Smartphone, Wi-Fi et vie privée : comment votre smartphone peut se révéler être votre pire ennemi

Mathieu Cunche

Smartphone, Wi-Fi et vie privée : *comment votre smartphone peut se révéler être votre pire ennemi.*

Mathieu Cunche
INSA-Lyon, CITI Lab., Inria, Privatics team
mathieu.cunche@inria.fr

Abstract

Nos smartphones et nos tablettes sont des objets qui nous accompagnent partout et qui sont pour la plupart équipés d'une interface Wi-Fi. Certaines spécificités de cette technologie font que ces compagnons de vie numériques se comportent comme de véritables mouchards en révélant des informations personnelles à qui veut bien tendre l'oreille (ou plutôt l'antenne). Nous faisons ici le point sur ces fuites de données et les dangers qu'elles représentent pour notre vie privée.

1 La technologie Wi-Fi

Le Wi-Fi, apparu en 1999, permet à nos appareils informatiques de communiquer entre eux par ondes radio. Au départ limité aux ordinateurs de bureau, le Wi-Fi est aujourd'hui intégré à tout type d'équipement et en particulier auxéquipements mobiles tels que les smartphones, les tablettes et les ordinateurs portables.

Du fait de l'utilisation d'un médium ouvert et partagé, les communications sans filsont susceptibles d'être l'objet d'écoutes illégitimes. La confidentialité des données et le contrôle d'accès au réseau sont deux éléments centraux de la sécurité du Wi-Fi. Depuis ses origines, cette technologie a été mise à rude épreuve par des attaques qui ont permis l'évolution des mécanismes de durcissement tels que l'abandon du WEP au profit de WPA et WPA2[1], si bien

Ce document est la version auteur de l'article *Smartphone, Wi-Fi et vie privée : comment votre smartphone peut se révéler être votre pire ennemi* paru dans Multi-system & Internet Security Cookbook (MISC), Editions Diamond, 2013, Apprenez à protéger votre vie privée (Hors-série), 8 `http://www.unixgarden.com/index.php/misc-hs/misc-hors-serie-n8-octobrenovembre-2013-en-kiosque`

[1] `http://fr.wikipedia.org/wiki/Wi-Fi_Protected_Access`

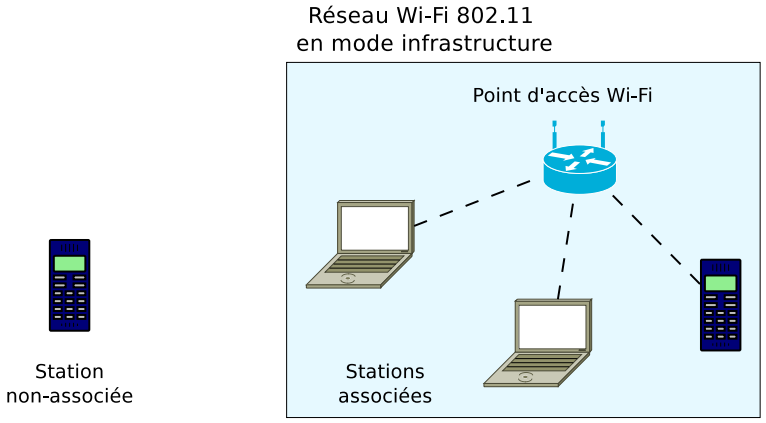

Figure 1: Exemple d'environnement Wi-Fi composé d'un point accès et de plusieurs stations.

qu'aujourd'hui un réseau Wi-Fi avec un système de protection à jour et correctement configuré garantissent un contrôle d'accès solide ainsi que la confidentialité des données qui transitent sur le réseau.

Malheureusement, en plus des problèmes de sécurité, la technologie Wi-Fi expose ses utilisateurs à des problèmes de vie privée. En effet un certain nombre d'informations à caractère personnel circulent en clair sur les ondes radio et cela même si le terminal n'est pas connecté à un réseau. Parmi les problèmes que nous aborderons dans cet article, nous pouvons citer : la diffusion de l'historique de connexion d'un terminal et la diffusion d'un identifiant unique permettant le traçage des déplacements des individus. Ces problèmes de vie privée sont exacerbés par la généralisation de la technologie Wi-Fi.

