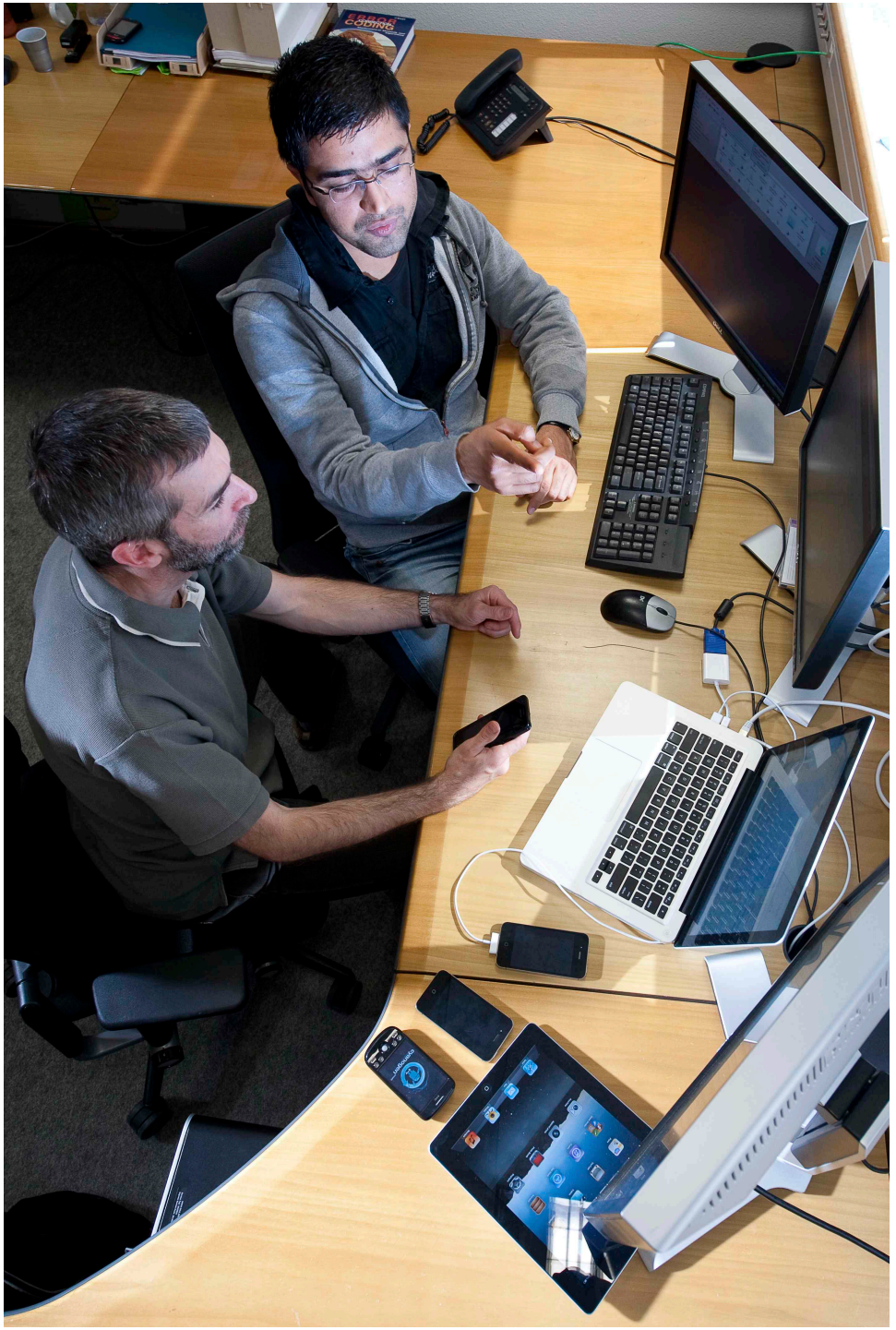
# Vie privée et smartphones : le projet Mobilitics Inria/CNIL

**Privatics team (Vincent Roca)**

Séminaire GUILDE

4 novembre 2014



informatiques mathématiques

# *Inria – équipe Privatics*

# Résumé

Un smartphone « voit passer » tout naturellement un grand nombre d'informations personnelles lorsque son utilisateur s'en sert pour communiquer, parcourir Internet et utiliser toutes sortes applications. Ce smartphone est également équipé d'un grand nombre de capteurs (GPS, gyroscope, accéléromètre, compas, détecteur de proximité, de mouvement, de lumière, mesure de température et d'humidité ambiante, baromètre, lecteur d'empreintes digitales, cardio-fréquencemètre, podomètre), sans compter les caméras et micros. Les données issues de ces capteurs vont s'ajouter à la masse d'informations personnelles que le smartphone « voit passer ». Et comme ce smartphone est de façon quasi permanente connecté à Internet et rarement éteint, il peut potentiellement révéler beaucoup d'informations sur son utilisateur. Enfin la multiplication des objets connectés tels que montre intelligente (souvent capable de mesurer des paramètres biologiques) ou de lunettes connectées (capables potentiellement de « voir » ce que l'utilisateur voit, et non ce que l'utilisateur lui « donne à voir »), ainsi que la croissance du domaine de l'auto-mesure (ou « Quantified Self ») pour un suivi personnel de son activité physique, voir de sa santé, ne font qu'accroître ce volume de données que le smartphone « voit passer ». Dans ce contexte, étant donné l'importance des intérêts économiques en jeux, il serait illusoire de penser que ces données restent sagement dans son smartphone, tout en étant éventuellement envoyées de façon anonyme avec le serveur de la société qui développe telle application ou tel objet connecté pour rendre le service attendu… La réalité est toute autre, sans même avoir besoin de parler d'espionnage étatique omniprésent « à la NSA » !

Cet exposé introduit le projet Mobilitics, mené conjointement par Inria/Privatics et la CNIL depuis 2012 et certains de ses résultats. Il en ressort que ces téléphones intelligents sont souvent de véritables mouchards de poche, d'autant plus dangereux que les informations personnelles ainsi captées sont la plupart du temps exfiltrées vers des serveurs à l'étranger, la législation Française (voire Européenne) devenant difficile à appliquer, pour être stockées, manipulées et échangées avec d'autres acteurs d'une façon totalement inconnue. Il en ressort un besoin de transparence auprès des citoyens et des autorités de régulation qui est pour le moment loin d'être acquis.

