Vie Privée et Smartphones

© Inria / Photo H. Raguet

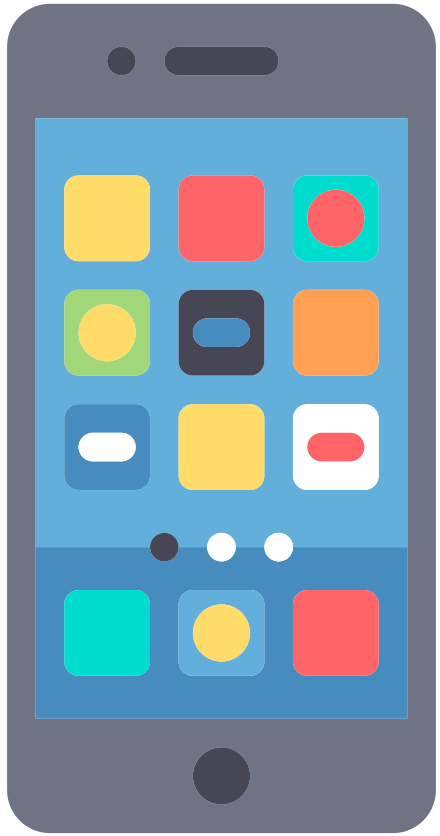
Vincent Roca, Inria PRIVATICS, vincent.roca@inria.fr

Colloque CAPPRIS, 7 novembre 2016

informatiques mathématiques

contact : vincent.roca@inria.fr

Ne soyons pas naïfs...



- tous les jours nous utilisons…
  - des **services** de qualité gratuits
  - des **applications** de qualité gratuites
- possible grâce à un modèle économique basé en grande partie sur la **publicité ciblée**
  - l'annonceur paie à la place de l'utilisateur
- …ce qui requiert un **profilage** des utilisateurs
  - afin de connaitre leurs centres d'intérêts

… mais tout de même !

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

- où sont les problèmes ?
- comment les détecter ?
- cela marche t-il ?

- **L'écosystème "smartphones"**

Le smartphone au centre des collectes

- notre “compagnon” de tous les jours…
  - …pratique, sympa, toujours connecté, facilement personnalisable

- …qui :

concentre

des infos personnelles

lorsqu’on l'utilise : téléphone, SMS, web, application bancaire, etc.

génère

des infos personnelles

GPS, NFC, WiFi, caméra, capteurs d’empreintes, capteurs de rythme cardiaque, etc.

Le smartphone au centre des collectes (2)

- il en sait beaucoup sur nos activités sur Internet **et** dans le monde physique
  - **les applications ont beaucoup d'opportunités pour faire fuiter des infos personnelles**
  - **explique que des sites web vous incitent à installer leur App**

Notre mouchard de poche préféré ?

Les multiples acteurs

développeur (first party)

inclue une lib. A&A dans l'App

$

Advertising & Analytics (A&A) (third party)

agrège et construit des profils utilisateurs

développe et gère une appli.

magasin d'applis

installe l'appli.

utilisateur

l'appli. envoie des données à la société A&A

pub

intéressé par un utilisateur « jeune & mode »?

