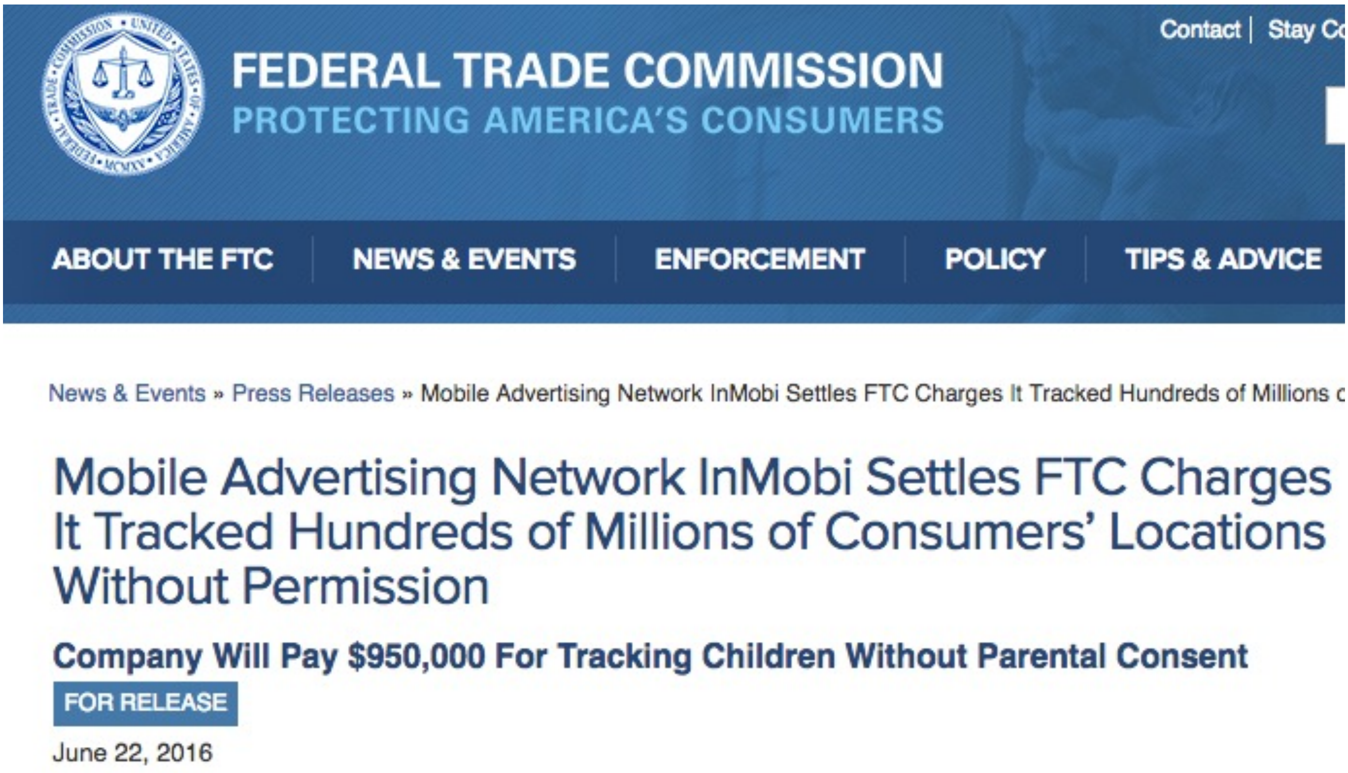
Wifi address + other info



catégories d'applis

J. Achara, M. Cunche, V. Roca, A. Fr
Implications of the ACCESS WIFI S
http://hal.inria.fr/hal-00997716/en/

Two outcomes



FEDERAL TRADE COMMISSION
PROTECTING AMERICA'S CONSUMERS

News & Events » Press Releases » Mobile Advertising Network InMobi Settles FTC Charges It Tracked Hundreds of Millions o

Mobile Advertising Network InMobi Settles FTC Charges It Tracked Hundreds of Millions of Consumers' Locations Without Permission

Company Will Pay $950,000 For Tracking Children Without Parental Consent

FOR RELEASE

June 22, 2016

Mobilitics triggered this enquiry ☺

Two outcomes… (2)