**contact : vincent.roca@inria.fr**

# Le projet Inria-CNIL Mobilitics

- démarré en janvier 2012


- s'intéresse à Android et iOS
  - ce sont les principaux OS, tout simplement

- analyse les fuites d'informations personnelles par les **Apps** et **les services de l'OS**
  - comparer Android/iOS
  - identifier les pratiques et tendances
  - études en labo et études « in vivo »

- **Deux exemples pour commencer…**

# *Exemple 1 : données de positionnement d'un opérateur téléphonique (2009)*

- Malte Spitz (Parti Vert allemand) a demandé à son opérateur les données le concernant
  - **enrichi avec des données publiques (ex. twitter)**
  - **mis en forme avec une application dédiée permettant de naviguer dans l'historique**
    - http://www.zeit.de/datenschutz/malte-spitz-data-retention/

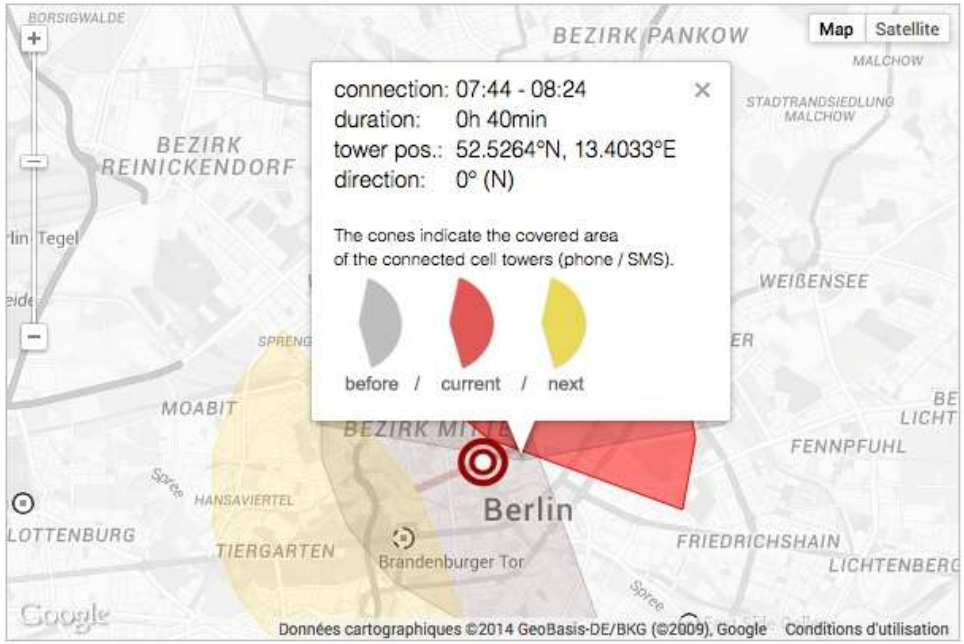
# Exemple 1… (suite)

## Tell-all telephone

deutsch | **english**

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.



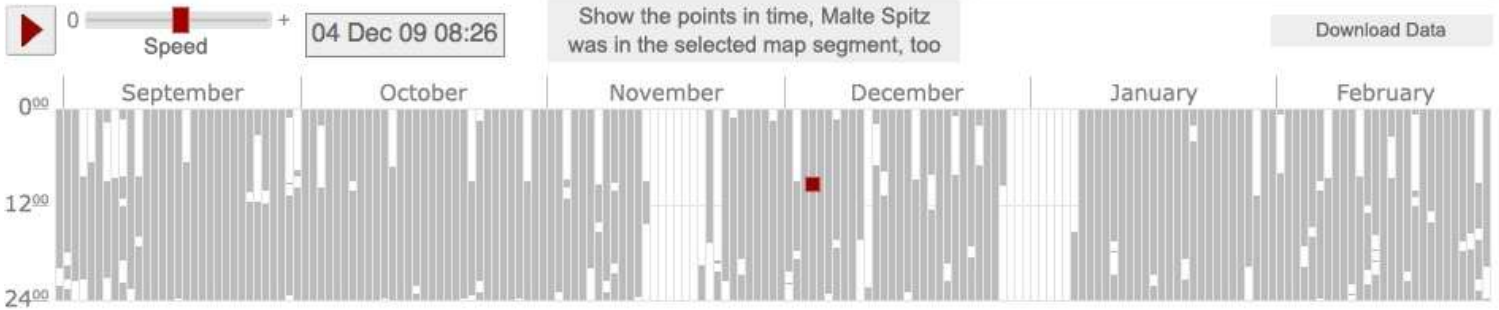

**Friday, 4 December 2009**

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




Data Protection: Betrayed by our own data

Implementation: OpenDataCity © ZEIT ONLINE

# *Exemple 1… (suite)*

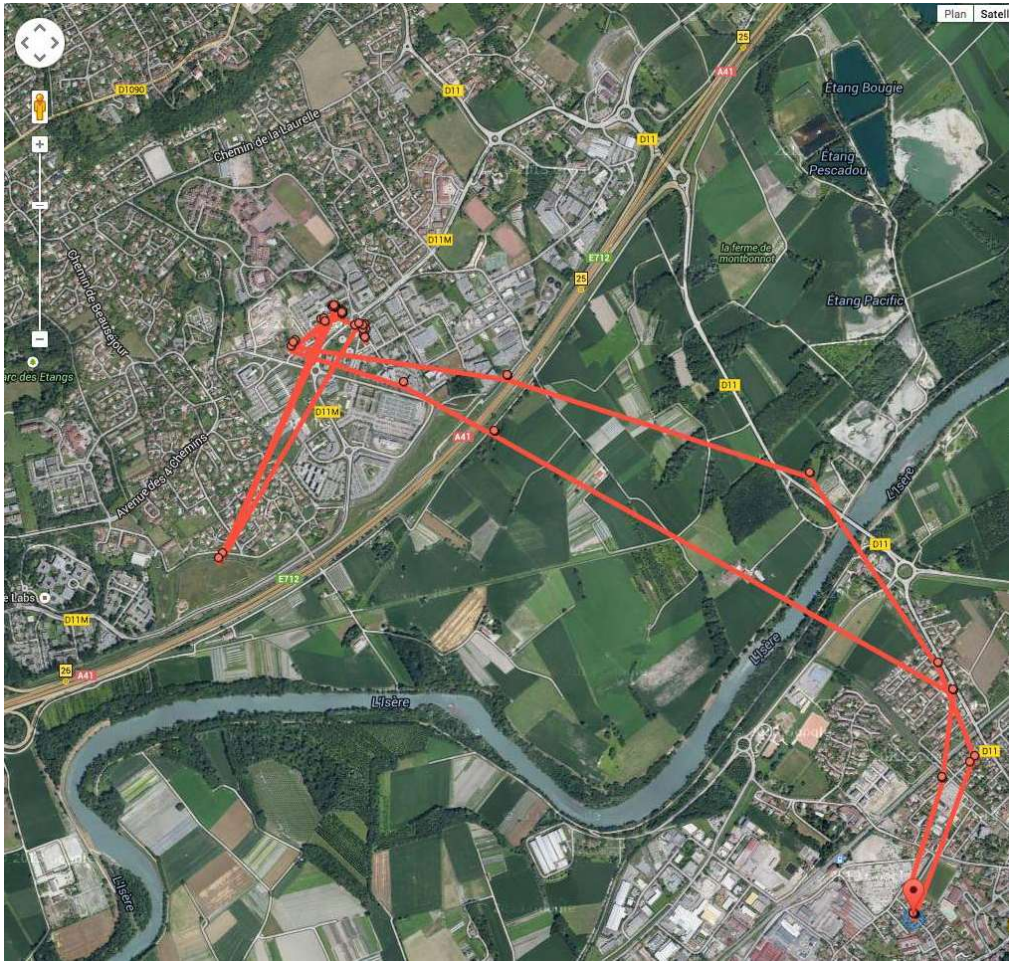
- okay, mais…
  - l'opérateur téléphonique a des obligations légales
  - données accessibles uniquement dans des conditions particulières, après requête officielle