pub + $$$

adidas

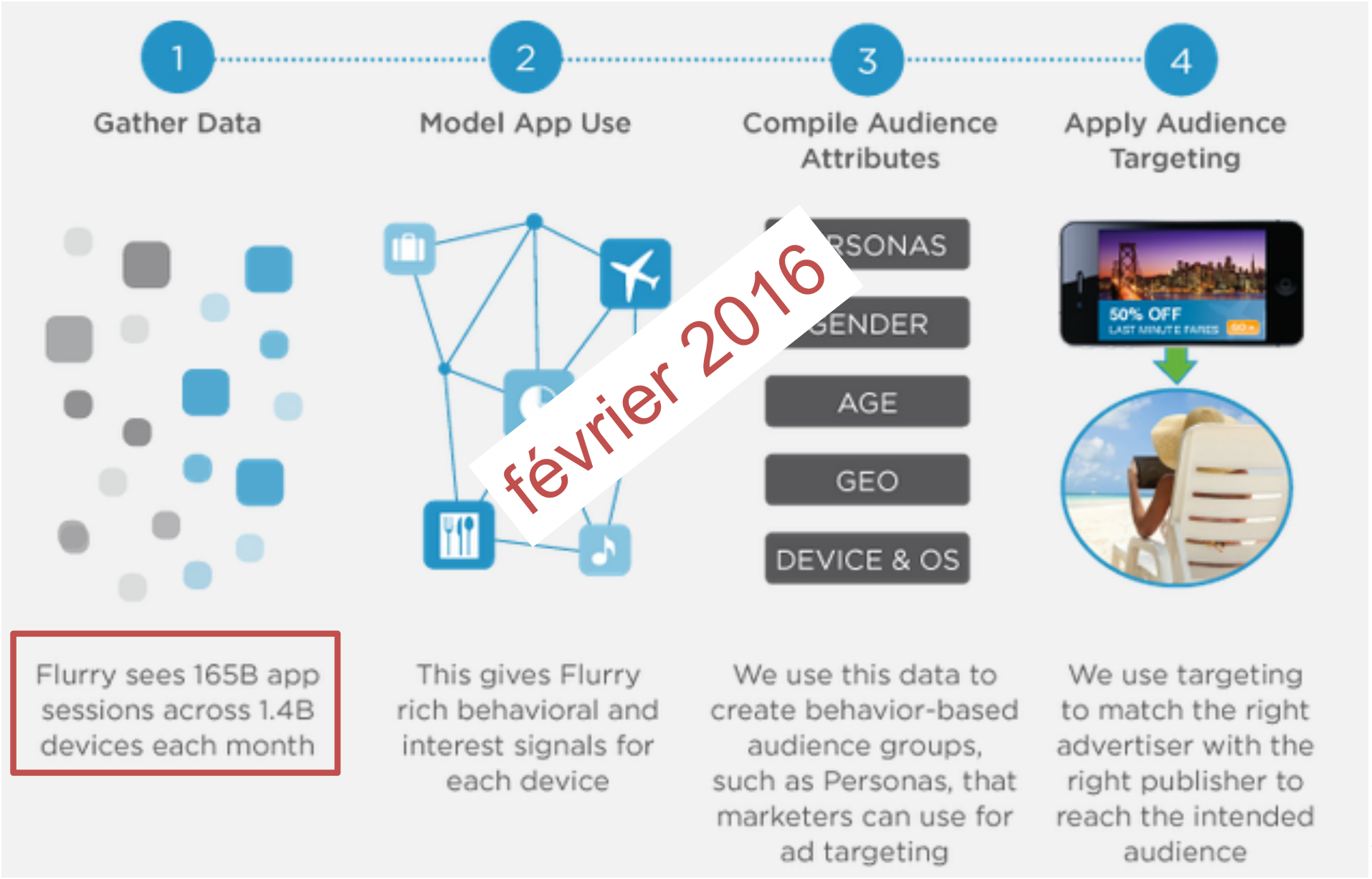
Annonceur

Les multiples acteurs (2)

- centré autour de la **régie publicitaire**
  - ou "Analytics & Advertising" (A&A), aussi "third party"
  - rôles :
    - **à l'interface entre annonceurs et applications**
    - **collecte les infos utilisateurs**
    - **trace les utilisateurs : quelles applis utilise un utilisateur, à quelle fréquence, à quel moment ?**
    - **crée et enrichit progressivement les profils utilisateurs**
    - **créent les enchères en temps réel lorsqu'un bandeau publicitaire est disponible**
  - manipule des petaoctets (10^{12}) de données utilisateurs !
  - fait pour la publicité ciblée, mais les données sont aisées à détourner

Un exemple : Flurry (from Yahoo)

http://www.flurry.com



Un modèle économique comme un autre ?

- gratuité des services/applications contre publicité ciblée
  - ce qui nécessite d'avoir des informations sur l'utilisateur
  - ce peut être un modèle économique raisonnable…

- actuellement il y a des **problèmes de fond**

1- Un écosystème complexe où il est impossible de faire confiance à tous

2- Des collectes qui peuvent être massives et disproportionnées

- un exemple : historique des positions enregistrées par mon smartphone Android pour les services Google

  - **se connecter avec le compte Gmail utilisé sur son smartphone Android à l'URL :**

    https://maps.google.com/locationhistory/

Est-ce bien raisonnable ?

- Google sait où je travaille, où j'habite, ce que je fais dans ma journée, comment je me déplace…
  - **vous aussi maintenant ;-)**

Est-ce bien raisonnable ?

- … ceci avec une précision incroyable
  - ci-contre la liste des positionnements dans la base de données Google
    - **un relevé toute les 5mn durant la nuit**
    - **… et toutes les minutes en période d'activité !**