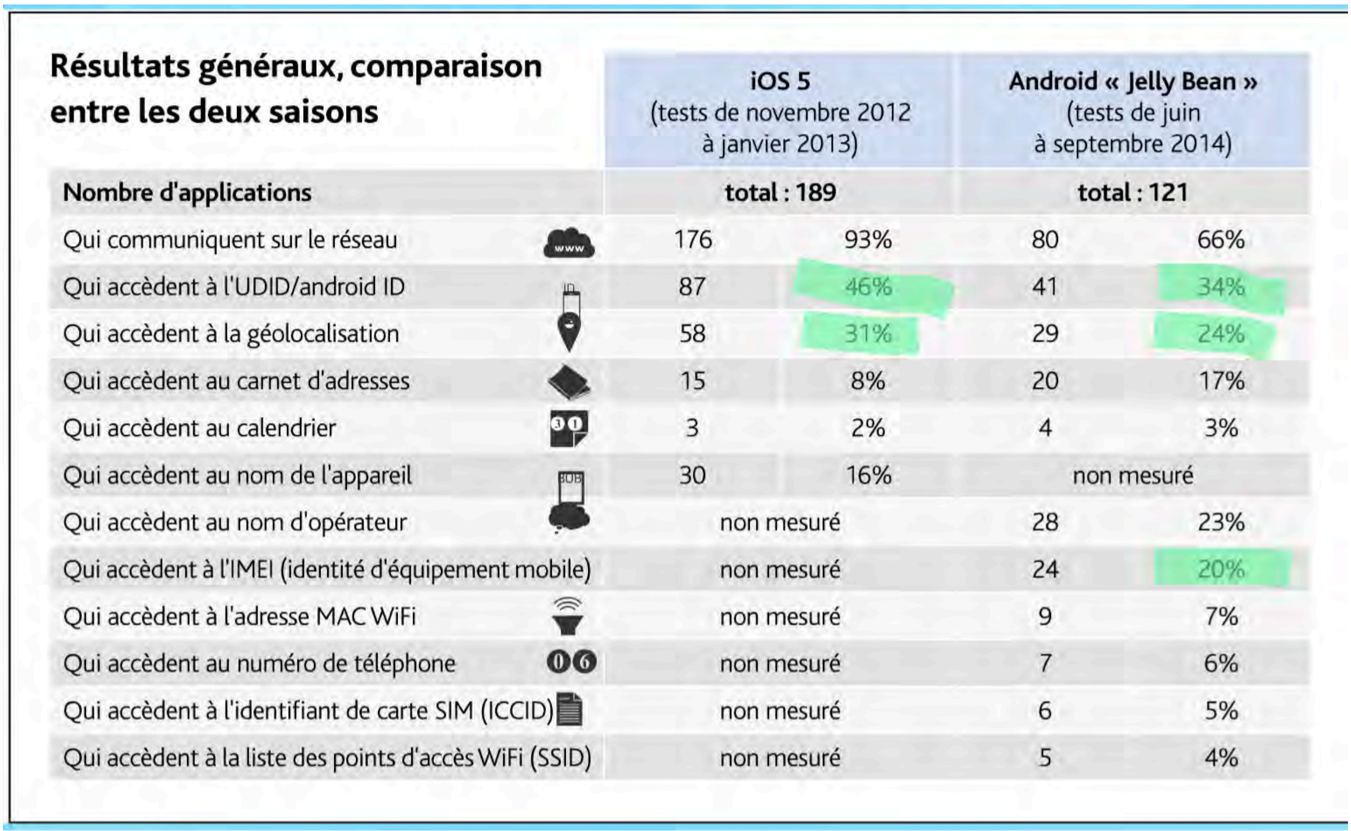
- mid-2016 Google changed a little bit the authorisation
  - ❍listening to Wifi network is now protected by the "geolocalisation" permission

Did Mobilitics triggered this enquiry?

A few ideas and results from Mobilitics

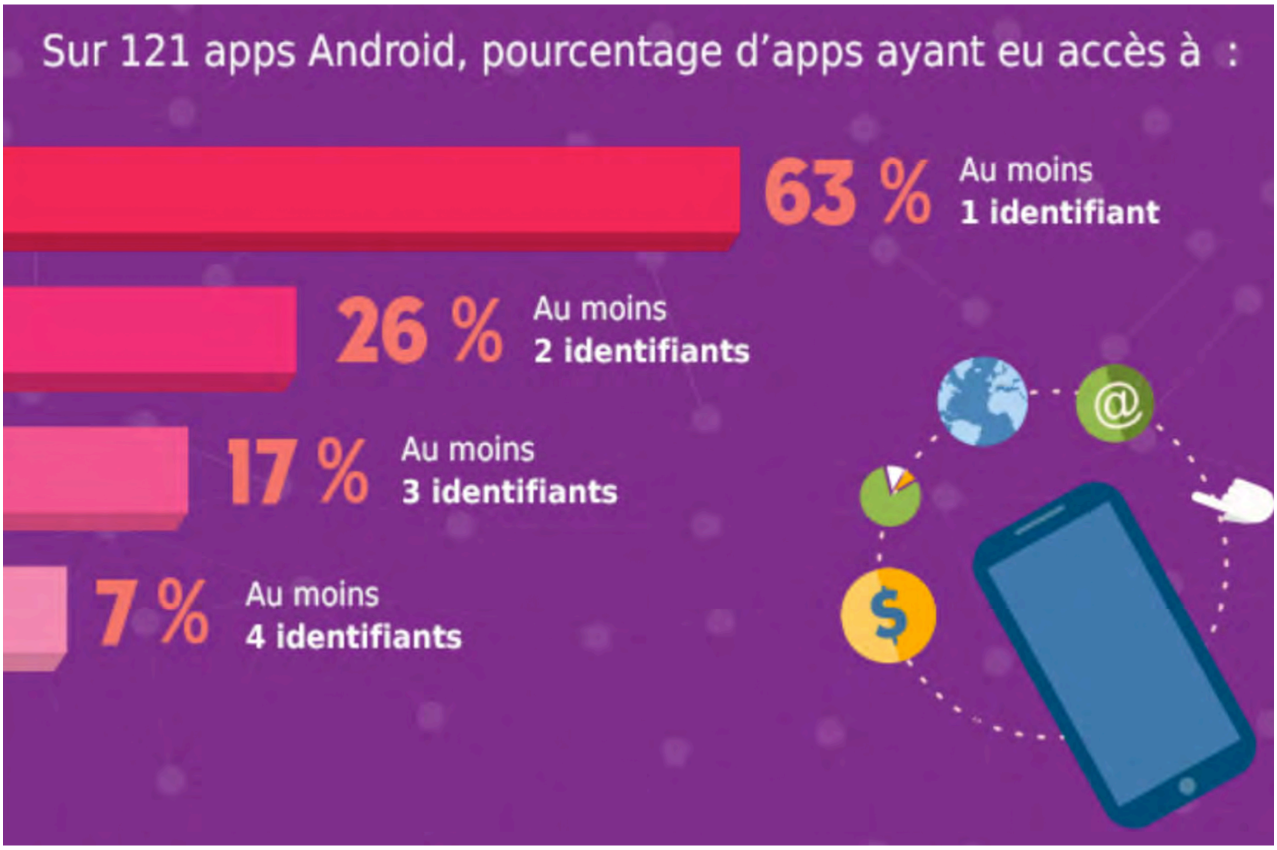
1. the OS manufacturer approaches to control PI
2. the case of the "ACCESS_WIFI_STATE" Android permission
3. applications: a rush towards stable identifiers
4. the RATP application, 2013 version
5. tracking in the physical world with the smartphone Wifi interface