# *Exemple 2 : données de positionnement version Google*

- historique des positions enregistrées par mon smartphone Android pour les services Google
  - on y a accès
    - nb: se connecter avec le compte gmail utilisé sur le smartphone Android

    https://maps.google.com/locationhistory/
  - ça vaut le coup d'aller y jeter un coup d'œil !
    - cliquer sur « afficher tous les points » pour tout avoir…

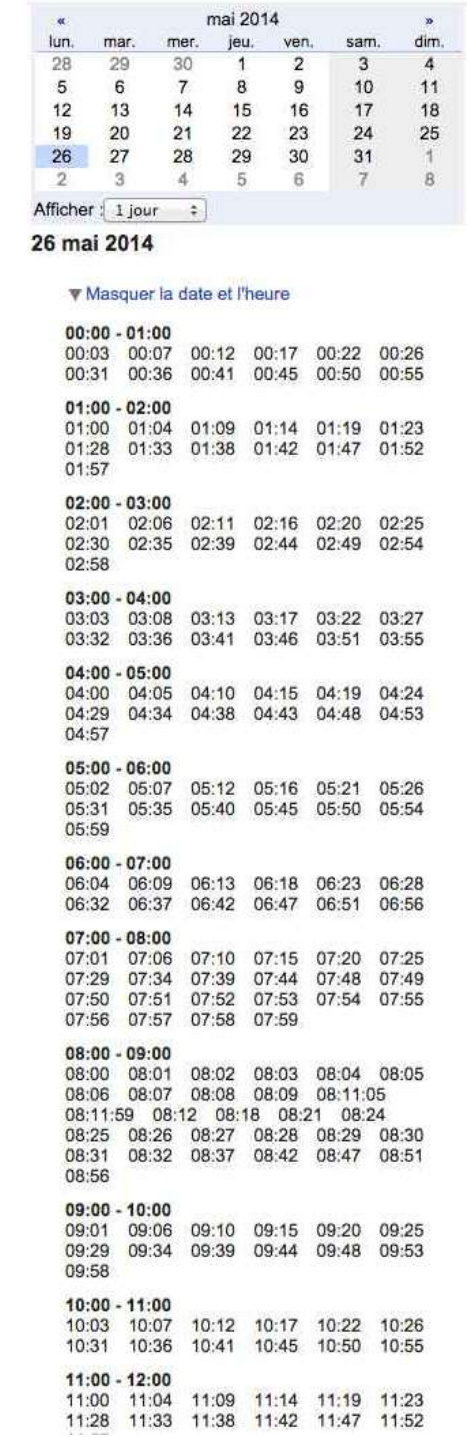
# Est-ce bien raisonnable ?




- 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… (suite)

- … 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 !**
  - ces données resteront enregistrées ainsi longtemps…



« 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

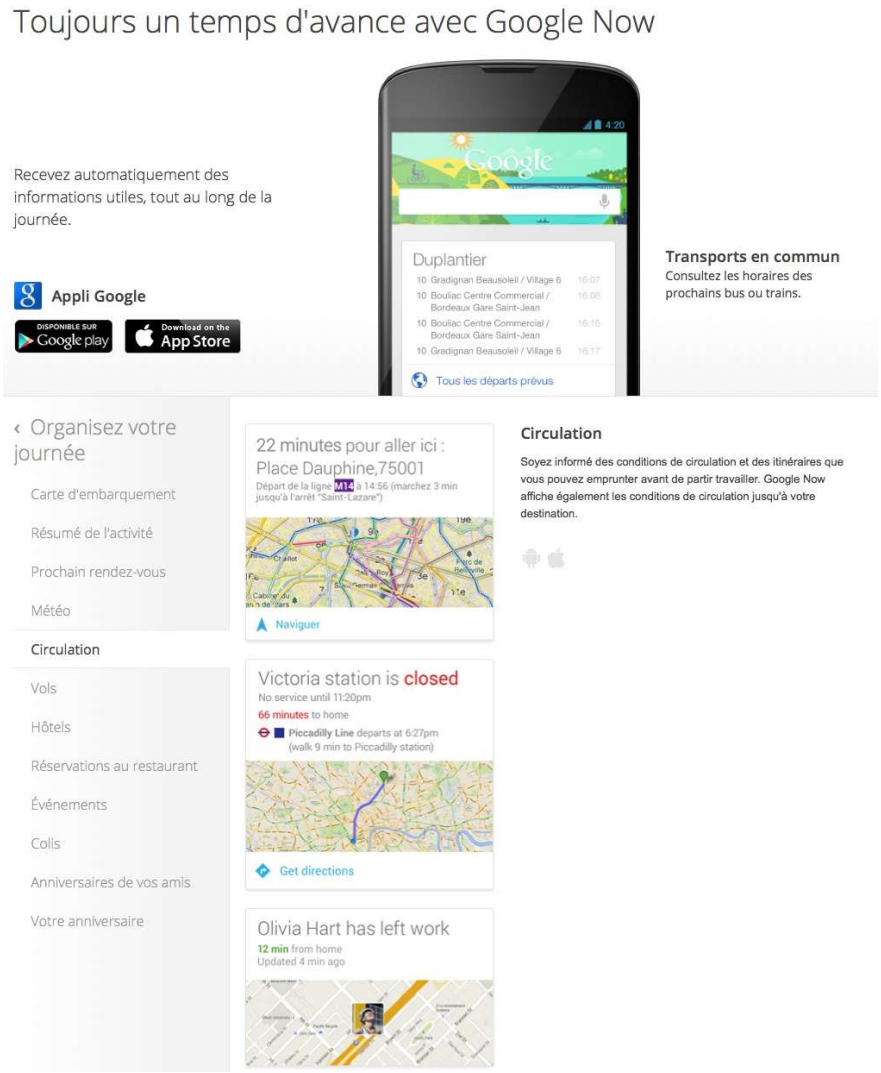
**09:00 - 10:00**
09:01 09:06 09:10 09:15 09:20 09:25
09:29 09:34 09:39 09:44 09:48 09:53
09:58

**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

# Est-ce bien raisonnable… (suite)

- comment suis-je arrivé là ?
  - j'ai activé Google Now : http://www.google.com/landing/now/… Argh !