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

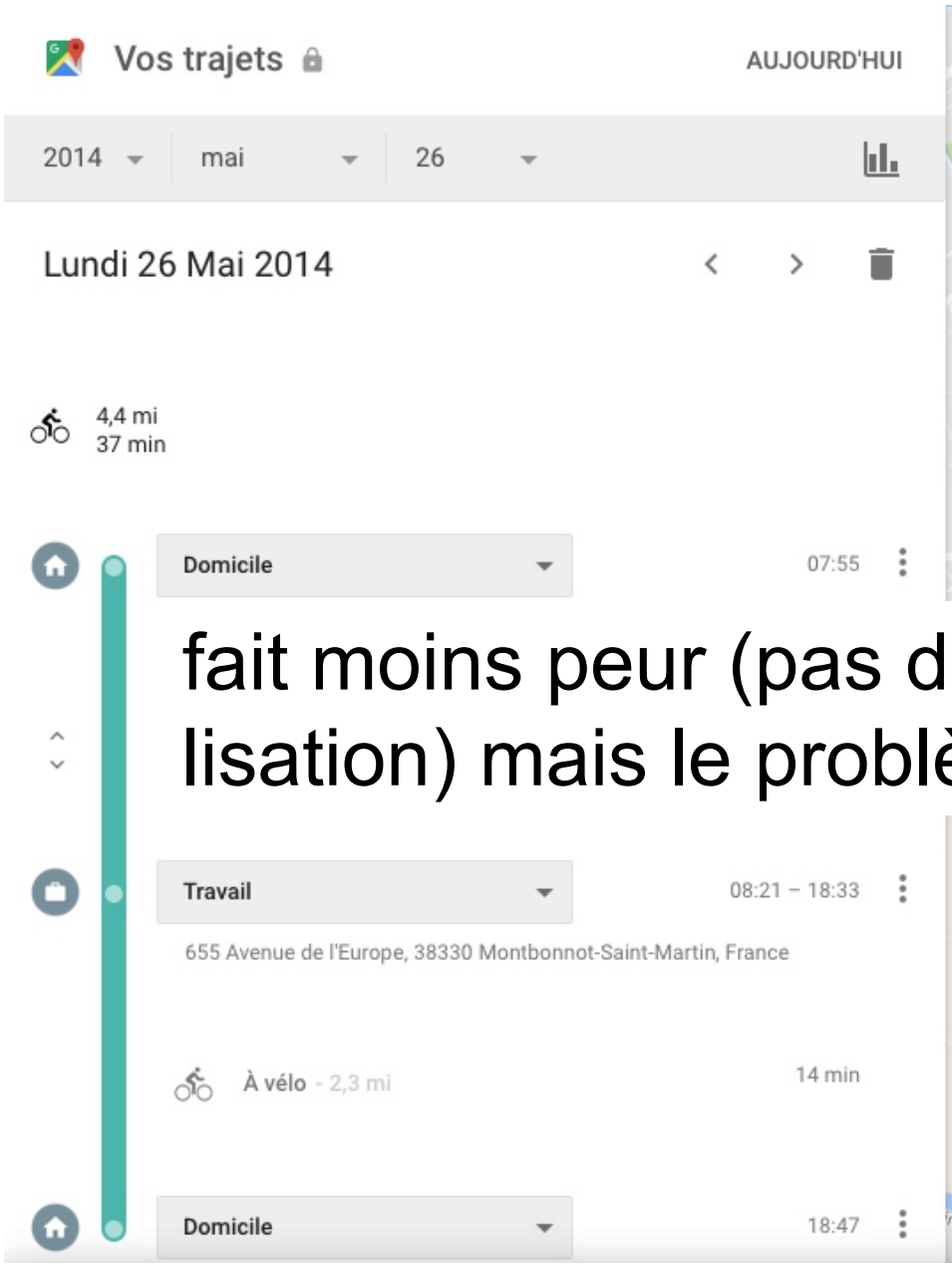
08:00 - 09:00
08:00 08:01 08:02 08:03 08:04 08:05
08:06 08:07 08:08 08:09 08:11:05
08:11:59 08:12 08:18 08:21 08:24
08:25 08:26 08:27 08:28 08:29 08:30
08:31 08:32 08:37 08:42 08:47 08:51
08:56

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

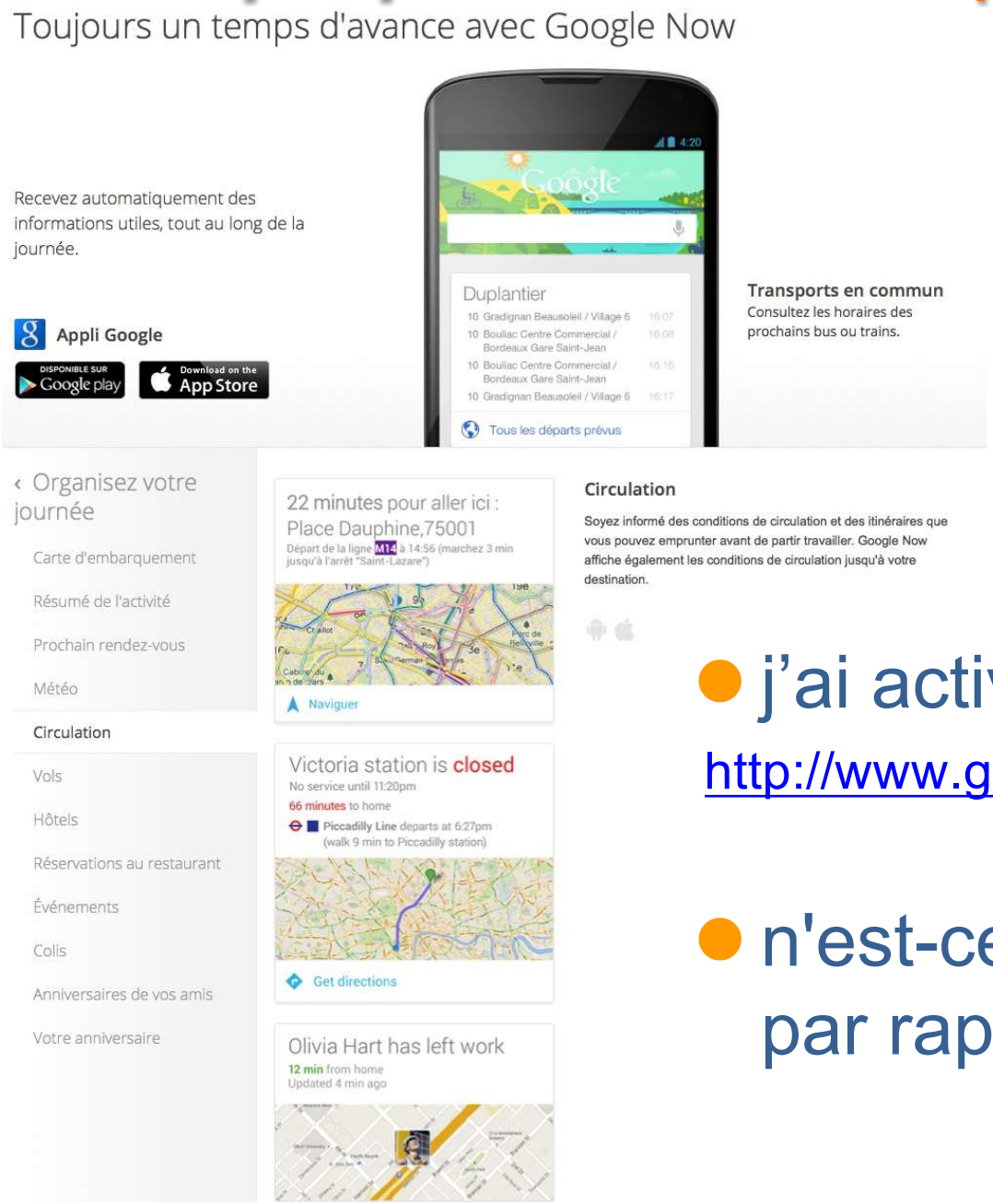
11:00 - 12:00
11:00 11:04 11:09 11:14 11:19 11:23
11:28 11:33 11:38 11:42 11:47 11:52