A rush towards stable identifiers



Résultats généraux, comparaison entre les deux saisons	iOS 5 (tests de novembre 2012 à janvier 2013)		Android « Jelly Bean » (tests de juin à septembre 2014)	
Nombre d'applications	total : 189		total : 121	
Qui communiquent sur le réseau	176	93%	80	66%
Qui accèdent à l'UDID/android ID	87	46%	41	34%
Qui accèdent à la géolocalisation	58	31%	29	24%
Qui accèdent au carnet d'adresses	15	8%	20	17%
Qui accèdent au calendrier	3	2%	4	3%
Qui accèdent au nom de l'appareil	30	16%	non mesuré	
Qui accèdent au nom d'opérateur	non mesuré		28	23%
Qui accèdent à l'IMEI (identité d'équipement mobile)	non mesuré		24	20%
Qui accèdent à l'adresse MAC WiFi	non mesuré		9	7%
Qui accèdent au numéro de téléphone	non mesuré		7	6%
Qui accèdent à l'identifiant de carte SIM (ICCID)	non mesuré		6	5%
Qui accèdent à la liste des points d'accès WiFi (SSID)	non mesuré		5	4%

A rush towards stable identifiers… (cont.')



About stable identifiers and their use

- AndroidID

  random number generated upon starting the smartphone for the first time and kept in a stable memory

- MAC address of Wifi (or Bluetooth) interface

  identifies uniquely the network interface (e.g., `68:a8:6d:28:ce:1f`)

- IMEI (International Mobile Equipment Identity)

  uniquely identifies a smartphone (used for instance to block a stolen phone)

- IMSI (International Mobile Subscriber Identity)

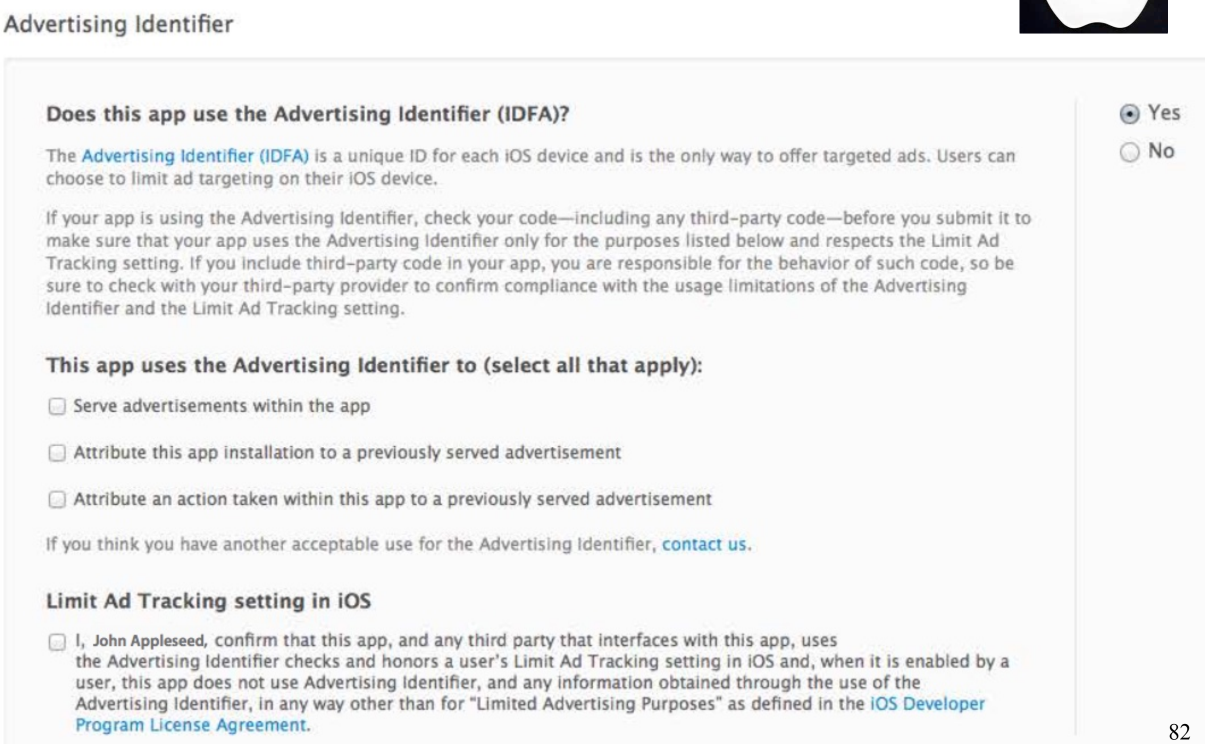
  identifies a user at his/her cell phone operator

- AdID (Advertising Identifier)

  special ID used for advertising tracking that a user can reset at any time to prevent long term tracking (in theory at least)

About the Advertising Identifier

- "Advertising Identifier" according to Apple
  - **be transparent and give control back to the user ☺**



Advertising Identifier

Does this app use the Advertising Identifier (IDFA)?