1.1 Le standard 802.11

La technologie Wi-Fi repose sur la famille de standards 802.11 qui spécifie la couche MAC (*Medium Access Control*) et la couche physique pour implémenter des réseaux locaux sans fils. Le protocole 802.11 se décline sous plusieurs variantes parmi lesquelles celles correspondant aux Wi-Fi sont le 802.11a, 802.11b, 802.11g et le 802.11n. Ces standards travaillent sur les bandes de fréquence de 2.4 GHz et 5 GHz qui sont décomposé en plusieurs canaux (de 11 à 13 en fonction des régions). Un réseau 802.11 peut avoir différentes topologies. La topologie de réseau la plus courante est celle du mode infrastructure dans laquelle un ou plusieurs points d'accès constituent une base auxquels viennent se connecter des stations. Un paysage Wi-Fi se compose alors de plusieurs points d'accès et de stations qui peuvent être connectés (on dit aussi associé) ou non à un point d'accès (voir Figure 1).

La couche MAC du protocole 802.11 correspond au niveau 2 de la pile OSI et ses datagrammes sont appelés des *trames* (voir Figure 2). Il existe 3 types

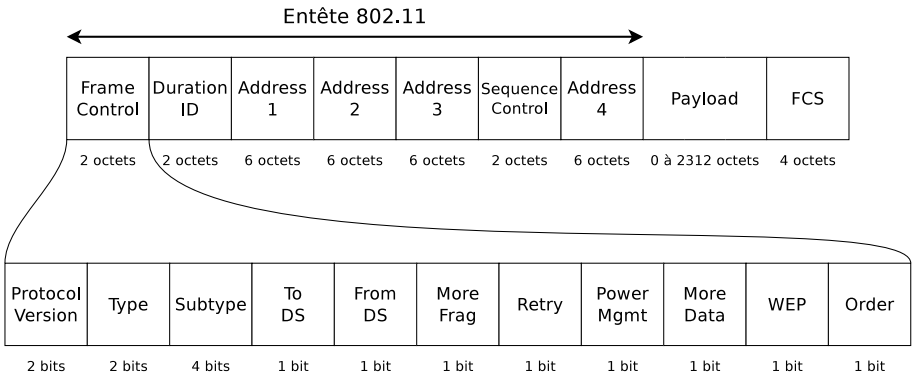

Figure 2: Décomposition d'une trame 802.11 et du champ *Frame Control*.

de trames : les trames *Data* qui contiennent les données utiles, les trames *Management* qui sont utilisées pour l'établissement et le maintien des connexions, et les trames *Control* qui permettent les acquittements des transmissions. Chaque trame est composée d'un entête de 30 octets, d'un corps (*payload*) de longueur comprise entre 0 et 2312 octets et se termine par une séquence de contrôle (FCS pour *Frame Control Sequence*) de 4 octets. L'entête d'une trame contient les informations nécessaires à son interprétation et à son acheminement, le corps de la trame contient les données correspondant aux datagrammes des protocoles des couches supérieures de la pile protocolaire (par exemple des paquets TCP/IP), et le champ FCS contient un CRC32 pour le contrôle d'intégrité.

L'entête d'une trame 802.11 est composé d'un champ de contrôle de trame (FC pour *Frame Control*) sur 2 octets, d'un champ Durée/Identifiant sur 2 octets, de quatre champs "adresse" de 48 bits, et d'un champ de contrôle de séquence. Le champ de contrôle de trame contient un ensemble de sous champs

utiles au fonctionnement des communications. Parmi ces champs, on peut noter les champs suivant dont nous reparlerons par la suite :

- **Type** et **Subtype** : Ces champs identifient le type de la trame (Data, Management ou Control) et son sous-type (*beacon*, *probe request*, *association request*, etc).
- **Power Management** : Ce bit indique l'état du terminal (en veille ou actif).
- **WEP** : Ce bit indique si le corps de la trame est chiffré.

Le rôle des champs "adresse" peutvarier en fonction du type de la trame. En règle générale, la 1ère adresse correspond au destinataire, la seconde à la source (on parle aussi d'émetteur) et la troisième correspond au BSSID qui identifie le point d'accès et donc le réseau auquel correspond cette trame. Enfin, le quatrième champ est optionnel et est uniquement utilisé dans des cas qui ne nous intéressent pas ici.

2 Des appareils très bavards

Les terminaux équipés d'une interface Wi-Fi émettent des messages même lorsqu'ils ne sont pas connectés à un réseau. Nous allons voir comment il est possible d'écouter ces messages et les raisons de ces transmissions intempestives.