Au passage, Google a revu sa présentation



fait moins peur (pas de liste détaillée de géoloca-lisation) mais le problème reste entier !

2- Des collectes qui peuvent être massives et disproportionnées… (5)



- j'ai activé Google Now !

  http://www.google.com/landing/now/

- n'est-ce pas disproportionné par rapport au service rendu ?

2- Des collectes qui peuvent être massives et disproportionnées… (6)

- les données de géolocalisation sont porteuses de sens !
  - Google sait si je fréquente des lieux de culte
  - Google sait si je fréquente des lieux de soin (hôpital, clinique, médecin)

  - ce sont des **données sensibles** au sens de la loi Informatique et Liberté de 1978
    - **NE PEUVENT PAS faire l'objet de collecte ou traitement !**

3- Des collectes discrètes

- on ne sait pas tout, loin de là…
- un exemple : l'application RATP **version 2013**

  - en principe, il n'y a aucune collecte..

Des collectes discrètes (2)

Envoyé à Sofialis, une régie publicitaire, non chiffré


```
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-A5C9-5841738374A2","bundle_identifier":
"com.ratp.ratp","system_os_version_name":"iPhone OS", "device_name":
"Jagdish's iPhone","bundle_executable":"RATP",
```

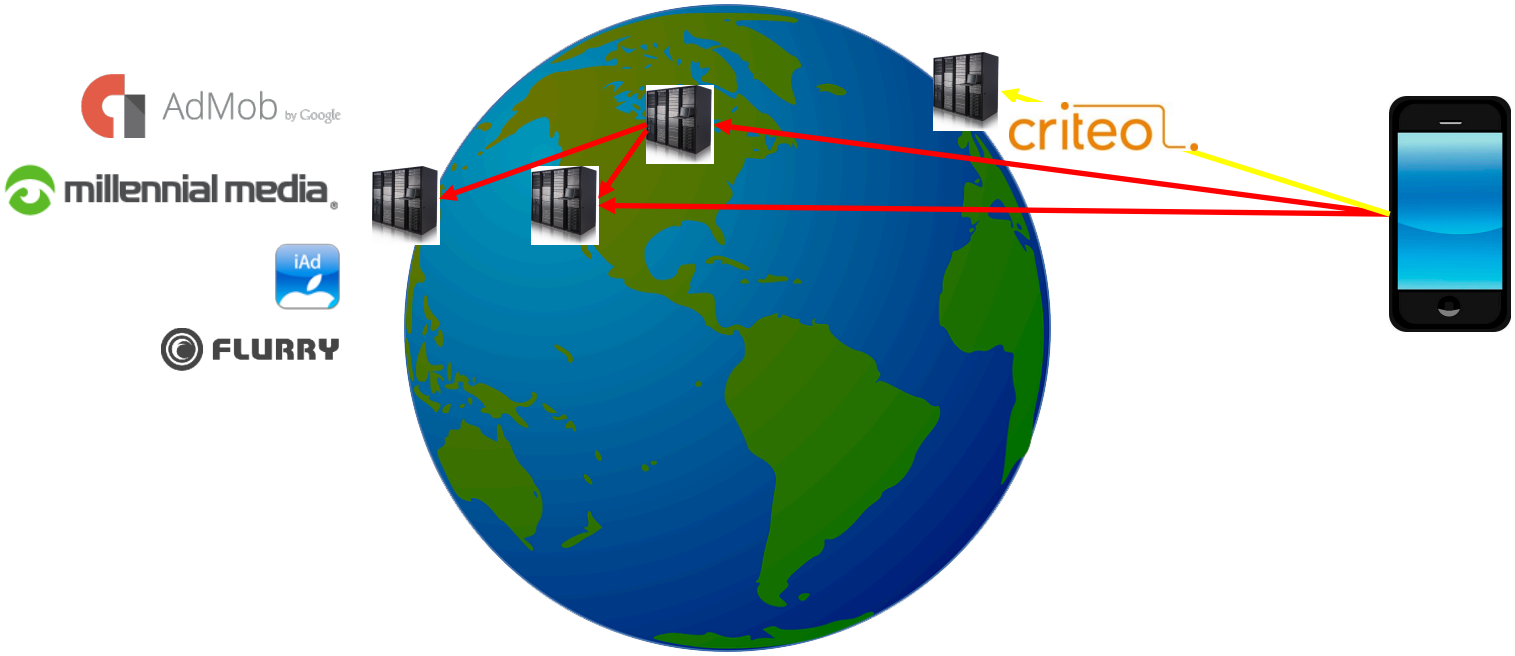

Envoyé à Adgoji, une régie publicitaire, chiffré

Des collectes discrètes (3)