● Yes
○ No

The Advertising Identifier (IDFA) is a unique ID for each iOS device and is the only way to offer targeted ads. Users can choose to limit ad targeting on their iOS device.

If your app is using the Advertising Identifier, check your code—including any third-party code—before you submit it to make sure that your app uses the Advertising Identifier only for the purposes listed below and respects the Limit Ad Tracking setting. If you include third-party code in your app, you are responsible for the behavior of such code, so be sure to check with your third-party provider to confirm compliance with the usage limitations of the Advertising Identifier and the Limit Ad Tracking setting.

This app uses the Advertising Identifier to (select all that apply):

□ Serve advertisements within the app

□ Attribute this app installation to a previously served advertisement

□ Attribute an action taken within this app to a previously served advertisement

If you think you have another acceptable use for the Advertising Identifier, contact us.

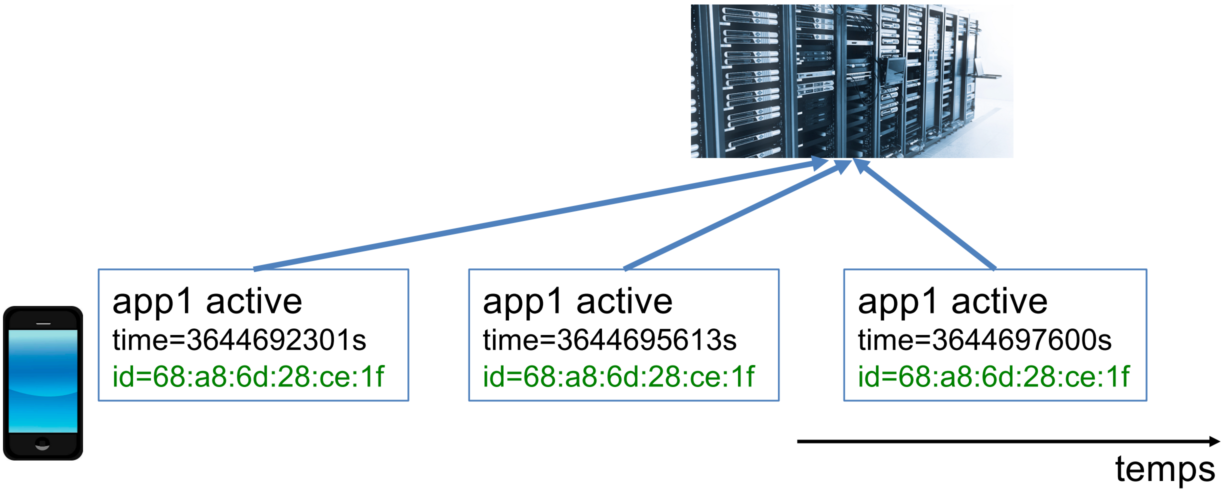
Limit Ad Tracking setting in iOS

□ I, John Appleseed, confirm that this app, and any third party that interfaces with this app, uses the Advertising Identifier checks and honors a user's Limit Ad Tracking setting in iOS and, when it is enabled by a user, this app does not use Advertising Identifier, and any information obtained through the use of the Advertising Identifier, in any way other than for "Limited Advertising Purposes" as defined in the iOS Developer Program License Agreement.

About stable IDs and their use… (cont.’)

- looks safe but…
  - considered as PI by FR/EU law

- **stable IDs** are perfect for **tracking** users on the long term



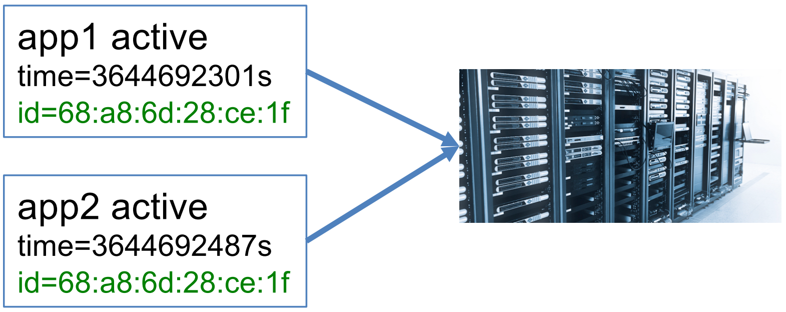
app1 active
time=3644692301s
id=68:a8:6d:28:ce:1f

app1 active
time=3644695613s
id=68:a8:6d:28:ce:1f

app1 active
time=3644697600s
id=68:a8:6d:28:ce:1f

About stable IDs and their use… (cont.’)

- **stable IDs** are perfect to **correlate** information collected from several Apps
  - **and therefore create a user profile**