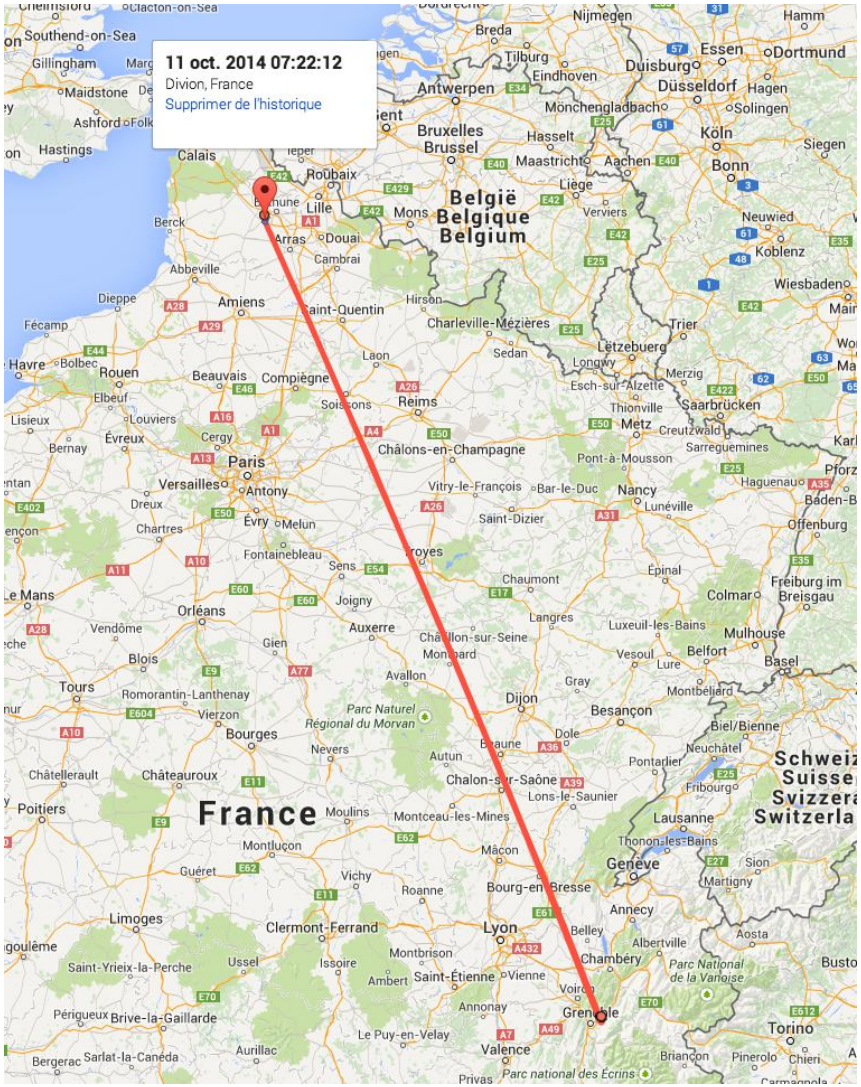
# *Est-ce bien raisonnable… (suite)*

- certes…
  - on peut désactiver Google Now (par défaut OFF)
  - on peut purger tout ou partie des données sur la page « historique des positions » de Google

- mais…
  - n'est-ce pas **disproportionné** par rapport au service rendu ?
    - **contraire au principe de base : « collecter le minimum de données nécessaires à la fourniture d'un service »**
    - **y a t-il besoin de conserver les données ?**
  - il y a aussi des **erreurs** de positionnement, parfois…

# Est-ce bien raisonnable… (suite)

- à Grenoble à 7h20, dans le nord à 7h22
  - **ici l'erreur est caricaturale, mais ce pourrait être crédible !**



- **Pourquoi en est-on arrivé là ?**

# *Une surveillance massive du monde*

- **un enregistrement systématique** des utilisateurs des outils de communication
  - nous laissons des traces à chaque fois que nous allons sur Internet ou utilisons notre smartphone…
    - **sur le web "visible"**
    - **sur le web "invisible"**
  - pour des raisons **économiques** ou **sécuritaires**

# *Surveillance sur le web "visible"*

- Foursquare sait **où vous êtes**
- Flickr sait **ce que vous voyez**
- Facebook sait **ce que vous faites**
- Linkedin sait **sur quoi/avec qui vous travaillez**
- Twitter sait **ce que vous dites**
- Amazon sait **ce que vous achetez**
- Google sait **ce que vous pensez**
- …

Si l'on croise ces infos, cela devient terrifiant…

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

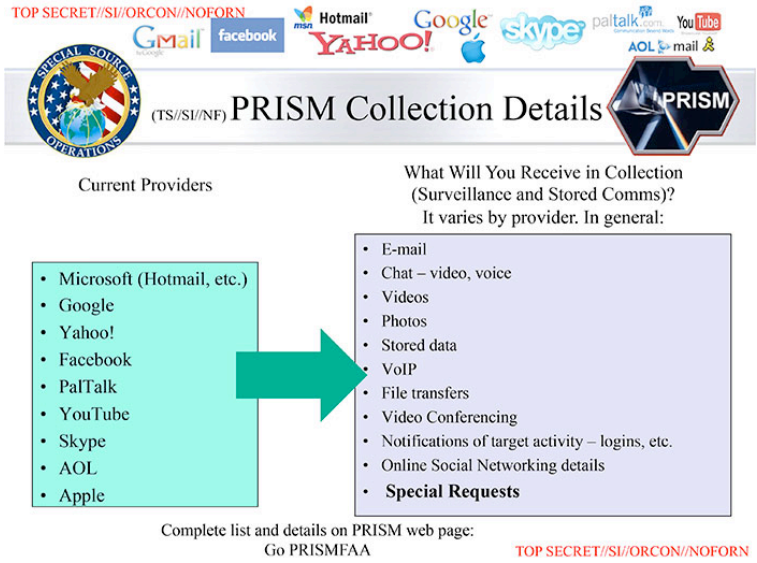
*Courtesy of F. Bancihlon

# Surveillance sur le web “invisible”

- grâce aux cookies, pixels, boutons “j’aime”, etc. des sites web
  - permet de **tracer** et **profiler** des utilisateurs

# Une situation qui peut conduire à des abus

- NSA…
  - le problème n'est pas de surveiller des cibles bien identifiées (activité « légitime »), mais :
    - **de surveiller tous les citoyens au niveau mondial**
    - **de compromettre la sécurité des outils de base**


- et ils ne sont pas les seuls

- **Revenons en aux smartphones…**

# *Les smartphones ont une responsabilité majeure*

- ils sont nos “compagnons” de tous les jours
  - pratiques, sympas, toujours connectés, faciles à personnaliser
- ils en savent beaucoup sur nos cyber-activités
  - ils **concentrent** des informations personnelles
    - **lorsqu’on les utilise (téléphone, sms, web, etc.)**
  - ils **génèrent** des informations personnelles
    - **GPS, NFC, WiFi, caméra, capteurs d’empreintes, capteurs de pression cardiaque, etc.**
  - les applications créent beaucoup d’opportunités pour faire fuiter des infos personnelles
    - **explique que des sites web vous incitent à installer leur App**

# *Les smartphones ont une responsabilité majeure… (suite)*

- notre smartphone devient notre **mouchard de poche préféré**

- il est constitué :
  - du système d'exploitation (OS)
    - **Android (Google), iOS (Apple), WindowsPhone, BBOS (BlackBerry), FirefoxOS, etc.**
  - d'applications (ou « Apps »)
  - d'un système complet (processeur + OS) de gestion des communications bas niveau (radio) totalement caché

notre sujet (Android/ iOS)