- l'application RATP a bien changée depuis cette version 2013, mais de nombreuses autres applications continuent !

4- Des collectes qui échappent à tout contrôle...

- immédiatement **exfiltrées** à l'étranger où la législation FR et EU aura du mal à s'appliquer



- **sans garantie** quant au stockage, sécurité, exploitation, échanges avec d'autres acteurs

- **le projet Mobilitics Inria-CNIL et les travaux Jagdish Achara (PhD) dans CAPPRIS**

Le projet Mobilitics Inria-CNIL

- 2012 - 2014

- s'est intéressé à Android et iOS
  - ce sont les principaux OS
  - en 2014, Android ≃ 82%, iOS ≃ 15%

informatiques mathématiques

Commission Nationale de l'Informatique et des Libertés

Le projet Mobilitics Inria-CNIL (2)

- a permis de **comparer** les écosystèmes

  - quelles sont les possibilités de captation de données ?
  - quels moyens a un utilisateur pour contrôler la situation ?

- a permis **d'exposer** les pratiques

"tracking the trackers"

  - la réputation est un puissant levier pour faire bouger les acteurs, complémentaire du volet légal et répressif
  - fournir des données techniques factuelles

Le projet Mobilitics Inria-CNIL (3)

- Mobilitics, c'est :
  - des versions instrumentées de iOS et Android
  - des outils d'analyse à postériori

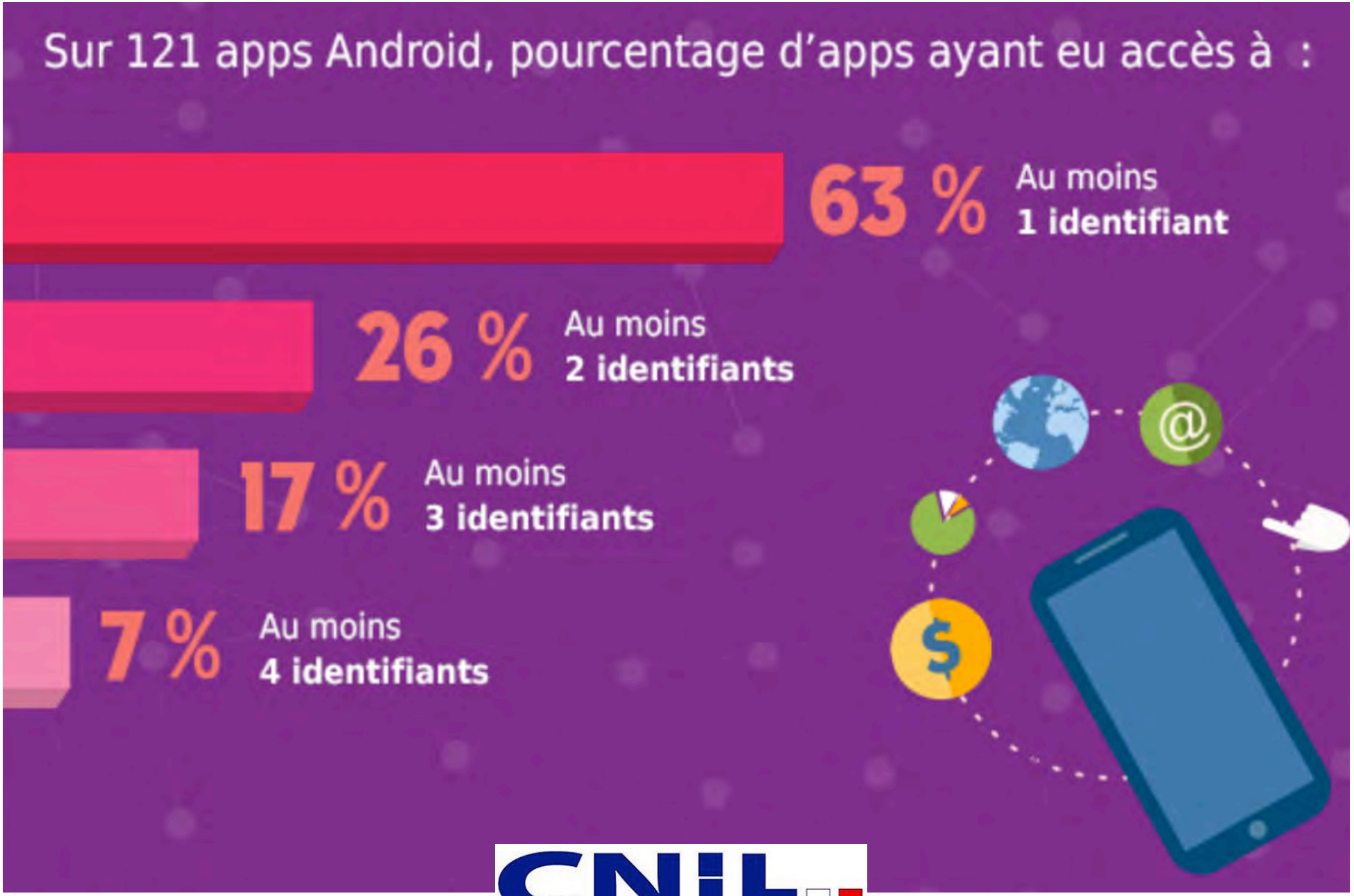
  - des expérimentations en labo…
  - des expérimentations « in vivo » avec des volontaires