app1 active
time=3644692301s
id=68:a8:6d:28:ce:1f

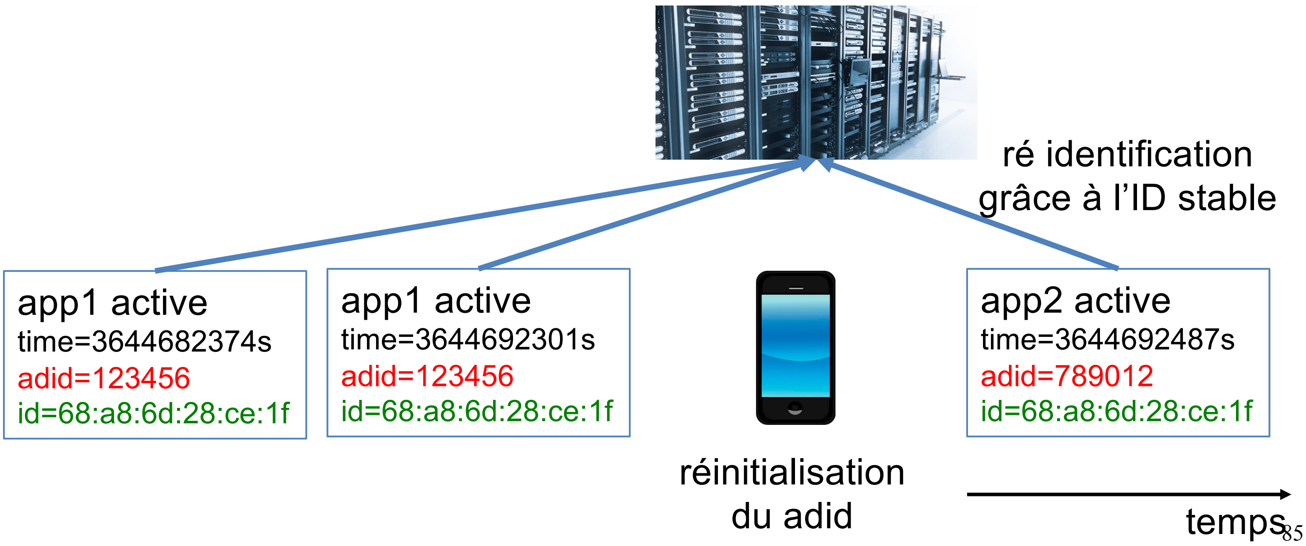
app2 active
time=3644692487s
id=68:a8:6d:28:ce:1f

From the same device? It’s sufficient to compare the IDs

If yes, we know a subset of Apps for this user and his/her centers of interest

About stable IDs and their use… (cont.')

- **stable IDs** are perfect to **bypass** the desired limits of advertising tracking
  - **if the user resets his Advertising ID, the A&A company can easily re-identify the user**



app1 active
time=3644682374s
adid=123456
id=68:a8:6d:28:ce:1f

app1 active
time=3644692301s
adid=123456
id=68:a8:6d:28:ce:1f

app2 active
time=3644692487s
adid=789012
id=68:a8:6d:28:ce:1f

To know more… (in French)

La lettre innovation et prospective de la CNIL

N°08 / novembre 2014

IP INNOVATION & PROSPECTIVE

Retrouvez-nous sur notre site [www.cnil.fr/ip] en flashant le code ou sur :

Mobilitics, saison 2 :
Les smartphones et leurs apps sous le microscope de la CNIL et d'Inria

La CNIL et Inria travaillent depuis maintenant 3 ans sur un projet de recherche et d'innovation ambitieux nommé Mobilitics. Son objectif : mieux connaître les smartphones, ces objets utilisés quotidiennement par des dizaines de millions de français et qui

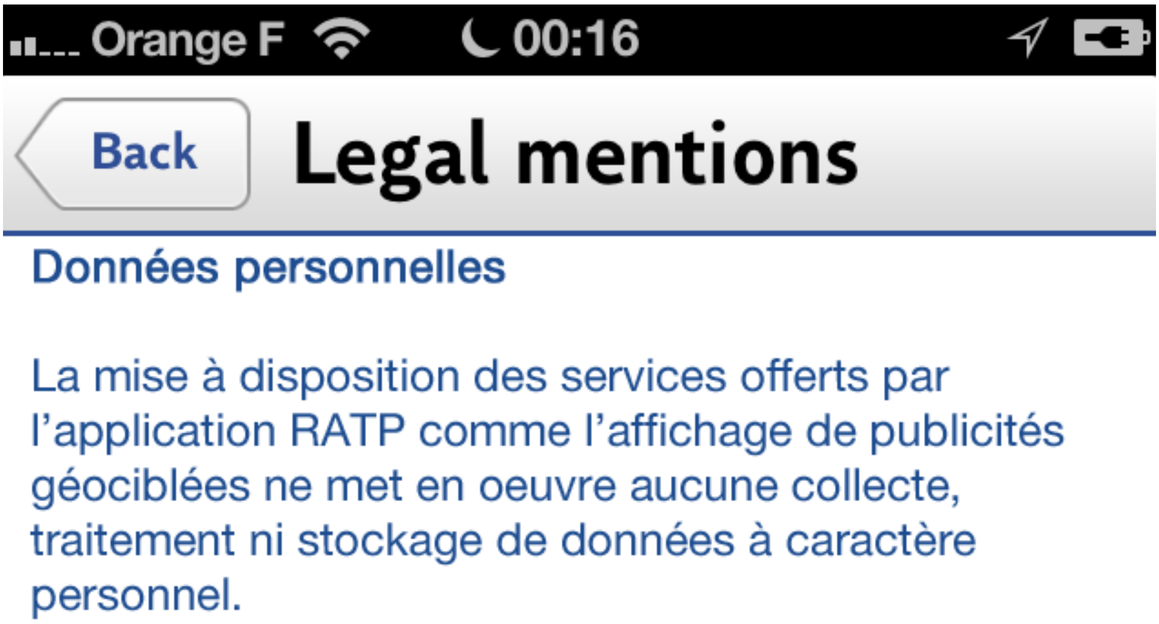
http://www.cnil.fr/fileadmin/documents/La_CNIL/publications/DEIP/Lettre_IP_N-8-Mobilitics.pdf

A few ideas and results from Mobilitics

1. the OS manufacturer approaches to control PI
2. the case of the "ACCESS_WIFI_STATE" Android permission
3. applications: a rush towards stable identifiers
4. the RATP application, 2013 version
5. tracking in the physical world with the smartphone Wifi interface

An example: the RATP App

- RATP application **version 2013**
  - according to the privacy policies, there's no collect..