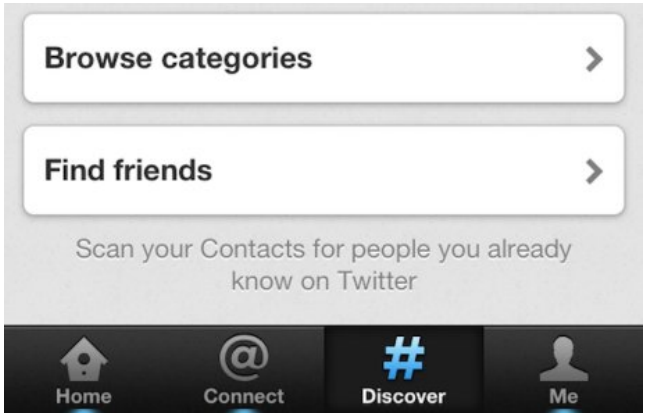
très difficile à étu-dier

# Notre mouchard de poche…

- situation **complexe** qui implique plusieurs acteurs
  - « first party » : **propriétaire de l'App**
    **⇒ celui qu'on voit**
  - « third party » : **Advertising and Accounting (A&A)**
    **⇒ celui que l'on ne voit jamais**
  - **le third party a lui même des clients (ex. régies publicitaires)**
  - **certains acteurs jouent plusieurs rôles (ex. Google, Facebook)**
  - il devient difficile de faire confiance à tout le monde !
- voyons deux exemples…

# Exemple 1 : fuites d'infos par "maladresse"

- Twitter (février 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."**




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

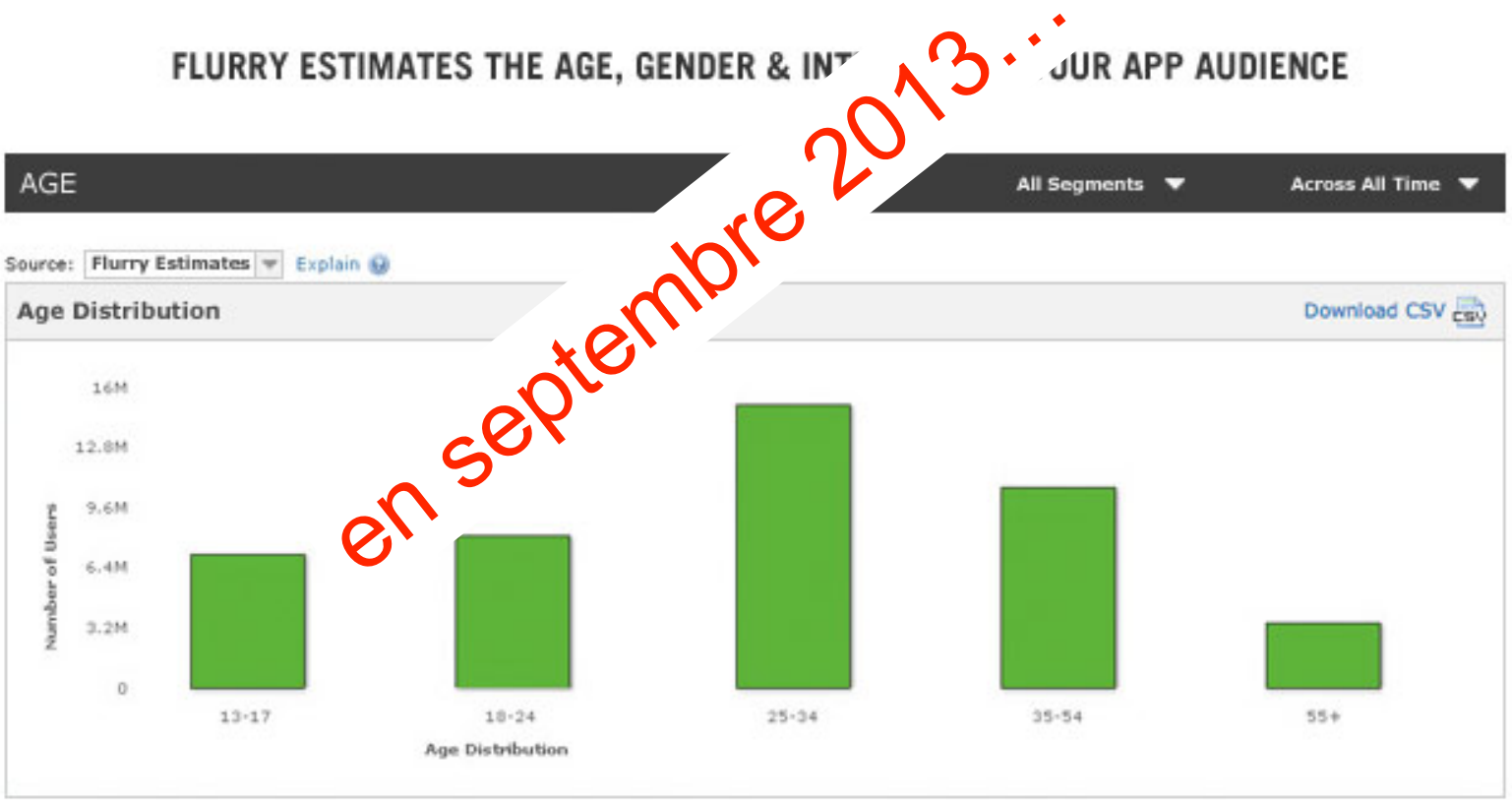
  - **des scandales similaires avec with LinkedIn et Path en 2012!**

- ce sont des erreurs stratégiques
  - **une société renommée a peu à gagner alors qu'elle risque très gros en termes d'images**
  - **rapidement corrigé dans la nouvelle version de l'App**

# Exemple 2 : collecte massive et organisée

- 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:




en septembre 2013…

# Exemple 2 : collecte massive… (suite)

- pourquoi faire ?
  - pour **tracer les utilisateurs**
    - **un même utilisateur revient il ? Utilise t-il plusieurs Apps ? Lesquelles ? A quelle fréquence ? Quand ?**
  - pour **profiler les utilisateurs**
    - **est-ce un homme d'âge moyen, technophile, accro au sport de canapé, lecteur de news, etc.**
  - afin de vendre des **publicités ciblées** sur le smartphone
    - **ciblée donc qui a des chances de faire mouche**
  - afin de se créer des **bases de données** qui peuvent être valorisées par ailleurs (ex. à venir)

- **Qui fait et gagne quoi dans cette histoire ?**

# Les multiples acteurs

**développeur (first party)**

**Advertising & Analytics (third party)**

inclue une lib. A&A dans l'App

$

agrège et construit des profils utilisateurs

développe et gère une App

**App Store**

installe l'App

**utilisateur**

l'App envoie des données à la société A&A

pub

intéressé par un utilisateur « jeune & mode »?

pub + $$$

adidas

# A propos de la publicité pour mobiles

- beaucoup de sociétés existent



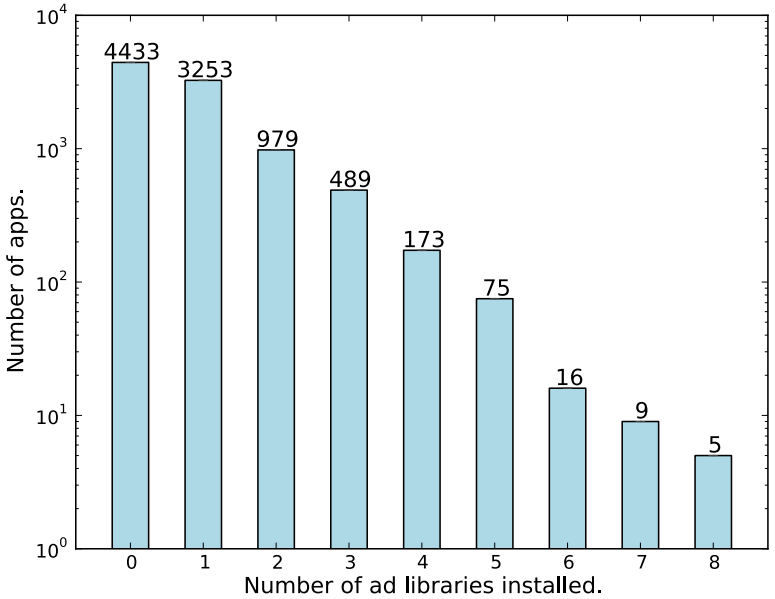



**et bien d'autres…**

- Google a dépassé **8 milliards de $** de recette pour la publicité mobile en 2013 !