2.1 Comment les écouter

Comme nous l'indiquions plus haut, la nature ouverte et partagée du médium fait que les communications Wi-Fi sont relativement faciles à intercepter. En temps normal, une interface Wi-Fi ne conserve que les trames qui lui sont adressées (trames pour lesquels le champ adresse destination correspond à sa propre adresse MAC) et ignore les autres. Pour observer les trames qui ne nous sont pas destinées il faut disposer d'une interface Wi-Fi supportant un mode de fonctionnement appelé *monitoring* qui permet d'observer l'ensemble des trames émises sur le canal.

L'activation de ce mode nécessite l'installation préalable de pilotes compatibles tel que le pilote rt2800usb utilisé pour les chipsets Ralink RT5370. Une fois cela effectué, nous pouvons utiliser les outils de la suite *aircrack-ng* [2], disponible pour la plupart des OS (Linux, Windows, MacOSX, OpenBSD), afin de configurer notre interface et intercepter le trafic. Cette suite est constituée d'un ensemble d'outils d'audit de sécurité des réseaux Wi-Fi qui permettent, entre autres, d'intercepter des communications et de tester la robustesse des mécanismes de sécurité (par exemple le cassage de clefs WEP et WPA).

La commande *airmon-ng* permet de passer une interface Wi-Fi en mode *monitoring*. Nous pouvons alors effectuer une capture en utilisant la commande

[2] `http://www.aircrack-ng.org`

airodump-ng ou en utilisant directement des outils d'analyse de trafic comme *tcpdump* ou *Wireshark*.

2.2 Un mot sur le chiffrement

Le standard 802.11 intègre un certain nombre de mécanismes de sécurité qui permettent l'authentification des stations et le chiffrement des données. Cependant, le chiffrement ne concerne que le corps des trames de type *Data*. L'entête et le corps des trames *Management* ne sont pas chiffrés.

Ainsi, malgré l'emploi de mécanismes de sécurité, une partie des données circulant sur un réseau Wi-Fi sont transmises en clair. Ce sont ces données qui vont nous permettre de collecter des informations personnelles sur les utilisateurs. Il est important de noter que l'accès à ces informations ne nécessite en aucun cas de casser un quelconque système de sécurité.

2.3 Adresse MAC : identifiant unique

Comme nous le présentions plus tôt, l'entête des trames Wi-Fi contient quatre champs "adresse" dont un identifie l'émetteur et un autre le destinataire de la trame. Ces adresses sont appelées des adresses MAC.

Une adresse MAC est composée de 48 bits et identifie de manière unique chaque interface Wi-Fi dans le monde. Les 24 premiers bits de l'adresse désignent le distributeur de l'interface. Ils permettent dans bien des cas d'identifier la marque du terminal correspondant (par exemple une adresse commençant par le préfixe `7C:C3:A1` désigne le distributeur *Apple Inc.*).

Une trame peut être adressée à toutes les interfaces à portée (on parle alors de *broadcast*). Dans ce cas, l'adresse destination est alors la valeur spéciale `ff:ff:ff:ff:ff:ff`. Cependant l'adresse MAC source correspond toujours à l'interface qui a émis la trame.

L'adresse MAC étant propre à une interface et donc à un appareil, elle permet d'identifier de manière unique un terminal tant que l'interface n'a pas été extraite du terminal et remplacée par une autre. Dans le cas de terminaux mobiles portés par des individus, cet identifiant peut être, par transitivité, utilisé pour identifier de manière unique le porteur du terminal tant que son propriétaire ne l'a pas revendu.

Etant présente en clair dans toutes les trames émises par le terminal, l'adresse MAC constitue donc un moyen idéal pour identifier et tracer le porteur de l'appareil.

Par ailleurs, une interface Wi-Fi ne se contente pas d'émettre des trames lorsqu'elle est utilisée pour échanger des données avec le réseau auquel elle serait connectée. En fait, un certain nombre de mécanismes du Wi-Fi que nous allons décrire l'amène à générer des trames en l'absence de trafic de données et même lorsqu'elle n'est pas connecté à un réseau.

2.4 Les mécanismes d'économie d'énergie

Le premier mécanisme causant l'émission de trame en l'absence de trafic trouve sa source dans l'économie d'énergie. En effet, les terminaux mobiles disposent d'une réserve d'énergie limitée. Afin d'économiser cette ressource, les interfaces Wi-Fi peuvent se désactiver temporairement. Une interface Wi-Fi peut alors se trouver dans deux états : *actif* ou *veille*. Pour ne pas perdre de messages qui lui seraient adressés lorsqu'elle est en veille, une interface avertira le point d'accès auquel elle est associée de chaque changement d'état. Le point d'accès se charge alors de mettre les paquets destinés à une station en veille dans une mémoire tampon et de les lui délivrer lorsque lorsqu'elle redeviendra active.