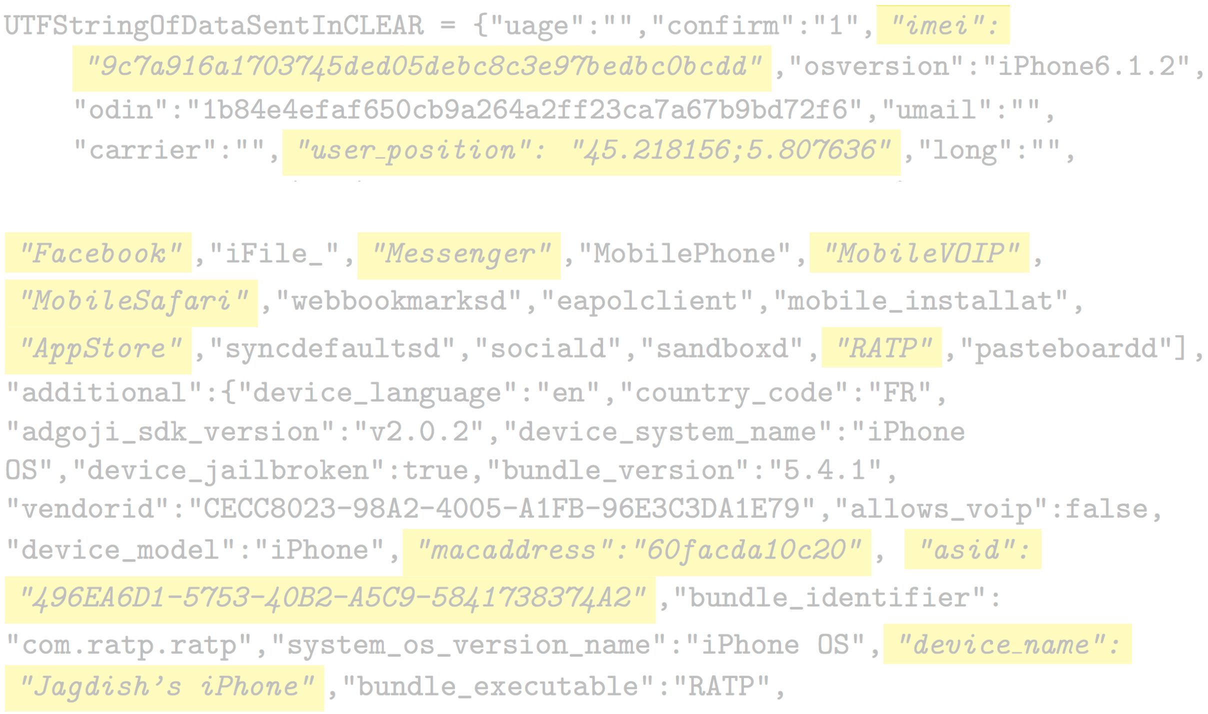


Données personnelles

La mise à disposition des services offerts par l'application RATP comme l'affichage de publicités géociblées ne met en oeuvre aucune collecte, traitement ni stockage de données à caractère personnel.

An example: the RATP App… (2)

Sent to Sofialis, an A&A, in clear-text



```
UTFStringOfDataSentInCLEAR = {"uage":"","confirm":"1", "imei":
    "9c7a916a1703745ded05debc8c3e97bedbc0bcdd","osversion":"iPhone6.1.2",
    "odin":"1b84e4efaf650cb9a264a2ff23ca7a67b9bd72f6","umail":"",
    "carrier":"", "user_position": "45.218156;5.807636","long":"",

"Facebook","iFile_", "Messenger","MobilePhone", "MobileVOIP",
"MobileSafari","webbookmarksd","eapolclient","mobile_installat",
"AppStore","syncdefaultsd","sociald","sandboxd", "RATP","pasteboardd"],
"additional":{"device_language":"en","country_code":"FR",
"adgoji_sdk_version":"v2.0.2","device_system_name":"iPhone
OS","device_jailbroken":true,"bundle_version":"5.4.1",
"vendorid":"CECC8023-98A2-4005-A1FB-96E3C3DA1E79","allows_voip":false,
"device_model":"iPhone", "macaddress":"60facda10c20", "asid":
"496EA6D1-5753-40B2-A5C9-584173837 4A2","bundle_identifier":
"com.ratp.ratp","system_os_version_name":"iPhone OS", "device_name":
"Jagdish's iPhone","bundle_executable":"RATP",
```

Sent to Adgoji, an A&A, encrypted

An example: the RATP App… (3)

- the RATP App changed quite a lot since the 2013 version, but many other applications continue…

Another example: My Talking Tom

« My Talking Tom » **accesses**

"imei": 8,
"network_code": 6,
"wifi_mac": 5,
"android_id": 12,
"operator_name": 8

« My Talking Tom » **transmits**

"android_id”:
    "85.195.69.168:(plain-text)",
    "162.217.102.42:(plain-text)",
    "vungle.com:(plain-text)",
    "sponsorpay.com:(plain-text)”
"imei”:
    "ws.tapjoyads.com:(SSL)",
    "1e100.net:(plain-text)",
    "85.195.69.168:(plain-text)",
    "outfit7.com:(plain-text)",
    "sponsorpay.com:(plain-text)”
"wifi_mac”:
    "85.195.69.168:(plain-text)",
    "vungle.com:(plain-text)",
    "sponsorpay.com:(plain-text)”