# A propos de la publicité mobile… (suite)

- quelques faits
  - "77% des 50 Apps gratuites pour Android étaient soutenues par de la publicité" on July 2011 [1]
  - 35% des Apps gratuites pour Android utilisent 2 ou plus bibliothèques publicitaires [2]
    - **un moyen d'accroitre ses revenus**





- **[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.**

# A propos de la publicité mobile… (suite)

- cela impacte vraiment le comportement de l'App
  - les bibliothèques A&A demandent des autorisations supplémentaires
    - **une App gratuite demande 2-3 permissions supplémentaires par rapport à une App payante équivalente [1]**

| Ad Library | Internet | NetworkState | ReadPhoneState | WriteExternalStorage | CoarseLocation | CallPhone |
|---|---|---|---|---|---|---|
| AdMob [22] | ✓ | ✓ | | | ○ | |
| Greystripe [25] | ✓ | ✓ | ✓ | | | |
| Millennial Media [36] | ✓ | ✓ | ✓ | ✓ | | |
| InMobi [29] | ✓ | ○ | | | ○ | ○ |
| MobClix [38] | ✓ | ○ | ✓ | | | |
| TapJoy [53] | ✓ | ✓ | ✓ | ✓ | | |
| JumpTap [32] | ✓ | ✓ | ✓ | | ○ | |

✓(required), ○ (optional)

*permissions per Ad lib [2]*

- **Là où cela pose vraiment problème…**

# *Le fond du problème*

- un modèle économique comme un autre ?
  - « si c'est gratuit, c'est vous le produit »
    - **le prix à payer pour des Apps gratuites ? Pourquoi pas…**

- mais
  - la collecte est **massive** et parfois **disproportionnée** au regard du service rendu
  - les données sont immédiatement **exfiltrées** hors du territoire national, pour être stockées, manipulées, échangées dans des conditions inconnues
    - **les lois de protection des données française et européennes sont difficiles à appliquer dans ces conditions**
  - enfin il y a un **impératif stratégique**
    - **« les données sont le pétrole de demain »**

- **Les approches « constructeurs » pour le contrôle des paramètres de vie privée**

# Des approches complémentaires

- divers modèles de contrôle du comportement des Apps vis à vis des données personnelles
  - **centrée marché :** vérification de l'App avant son acceptation sur le marché par le propriétaire du marché



  - **centrée utilisateur :** demander le consentement utilisateur
    - **soit à l'installation de l'App**

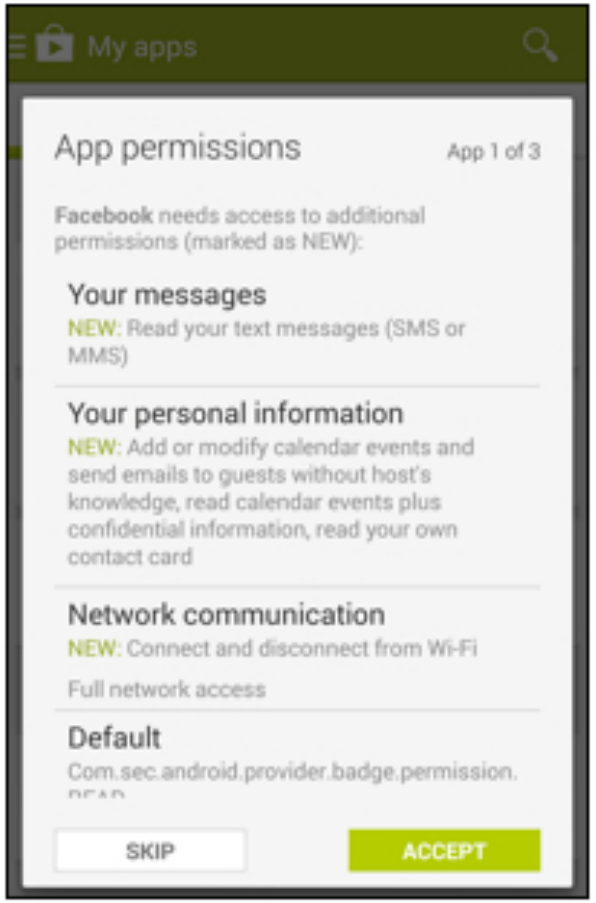

    - **soit dynamiquement, à l'usage de l'App**

# A propos des autorisation à l'installation

*Rappel : stratégie Google/Android*

- une App ayant des besoins particuliers demande le consentement utilisateur avant de procéder à l'installation
  - transfert la **responsabilité**