Pour avertir le point d'accès de son changement d'état, la station utilise un bit de l'entête des trames Wi-Fi : le drapeau *Power Management*. Ce drapeau peut être utilisé dans n'importe qu'elle trame. Cependant, en cas d'absence de données à transmettre, la station utilisera une trame de type *DATA* ne contenant aucune donnée. On parle alors de trame *NULL DATA*.

Une station émet donc au moins une trame à chaque changement d'état. La fréquence de ces changements d'état peut varier d'un terminal à l'autre mais les terminaux que nous avons observés changent d'état plusieurs fois par minute. Ainsi, lorsqu'un terminal est connecté à un réseau, il émet plusieurs fois par minute des trames contenant son adresse MAC, et cela même en l'absence de trafic.

2.5 Mécanisme de découverte de service

La seconde cause d'émission intempestive de trames par les appareils Wi-Fi provient des mécanismes de découverte de service. Ces derniers permettent à un appareil Wi-Fi de reconnaître son environnement en détectant les points d'accès qui sont à portée. Un terminal peut alors proposer à son utilisateur une liste de réseaux Wi-Fi auquel il peut se connecter.

La découverte de service est également utilisée par les stations déjà connectées à un point d'accès pour en trouver un autre appartenant au même réseau mais avec une meilleure qualité de réception. C'est une fonctionnalité particulièrement utile pour les terminaux mobiles car au gré des déplacements de leurs porteurs, ils doivent souvent migrer d'un point d'accès à un autre.

La découverte de service se décline en deux modes indépendants : un mode passif dans lequel les terminaux écoutent les trames de type *beacon* émises par les points d'accès alentour pour déclarer leur présence, et un mode actif dans lequel les terminaux Wi-Fi scannent leur environnement en envoyant des sollicitations aux points d'accès à portée.

Le mode passif : Le mode de découverte de service passif repose sur l'émission de trames de type *beacon* par les points d'accès. Elles appartiennent à la classe *Management* du protocole 802.11 et contiennent des informations sur la configuration du point d'accès : son canal, son identifiant unique (le BSSID), le nom du réseau auquel il appartient (le SSID), les modes de sécurité supportés, etc.

Chaque point d'accès émet périodiquement des trames *beacon* sur son canal. Ces émissions sont effectuées au moins plusieurs dizaines de fois par seconde. Une station souhaitant découvrir l'ensemble des points d'accès à sa portée doit écouter les trames beacon pendant une période pouvant aller jusqu'à plusieurs centaines de millisecondes. Pour avoir un panorama complet de son voisinage Wi-Fi, cette opération devra en plus être réitérée sur l'ensemble des canaux.

On peut utiliser *tcpdump* pour capturer et afficher les trames beacon (en rouge les BSSIDs et les SSIDs) :

```
$ sudo tcpdump -i mon0 -e 'type mgt subtype beacon' -vvv
tcpdump: listening on mon0, link-type IEEE802_11_RADIO (802.11 plus
     radiotap header), capture size 65535 bytes
17:20:26.526104 1.0 Mb/s 2457 MHz 11b -80dB signal antenna 7 0us
    BSSID:d2:17:33:d1:4e:a5 (oui Unknown) DA:Broadcast SA:d2:17:33:d1
    :4e:a5 (oui Unknown) Beacon (SFR WiFi FON) [1.0* 2.0* 5.5* 11.0*
    18.0 24.0 36.0 54.0 Mbit] ESS CH: 11
17:20:26.527457 1.0 Mb/s 2457 MHz 11b -80dB signal antenna 7 0us
    BSSID:d2:17:33:d1:4e:a7 (oui Unknown) DA:Broadcast SA:d2:17:33:d1
    :4e:a7 (oui Unknown) Beacon (SFR WiFi Mobile) [1.0* 2.0* 5.5*
    11.0* 18.0 24.0 36.0 54.0 Mbit] ESS CH: 11, PRIVACY
17:20:26.529053 1.0 Mb/s 2457 MHz 11b -81dB signal antenna 7 0us
    BSSID:00:17:33:d1:4e:a4 (oui Unknown) DA:Broadcast SA:00:17:33:d1
    :4e:a4 (oui Unknown) Beacon (NEUF_4EA0) [1.0* 2.0* 5.5* 11.0*
    18.0 24.0 36.0 54.0 Mbit] ESS CH: 11, PRIVACY
```