Quelques résultats

- applications : la course aux identifiants

- Android : les limites du système de permissions, en particulier "ACCESS_WIFI_STATE"

La course aux identifiants



Quelques exemples d'identifiants

- AndroidID

  nombre aléatoire généré lors du premier démarrage d'un téléphone Android et conservé dans une mémoire stable

- adresse MAC de l'interface Wifi (ou Bluetooth)

  séquence de 6 nombres qui identifie de façon unique le composant Wifi du terminal (par ex. : `68:a8:6d:28:ce:1f`)

- IMEI (International Mobile Equipment Identity)

  identifie un téléphone et utilisée par exemple pour bloquer un téléphone volé

- IMSI (International Mobile Subscriber Identity)

  identifie un abonné chez son opérateur

Des infos pas si anodines

- semble anodin, et pourtant…
  - ce sont des **données personnelles** au sens de la loi I&L
  - des identifiants **stables** dans le temps, idéaux pour pour **tracer** un utilisateur sur le long terme

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

temps

Des infos pas si anodines… (suite)

- des identifiants **stables**, idéaux pour **corréler** des collectes issues d'applications distinctes
  - et affiner le **profil utilisateur**

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

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

?

Issues du même équipement ?
Il suffit de comparer les ID…

Si oui, on connait une liste d'applications utilisées… Et certains centres d'intérêts de l'utilisateur…

Des infos pas si anodines… (suite)

- des identifiants **stables**, idéaux pour **contourner** tout système de limitation de suivi publicitaire
  - *Advertising Identifier* **: un identifiant spécialisé pour le traçage publicitaire, que l'utilisateur peut réinitialiser à tout moment pour « contrôler » son traçage**

ré identification grâce à l'ID stable

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

réinitialisation du adid

temps

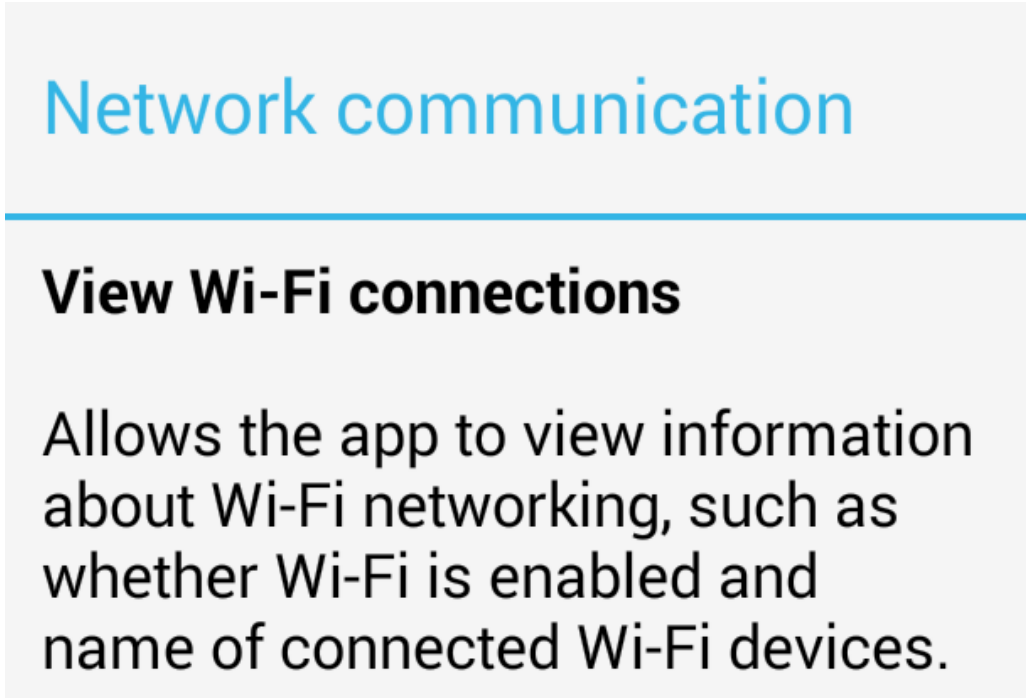
Quelques résultats (2)

- applications : la course aux identifiants

- **Android : les limites du système de permissions, en particulier "ACCESS_WIFI_STATE"**