    à l'utilisateur



# A propos des autorisations dynamiques

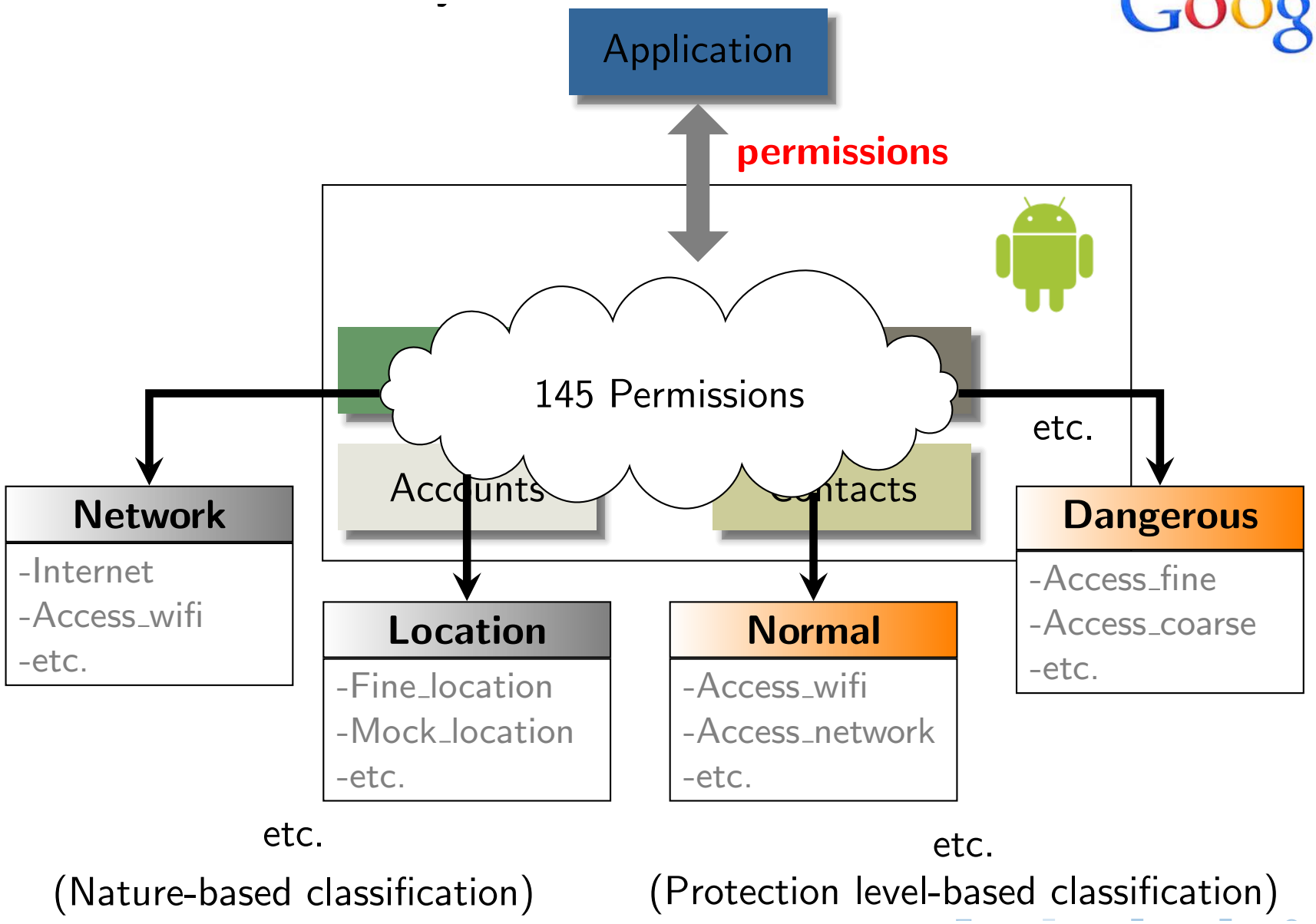
## *Rappel : stratégie Apple/iOS*

*(bizarrement aussi dans Android 4.3, puis retiré)*

- un panneau de contrôle vie privée permet à l'utilisateur d'autoriser/ interdire l'accès à des informations personnelles pour des applications
  - l'utilisateur a le **contrôle**
  - présent depuis iOS 6
    - **amélioré progressivement**

- **Les limites de ces approches sont multiples**

# Les autorisations Android sont un tantinet complexe…



# *Les autorisations sont complexes… (suite)*

- un utilisateur n'est pas toujours en mesure de comprendre les **implications**
  - **exemple : ACCESS_WIFI_STATE**
    - permet d'inférer un grand nombre de données personnelles (localisation, historique de déplacement, liens sociaux, identifiant stable pour le traçage) sans que ce soit évident
  - **(cf. plus loin)**

- ne permet pas un contrôle sur la **composition** des autorisations
  - **autoriser l'App à accéder à mes contacts et à Internet ne signifie pas que je l'autorise à transmettre mes contacts à des serveurs distants**

# *…tout en étant trop limitées*

- à prendre ou à laisser pour installer une App
  - **on ne vit pas dans un monde binaire !**
- aucun contrôle **comportemental** de l'App
  - **autoriser une App à accéder à ma localisation et à Internet pour rendre un service ponctuel ne signifie pas que je l'autorise à accéder à ma localisation toute les minutes et à envoyer cette localisation sur des serveurs distants**

# *Quid du système Apple ?*

- c'est mieux, mais pas encore suffisant

- aucun contrôle **comportemental** de l'App
  - **idem**
  - **autoriser un accès à une information personnelle ne signifie pas que j'autorise n'importe quelle modalité d'accès et de traitement de cette information !**

## trois exemples rapides

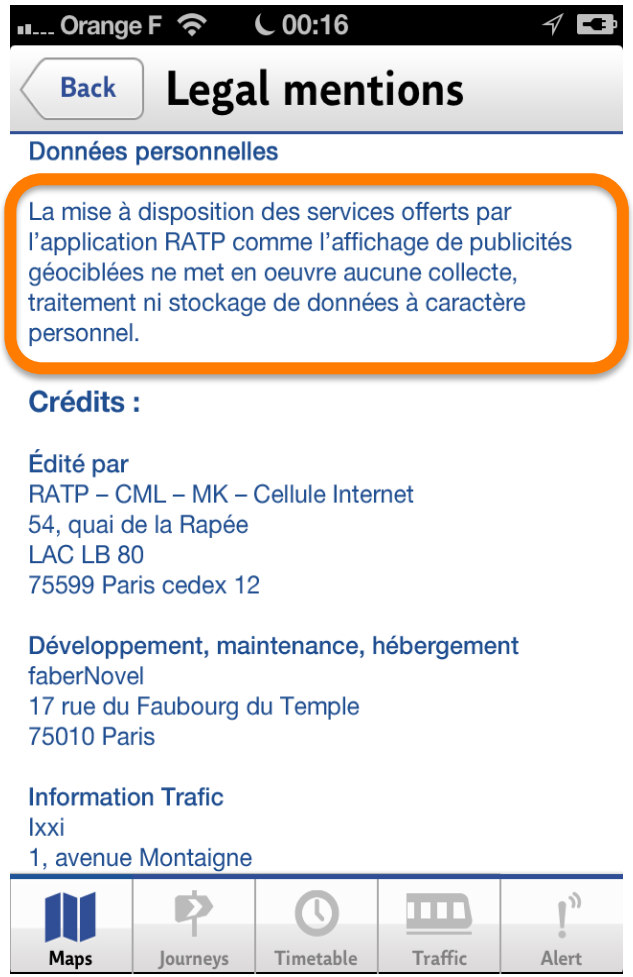
- l'application **RATP**, version de mi 2013
- J. P. Achara, M. Cunche, V. Roca, A. Francillon, **« Short paper: WifiLeaks: Underestimated Privacy Implications of the ACCESS WIFI STATE Android Permission »**, IEEE WiSec'14.
- **à propos du traçage des utilisateurs grâce à leur smartphone / Wifi,** présentation de M. Cunche (journées du LERTI)

# Un cas d'école : l'App RATP version 5.4.1

- « Y'a pas de problèmes » dixit la RATP

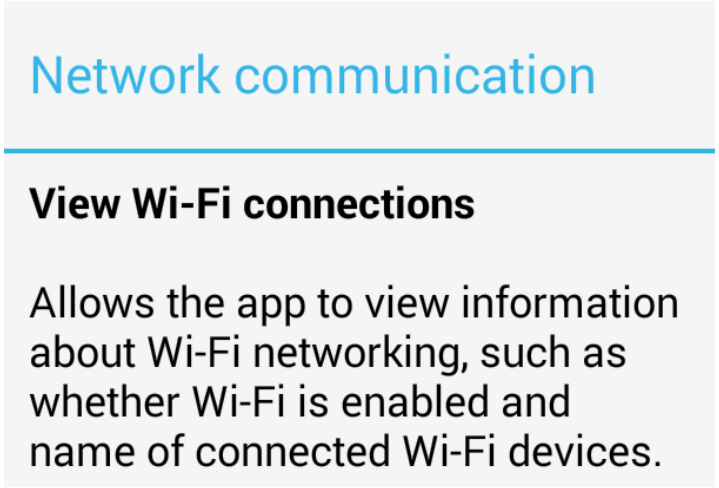
- Vraiment ?
  - la liste des Apps actives, mon adresse MAC, le nom de mon téléphone, ma position géographique précise (à 20m près), un identifiant permanent sont envoyés à Adgoji (chiffré) et sofialys (en clair !)