Le mode actif : Dans le mode de découverte de service actif, les terminaux Wi-Fi effectuent une recherche en émettant des trames de type *probe request*. Ces trames de type *Management* contiennent un champ SSID qui désigne le nom du réseau auquel ces requêtes sont destinées. A la réception d'une trame *probe request*, un point d'accès appartenant au réseau désigné par le SSID répondra au terminal en envoyant une trame de type *probe response*, déclarant ainsi sa présence.

Comme pour le mode passif, cette opération doit être effectuée sur chaque canal. Dans le mode actif, les réponses potentielles du point d'accès interviennent juste après l'émission de la *probe request*. Ainsi le terminal ne doit écouter le canal que pendant un intervalle de temps très bref. Le coût énergétique d'une émission de trame étant négligeable devant celui de la réception, le mode actif est moins gourmand en énergie que le mode passif et est donc privilégié par les terminaux mobiles.

Il est possible de capturer les trames *probe request* grâce à la commande *tcpdump* :

Dans le mode de découverte de service actif, les terminaux Wi-Fi effectuent une recherche en émettant des trames de type *probe request* (voir Figure 3). Ces trames de type *management* contiennent un champ SSID qui désigne le nom du réseau auquel ces requêtes sont destinées. A la réception d'une trame *probe request*, un point d'accès appartenant au réseau désigné par le SSID répondra au terminal en envoyant une trame de type *probe response*, déclarant ainsi sa présence. Comme pour le mode passif, cette opération doit être ef-

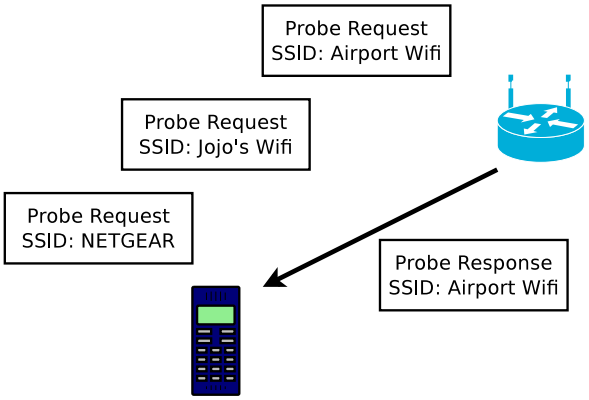

Figure 3: Mode de découverte de service actif.

fectuée sur chaque canal. Dans le mode actif, les réponses potentielles du point d'accès interviennent juste après l'émission de la *probe request*. Ainsi le terminal ne doit écouter le canal que pendant un intervalle de temps très bref. Le coût énergétique d'une émission de trame étant négligeable devant celui de la réception, le mode actif est moins gourmand en énergie que le mode passif et est donc privilégié par les terminaux mobiles.

La commande suivante permet à l'aide de tcpdump de capturer les trames *probe request* (en rouge les adresses MAC source et les SSIDs):

```
$ sudo tcpdump -i mon0 -e 'type mgt subtype probe-req' -vvv
10:49:19.754014 1.0 Mb/s 2462 MHz 11b -85dB signal antenna 7 0us
    BSSID:Broadcast DA:Broadcast SA:e8:40:f2:f9:ff:12 (oui Unknown)
    Probe Request () [1.0* 2.0* 5.5* 11.0* 9.0 18.0 36.0 54.0 Mbit]
10:49:25.805504 1.0 Mb/s 2462 MHz 11b -83dB signal antenna 7 0us
    BSSID:Broadcast DA:Broadcast SA:e8:40:f2:f9:ff:12 (oui Unknown)
    Probe Request () [1.0* 2.0* 5.5* 11.0* 9.0 18.0 36.0 54.0 Mbit]
10:49:28.340537 1.0 Mb/s 2462 MHz 11b -90dB signal antenna 7 0us
    BSSID:Broadcast DA:Broadcast SA:4c:72:b9:f4:52:cd (oui Unknown)
    Probe Request () [1.0* 2.0* 5.5* 11.0* 9.0 18.0 36.0 54.0 Mbit]
10:50:10.401049 1.0 Mb/s [bit 15] 0us BSSID:Broadcast DA:Broadcast
    SA:00:24:d7:59:e0:dc (oui Unknown) Probe Request (FreeWifi) [1.0
    2.0 5.5 11.0 6.0 9.0 12.0 18.0 Mbit]
10:50:10.600625 1.0 Mb/s [bit 15] 0us BSSID:Broadcast DA:Broadcast
    SA:00:24:d7:59:e0:dc (oui Unknown) Probe Request (Colubris) [1.0
    2.0 5.5 11.0 6.0 9.0 12.0 18.0 Mbit]
```