Limites du système de permissions Android

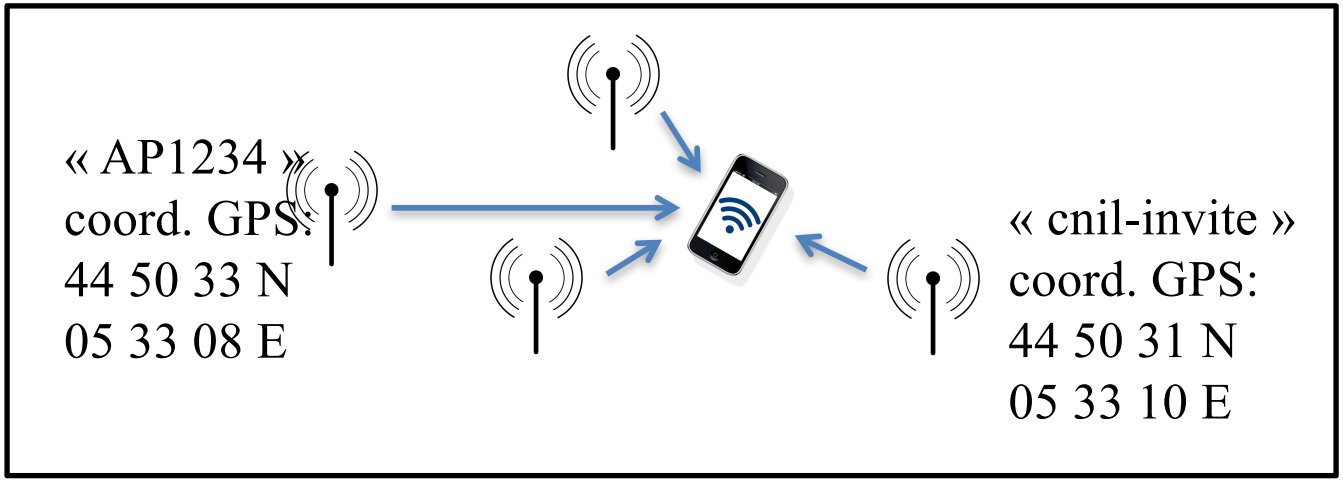
- la permission **ACCESS_WIFI_STATE**, d'apparence anodine
  - aaucun utilisateur, même attentif, ne peut imaginer toutes les implications de cette permission qualifiée de "normale"



Une permission pas si anodine que cela…

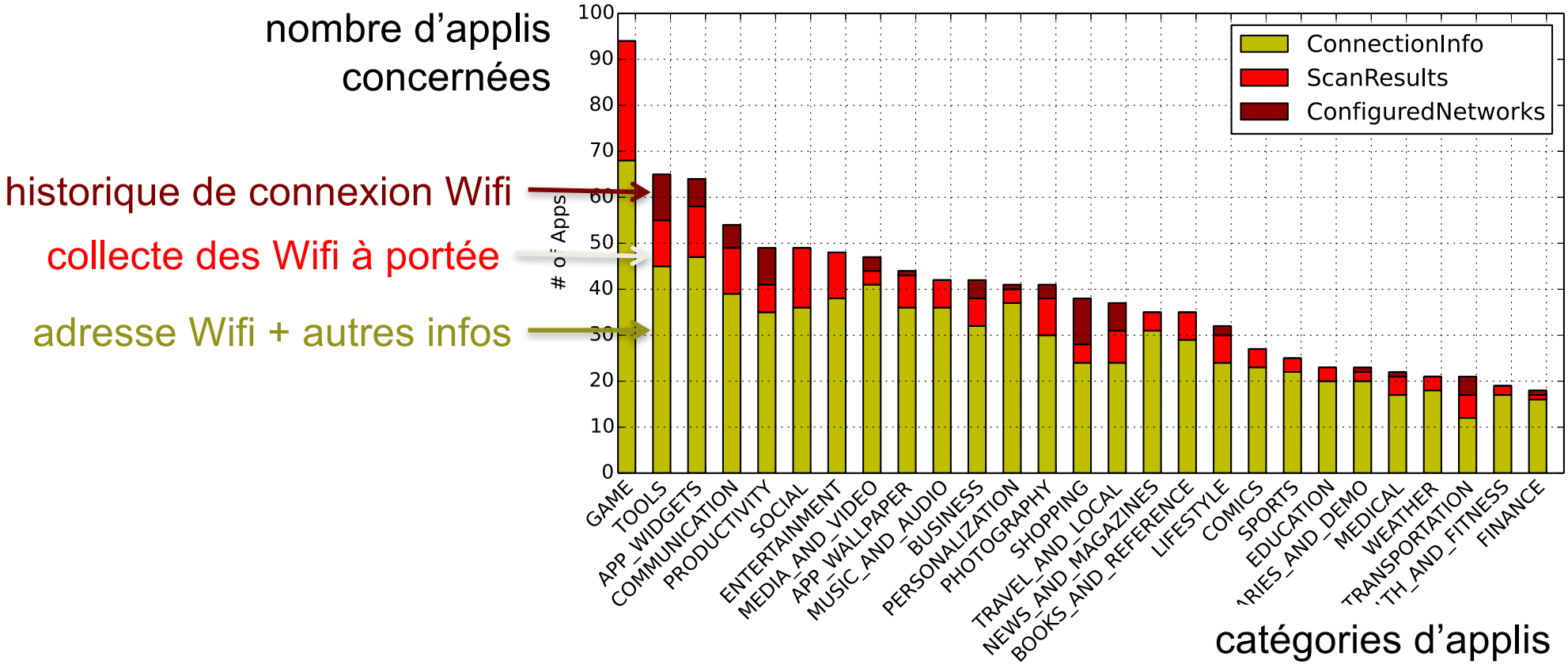
- permet de :
  - récupérer un **identifiant stable** pour tracer l'utilisateur : l'adresse MAC de l'interface Wifi
  - connaître une partie de vos déplacements via **l'historique** des réseaux Wifi où l'on s'est connecté et de vous profiler
  - vous localiser par écoute des **réseaux Wifi à portée**