A few ideas and results from Mobilitics

1. the OS manufacturer approaches to control PI
2. the case of the "ACCESS_WIFI_STATE" Android permission
3. applications: a rush towards stable identifiers
4. the RATP application, 2013 version
5. tracking in the physical world with the smartphone Wifi interface

Tracking users in physical world thanks to their smartphone Wifi interface

- Wi-Fi tracking system[11]
  - Set of sensors collect Wi-Fi signal
  - Detect and track Wi-Fi devices and their owners
  - MAC address used as identifier

M. Cunche slide
(Inria, Privatics)

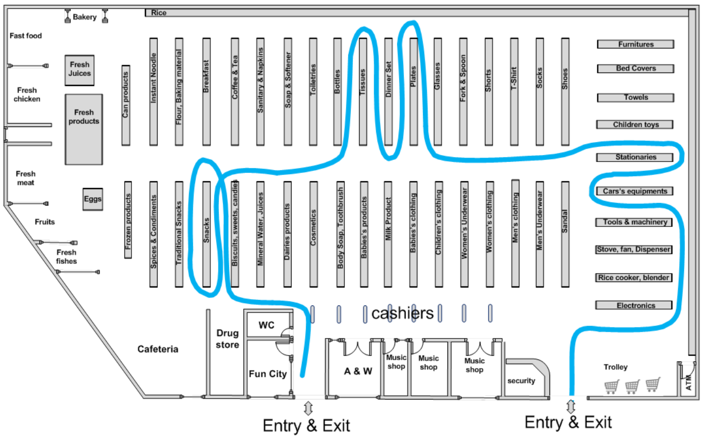
Smartphone Detection
Bluetooth
Wifi
SHOP
DEPARTMENT STORE
WiFi

[11] A. B. M. Musa and Jakob Eriksson. "Tracking unmodified smartphones using Wi-Fi monitors". In: *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. 2012.

Tracking users... (2)

- Physical analytics
  - Similar to Web Analytics
  - Frequency and length of visit, number of visitor, peak hour ....
- Trajectory reconstruction
  - Signal received by several sensors
  - Triangulation based on signal strength

M. Cunche slide
(Inria, Privatics)



© Inria / Photo H. Raguet

- **Conclusions**

The case of Google

- **Google business model relies on advertisements**
- …and Google needs PI for that
  - ❍**Apps have an easy access to (stable) identifiers needed to track users**
    - sometimes without having to ask user authorization
  - ❍**very limited motivation to change the situation**
    - since August 2014, new Apps are supposed to only use the "Advertising ID" for targeted advertising…
    - … but it will take time and other identifiers still remain
    - current strategy remains to collect as many IDs as possible
  - ❍**and contrary indicators exist**
    - Android 4.3 proposed a privacy dashboard… Removed from the following Android versions!

The case of Google… (cont.')

- but this is (partially) an open-source OS
  - building secure versions is possible ☺
    - **BlackPhone2 (Silent Circle)** 500 $
      - https://silentcircle.com/services#blackphone

    - **CryptoPhone 500 (GSMK)** 3500 $
      - http://www.cryptophone.de/en/products/mobile/cp500/
      - can identify faked cell towers
        - http://www.popsci.com/article/technology/mysterious-phony-cell-towers-could-be-intercepting-your-calls
        - http://www.aftenposten.no/nyheter/iriks/Secret-surveillance-of-Norways-leaders-detected-7825278.html
      - usually those are "IMSI catchers"

The case of Apple

- **Apple sells (costly) hardware and softwares**
- … and communicates a lot on privacy

  Tim Cook, PDG Apple : « Notre activité ne repose pas sur le fait de détenir des informations sur vous. Vous n'êtes pas notre produit »

  - **even if the situation is not perfect, there is are clear improvements across iOS versions**
    - many stable identifiers have been removed from the latest iOS versions
    - the AdID that a user can re-initialize is key to limit tracking
- don't be naïve… the goal is to sell more devices!
  - **but the company's position matches that of the citizen (for the moment)**

The user can also

- limit the number of Apps
  - be careful W.R.T. the App permissions asked or the privacy control dashboard
  - … and remove unused Apps
  - think it twice before using a daily assistant like “Google Now”
- use official App stores
  - Apps are checked (up to a certain point) by the store owner
- switch off the Wifi interface if not used…
  - to avoid physical tracking by stores (and others)
- …and if you can, switch off data communications
  - when not used

The user can also… (cont.’)

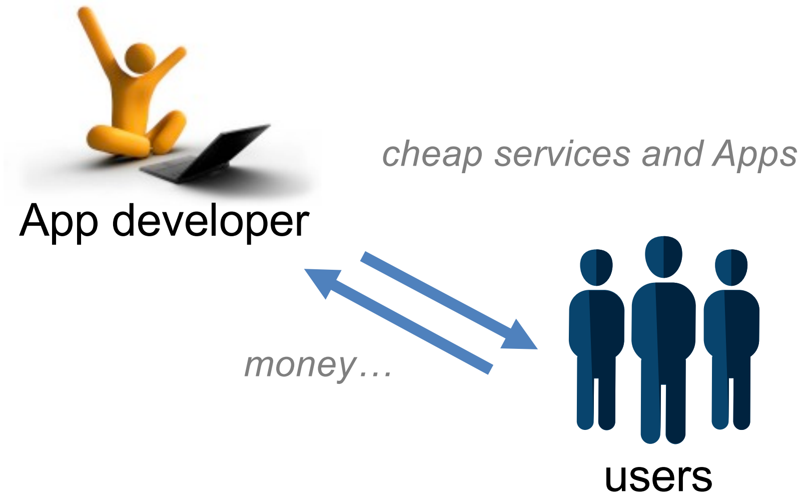
- explicitly stop Apps
  - instead of leaving them running in background
- set appropriate geolocation parameters
- limit advertising tracking / reset the AdvertisingID
  - with iOS, in case of Android it's useless
- “last but not least”, do not jaibreak/root your phone
  - otherwise any App has a full access to smartphone