Les SSIDs contenus dans les trames probe request correspondent aux réseaux connus du terminal. Ce sont les réseaux auquel le terminal s'est déjà connecté et qui sont conservés en mémoire par le système d'exploitation dans ce que l'on appelle la *liste des réseaux configurés*. Lors d'une recherche en mode actif, le terminal émettra une trame *probe request* pour chaque réseau enregistré, et ceci sur chaque canal.

Time	Src address	Dest address	SS	SSID
Aug 26, 2013 11:34:09.634222000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-83	'Freebox-617A41'
Aug 26, 2013 11:34:17.710016000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-83	'SFR WiFi Public'
Aug 26, 2013 11:34:25.777786000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-82	'INRIA-guest'
Aug 26, 2013 11:34:26.148042000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-87	''
Aug 26, 2013 11:34:27.796123000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-83	'eduroam'
Aug 26, 2013 11:34:30.183995000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-87	''
Aug 26, 2013 11:34:33.845567000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-83	'eduroam'
Aug 26, 2013 11:35:20.609544000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-88	''
Aug 26, 2013 11:36:13.058027000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-88	''
Aug 26, 2013 11:37:49.875385000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-83	''
Aug 26, 2013 11:37:55.926222000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-84	''

Figure 4: Résumé d'une capture de trames *probe requests*, montrant pour chaque trame, l'heure de capture, l'adresse source, l'adresse destination, la puissance du signal (en dB) ainsi que le SSID lorsqu'il est présent.

Lorsqu'il n'est pas associé à un réseau Wi-Fi, un terminal utilisant le mode actif diffuse plusieurs fois par minute des trames non chiffrées contenant son adresse MAC et les SSIDs des réseaux auquel il s'est précédemment connecté . Ainsi ce terminal diffuse en clair son historique de connexion sous la forme d'une liste de SSID que nous appellerons l'*empreinte Wi-Fi* du terminal. Ces trames étant émises en rafale plusieurs fois par minute, se trouver à portée d'un tel terminal pendant une durée de l'ordre de la minute est suffisant pour récupérer son empreinte.

Pour faire face à ce problème évident de vie privée, une modification du protocole a été proposé afin que les trames *probe request* ne désignent plus de SSID. Dans cette variante, le champ SSID des trames *probe request* contient une chaîne vide, désignant n'importe quel SSID. Les trames probe request correspondantes sont qualifiées de *Broadcast*, et à la réception d'une telle trame, tout point d'accès doit répondre par une trame *probe response*. Cette variante est progressivement adoptée par les distributeurs de terminaux mobiles, mais il existe encore aujourd'hui une part significative des terminaux utilisant le mode actif qui utilisent encore la version diffusant des trames probe request avec un SSID.

Par ailleurs, le mode actif est aussi utile dans le cas des réseaux Wi-Fi dit *cachés*. En effet, ces réseaux dissimulent leur présence en ne diffusant pas de beacons et en ignorant les *probe request broadcast*; le seul moyen de les détecter consiste à utiliser des trames *probe request* ciblées, c'est à dire des requêtes contenant le SSID du réseau caché. Ainsi, un utilisateur souhaitant utiliser le mode caché pour rendre son réseau Wi-Fi plus discret (certains pensent même que cela améliore la sécurité du réseau) va en fait forcer ses appareils à annoncer en permanence leur association au réseau caché dans les trames *probe request* qu'ils émettent.

Le mode de découverte de service actif transforme donc nos terminaux mobiles non connectés en véritables balises radio qui ne cessent d'émettre un identifiant unique (l'adresse MAC) ainsi que leur empreinte Wi-Fi tant qu'ils ne sont pas associés à un réseau auquel ils se sont précédemment connectés (voir Figure 4).

3 Que révèlent les SSIDs ?

Nous parlions plus tôt des SSIDs diffusés par le mode de découverte de service actif du Wi-Fi. Nous allons maintenant voir ce qu'ils peuvent révéler sur le propriétaire du terminal.