terminal 68:a8:6d:28:ce:1f
eduroam
Inria
monwifiamoi
aeroportdelyon
hilton
globecom2014
cnil-invite



Ces possibilités sont elles exploitées ?

- **Oui**… Sur les 2700 applis les plus populaires, 41% demandent ces permissions et bon nombre en font usage



J. Achara, M. Cunche, V. Roca, A. Francillon, « **Short paper: WifiLeaks: Underestimated Privacy Implications of the ACCESS WIFI STATE Android Permission »,** IEEE WiSec'14, juillet 2014. http://hal.inria.fr/hal-00997716/en/

Deux retombées

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

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

C'est Mobilitics qui a déclenché cette enquête ☺

Deux retombées (2)

- mi-2016 Google a changé son système
  - l'écoute des réseaux Wifi à portée est protégé par la permission "accès à la localisation"

Est-ce Mobilitics qui en est à l'origine ?

- **Pour conclure**

Conclusion

- l'éditeur de l'OS a un rôle majeur
  - **c'est lui qui fixe les règles du jeu**
  - **on voit une différence importante entre Google et Apple… Est-ce surprenant ?**
- l'utilisateur a une maitrise limitée de la situation
  - **règles de bon sens pour limiter les risques… mais cela a des limites sur Android**
- le législateur a un rôle important
  - **Mobilitics apporte des données factuelles qui vont aider**
  - **la nouvelle réglementation Européenne va renforcer le poids de l'Europe vis à vis des acteurs étrangers**

Pour un cercle vertueux

développeurs d'applis

soutien financier

services et applis peu coûteuses

annonceurs

adidas

utilisateurs

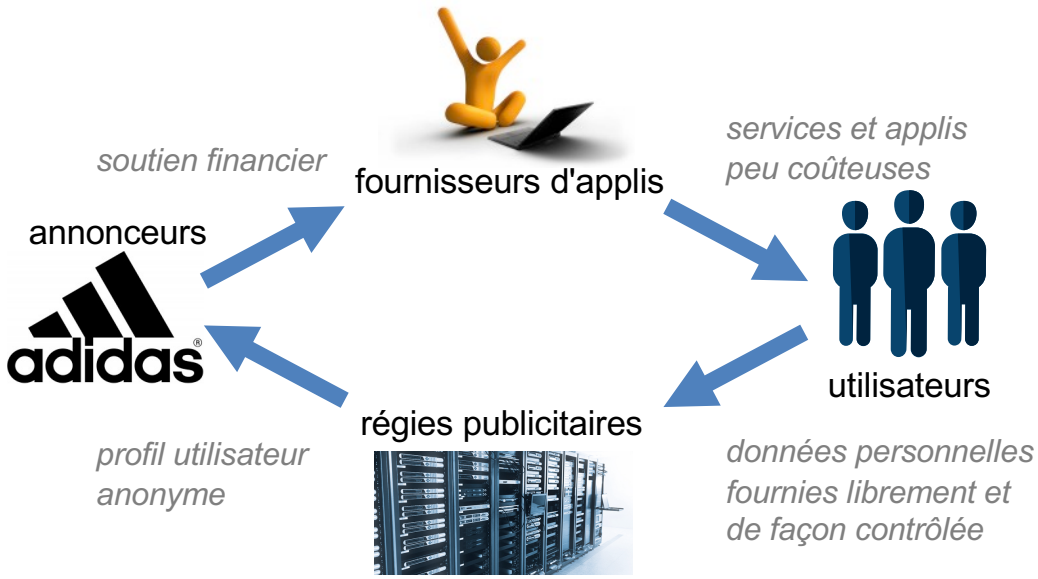
régies publicitaires

profil utilisateur anonyme

données personnelles fournies librement, sciemment et de façon contrôlée

Pour un cercle vertueux (2)

- les utilisateurs
  - ont le **contrôle** sur les informations fournies
- chaque acteur
  - est **transparent** sur ses pratiques ("transparency")
  - peut **prouver** ses pratiques ("accountability")
- des tiers de confiance sont nécessaires
  - peuvent **vérifier** les pratiques



Une utopie ?

- pas sûr !

- en économie les marchés avec une forte asymétrie d'information sont connus pour être fragiles
  - ça ne peut pas marcher longtemps

- … il en va de l'intérêt de tous

Merci… ☺

vincent.roca@inria.fr