- Voir notre blog : part-1 et part-2: https://team.inria.fr/privatics/



# Ex. 2 : à propos de ACCESS_WIFI_STATE

- une autorisation **Android**, de type réseau / normal
  - **« accéder à l'état du composant Wifi »**
  - **d'apparence anodine…**




  - **… alors qu'elle permet d'inférer**
    - la **localisation** par triangulation de réseaux Wifi à portée
    - l'historique des réseaux Wifi où l'on s'est connectés (ex. MacDO, wifi d'un hôtel, d'un aéroport, etc.) et donc une partie de l'**historique de déplacements**
    - des **liens sociaux** entre utilisateurs en comparant les historiques de connexion Wifi des utilisateurs
  - **fournit un identifiant stable utile pour tracer l'utilisateur**

# Ex. 3 : à propos du traçage des utilisateurs dans le monde physique

transparent M. Cunche (Inria, Privatics)

- 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


[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.

# Ex. 3 : traçage des utilisateurs... (suite)

- 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

transparent M. Cunche (Inria, Privatics)

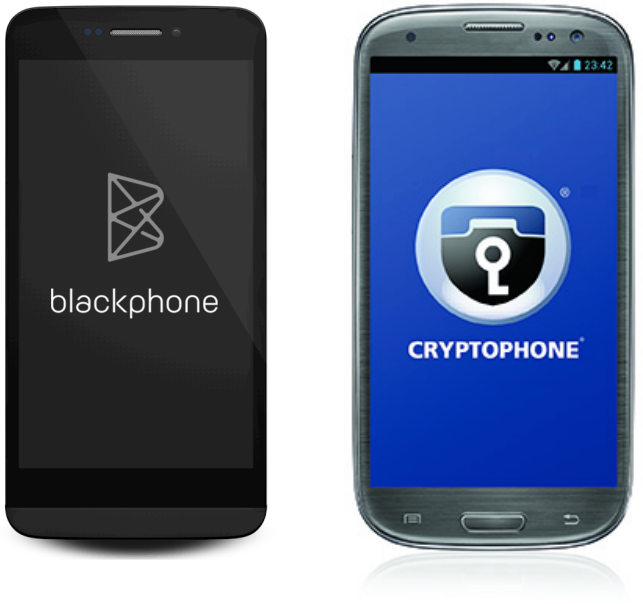

# Conclusions

# *Le cas de Google*

- **Google vend de la publicité**… et à besoin de données précises sur ses utilisateurs
  - **les Apps ont à disposition tous les identifiants stables nécessaires pour tracer les utilisateurs**
    - souvent sans avoir la moindre autorisation à demander !
  - **très peu de publicité pour améliorer la situation**
    - depuis août 2014, les nouvelles Apps sont censées utiliser uniquement « l'Advertising ID » pour tracer à des fins pubs…
    - … mais cela prendra du temps (fragmentation du marché Android) et Google ne communique pas dessus (!)
    - en attendant une App collecte tous les identifiants possibles
  - **il y a même eu des indicateurs contraires**
    - Android 4.3 intégrait un panneau de contrôle « vie privée » qui a été retiré ensuite

# Le cas de Google… (cont')

- mais c'est un OS en partie open-source
  - possible de construire des versions sécurisées…
    - **BlackPhone (de Silent Circle)** approx. 600 $
      - https://silentcircle.com/services#blackphone
    - **CryptoPhone 500 (de GSMK)** approx. 3500 $
      - http://www.cryptophone.de/en/products/mobile/cp500/
      - capable d'identifier les fausses antennes relais
        - http://www.popsci.com/article/technology/mysterious-phony-cell-towers-could-be-intercepting-your-calls



# *Le cas d'Apple*

- **Apple vend (cher) du matériel et du logiciel**… et communique sur le respect de la vie privée

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

❍ **on peut apporter des bémols**

- le mécanisme d'adresses MAC Wifi aléatoires est peu utile en l'état

❍ **mais la situation d'iOS s'améliore**

- beaucoup d'identifiants stables sont désormais inaccessibles dans iOS 8
- le mécanisme d'identifiant publicitaire ré-initialisable par l'utilisateur devient incontournable

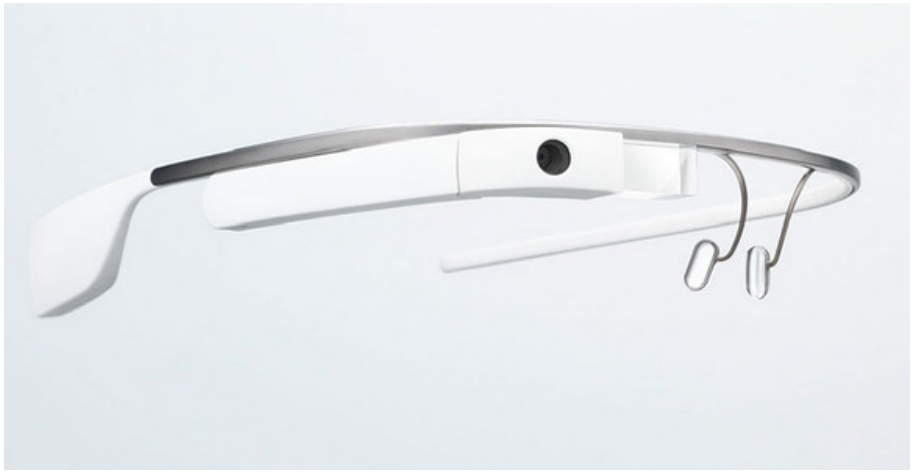
# *L'utilisateur peut aussi limiter la casse*

- limiter les Apps installées
  - être vigilent vis à vis des autorisations à l'installation (Android) ou via le panneau de contrôle (iOS)
  - … désinstaller celles qui ne sont pas utilisées
  - … et peut être éviter Google Now !
- utiliser les magasins d'App officiels
  - bénéficie du filtrage opéré par le propriétaire du magasin
- éteindre le Wifi dès que possible…
  - pour éviter le traçage physique (magasins)
- …et couper les connexions données si possible
  - limite certaines communications indésirables

# *Et le législateur a un réel pouvoir*

- la législation Européenne continue d'évoluer dans le bon sens
  - **avec des moyens coercitifs sur les sociétés (réputation et/ou financier)**
  - **coordination des agences de protection des données européennes (ex. CNIL en France) au sein du G29**

- mais il faut des outils pour des contrôles indépendants
  - les solutions Mobilitics Inria-CNIL en font partie



# Merci… ☺