Ces SSIDs peuvent être analysés de manière sémantique, c'est à dire en interprétant leurs sens, ou par une analyse systématique en les considérant comme un simple identifiant. Nous verrons qu'ils peuvent révéler des informations telles que des coordonnées géographiques, des noms de personnes et même des liens sociaux. Pour illustrer ce propos, nous prendrons des exemples dans une base de 20.000 SSIDs collectés en écoutant des trames *probes request* [6, 7].

3.1 SSIDs nominatifs

Chaque réseau Wi-Fi a un SSID. C'est son nom. Il va permettre à ses utilisateurs de l'identifier. Il doit être suffisamment explicite pour éviter toute ambiguïté, et comme il est le plus souvent configurable, certains propriétaires choisissent de personnaliser le SSID de leur réseau Wi-Fi. Ainsi, un nombre de SSIDs contiennent des informations nominatives que l'on peut classer de la manière suivante :

- **Nom d'entreprise ou d'organisation**. Exemple : *Global corp.*, *Université de Grenoble*
- **Nom de lieux**. Exemple : *Wi-Fi ville de Genève*, *Aéroport Charles de Gaule*, *Hilton Hotel - Paris*, *Hotel SANA Lisboa Network*
- **Nom d'événements** (conférence, salon, festival). Exemple : *GreHack13*, *Eurockéennes*, *GreHack*.
- **Nom et/ou prénom du propriétaire**. Exemple : *John Doe's Network*, *Famille_Snowden*, *Wi-Fi de Michel*
- **Adresse physique d'installation**. Exemple : *12 George St*, *34 avenue de la République appt 42*.

Un observateur avisé sera capable d'interpréter ces informations. Ici on pourra déduire des SSIDs que l'utilisateur du terminal s'est rendu à Genève et à l'*aéroport CDG*, qu'il a un lien avec l'entreprise *Global corp.*, qu'il s'est rendu à la conférence *GreHack13* et qu'il s'est connecté au réseau de *John Doe*.

Il faut également noter le cas des réseaux Wi-Fi hotspot créés par les terminaux mobiles pour partager une connexion avec d'autres appareils. Dans le cas des terminaux Apple, le SSID du réseau créé équivaut au nom affecté au terminal. Or ce nom est par défaut celui de l'utilisateur principal de l'équipement. Si ce dernier a utilisé sa véritable identité pour créer son compte, on voit alors apparaître des SSIDs du type *John Doe's iPhone* ou *MacBook Pro de Pierre Martin*.

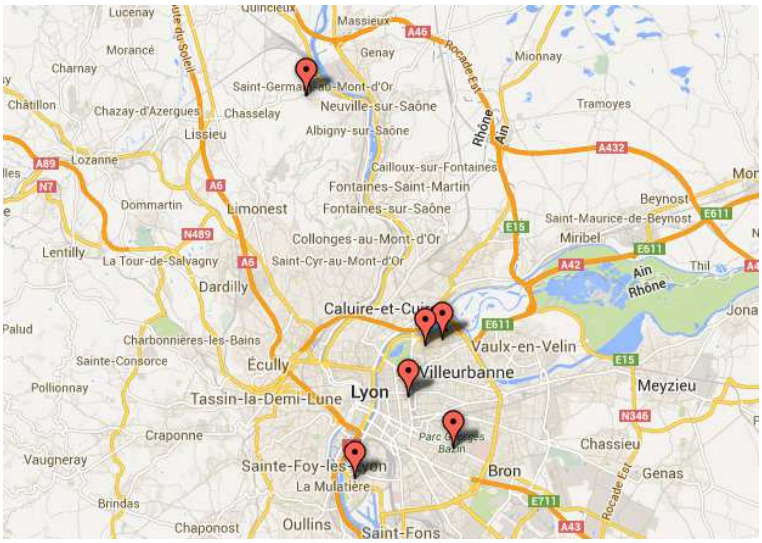

Figure 5: Carte montrant des coordonnées géographiques obtenues à partir de SSIDs.

3.2 Coordonnées géographiques

Ainsi que nous l'avons vu dans la section précédente, un SSID peut être suffisamment explicite pour indiquer une position géographique plus ou moins précise. Il est également possible d'obtenir de telles informations à partir d'un SSID sans passer par une analyse sémantique de celui-ci en s'appuyant sur des bases de données répertoriant les points d'accès Wi-Fi à l'échelle mondiale.