Fortunately the regulator has a real power

- the **EU laws** continue to evolve in the right direction
    - ❍new EU regulation on data protection
    - ❍true impacts on companies
    - ❍EU data protection agencies (e.g., CNIL in France) discuss in the G29 group

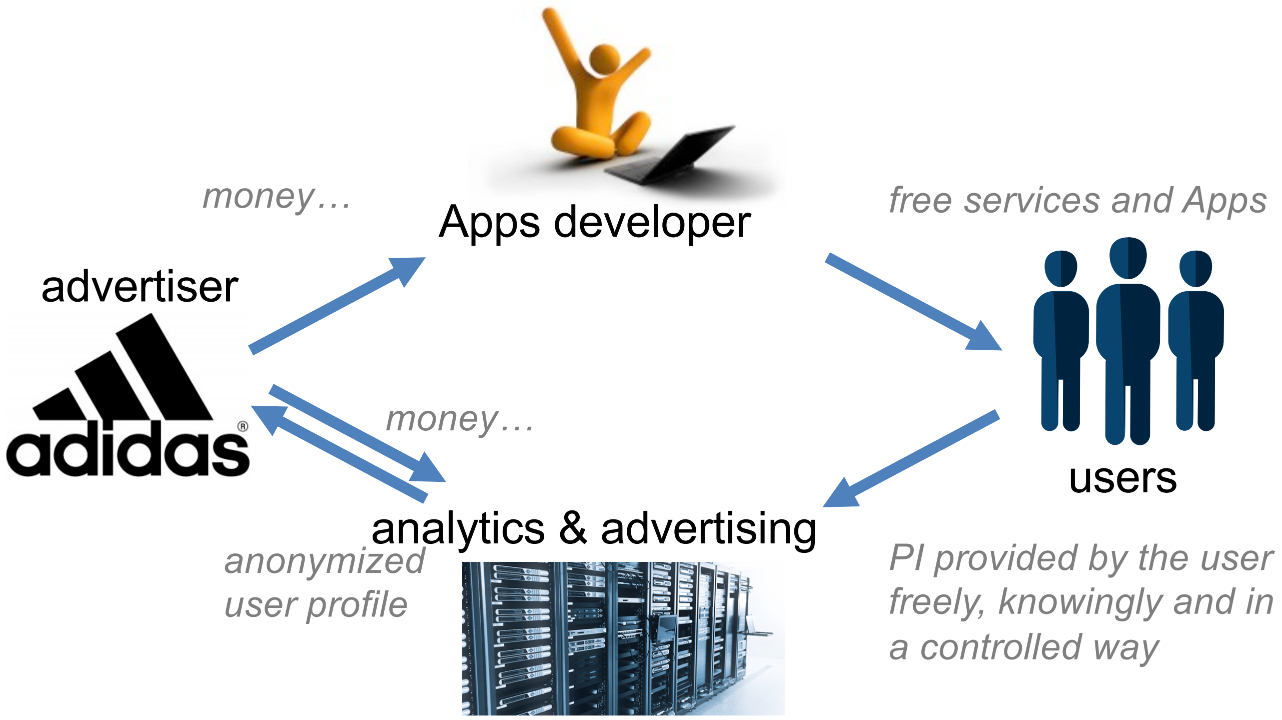
Toward a virtuous circle

Paying model



Toward a virtuous circle… (2)

"Free" model



money…

Apps developer

free services and Apps

advertiser

users

money…

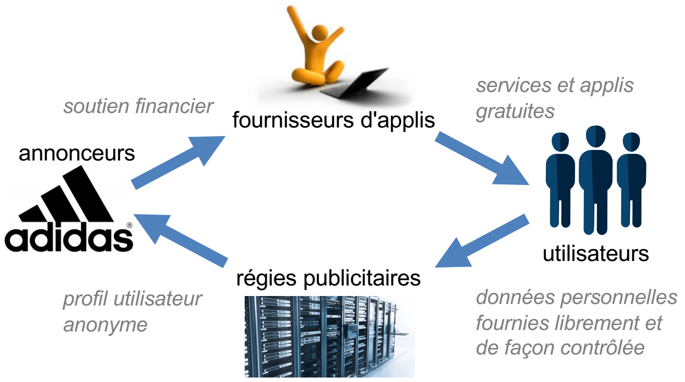
analytics & advertising

anonymized user profile

PI provided by the user freely, knowingly and in a controlled way

There are a few preliminary conditions

- users
  - should have **control** over information they provide
- each actor
  - should be **transparent** WRT practices ("transparency")
  - should be able to **prove** practices ("accountability")
- trusted third parties are needed
  - in order to **check** practices



soutien financier

fournisseurs d'applis

services et applis gratuites

annonceurs

utilisateurs

régies publicitaires

profil utilisateur anonyme

données personnelles fournies librement et de façon contrôlée

An utopia?

- not necessarily!
- market with a strong information asymmetry are known to be fragile
  - it cannot work for long periods
- … it's everybody's interest

Thank you… ☺

vincent.roca@inria.fr

Hey! What about my privacy??
They say that people who worry about their privacy have something to hide ...
sangrea.net