Ces bases, dont certaines sont libres d'accès telles que WiGLe (`http://wigle.net/`) et Openbmap (`http://openbmap.org/`), contiennent des informations telles que l'identifiant unique du point d'accès (le BSSID), le nom du réseau (le SSID), son canal, ses coordonnées GPS et d'autres éléments relatifs aux mécanismes de sécurité.

A partir d'un SSID il est possible de retrouver les coordonnées géographiques du réseau auquel il a été affecté en cherchant dans ces bases. De ces points géographiques, il serait possible de déduire d'autres informations personnelles telles que le lieu de domicile, le lieu de travail, des lieux de voyage professionnels ou personnels. On pourrait ensuite utiliser ces informations pour identifier le possesseur du terminal. En effet la paire de points géographiques domicile/travail constitue un excellent identifiant ; aux États-Unis cet identifiant est unique dans la plupart des cas [9]. La figure 5 montre l'exemple d'un nuage de points obtenu à partir des SSIDs diffusés par le smartphone de l'auteur.

Cette approche d'identification de points géographiques à partir de SSIDs a plusieurs limites. Premièrement, contrairement au BSSID, le SSID n'est pas un identifiant unique. Plusieurs réseaux distincts peuvent partager le même SSID. Ainsi pour un SSID donné, plusieurs points d'accès et donc plusieurs

coordonnées géographiques peuvent correspondre. C'est par exemple le cas des SSIDs communs tels que ceux configurés par défaut dans les routeurs (*linksys, NETGEAR, ...*) ou les hotspots (*MacDonald's HotSpot, FreeWifi, SFR HotSpot,* etc ...).

Le fait d'être associé à plusieurs points d'accès va ainsi réduire l'utilité de l'information géographique fournie par un SSID. Cette utilité dépendra de la taille de la zone géographique dans laquelle ces points d'accès sont répartis. Ainsi le SSID d'un réseau Wi-Fi d'un campus (ex : *UCBL*) fournira une information géographique avec une précision de l'ordre de quelques kilomètres. A l'opposé, aucune information géographique ne pourra être inférée à partir d'un SSID par défaut de routeur comme par exemple *NETGEAR*, puisque ces réseaux sont présents sur l'ensemble du globe. Au final, plus un SSID est rare, plus on pourra espérer obtenir une d'information géographique précise.

La seconde limite est que les bases de données ouvertes dont nous parlions plus haut (WiGLe et Openbmap) ne sont pas exhaustives et tous les points d'accès n'y sont pas répertoriés. En effet ces projets reposent sur des données collectées par des volontaires. Ce mode de fonctionnement ne permet qu'une couverture partielle du paysage mondial des points d'accès Wi-Fi.

Cependant, il existe des bases de données non publiques qui répertorient les points d'accès Wi-Fi avec leurs coordonnées géographiques. Il s'agit des bases de données des systèmes de géolocalisation basée sur le Wi-Fi tels que *Google Maps Geolocation* et *Skyhook*.

Ces systèmes constituent une alternative au système GPS. En leur soumettant la liste des points d'accès visibles depuis notre position ainsi que la force de signal pour chacun de ces points d'accès, ils renvoient un positionnement avec une précision de quelques dizaines de mètres. Pour arriver à ce résultat, ils maintiennent une base de données de points d'accès, qui se met à jour et s'autocorrige au fur et à mesure de son utilisation. Ces bases sont plus complètes que les bases de données ouvertes, mais elles ne sont pas en libre accès. On peut supposer qu'un accès direct à ces bases permettrait d'inférer d'avantage d'informations géographiques que celles obtenues à partir de leurs variantes ouvertes.

3.3 Liens sociaux

Une information inattendue que l'on peut obtenir à partir des SSIDs diffusés par des terminaux Wi-Fi est l'existence de liens sociaux entre les possesseurs de ces terminaux. En effet, nous avons montré [6, 7] que des liens sociaux pouvaient être identifiés en comparant des *empreintes Wi-Fi* (la liste de SSIDs diffusés par un terminal).

L'idée derrière cette approche est que des personnes ayant un lien social auront tendance à utiliser les mêmes réseaux sans fils : réseaux Wi-Fi personnels respectifs, ou alors des réseaux Wi-Fi utilisé lors d'activités sociales. Ainsi, des personnes ayant un grand nombre de SSID en commun sont probablement liées par un lien social. Il faut également prendre en compte la rareté des SSIDs partagés. En effet, partager un réseau personnel avec un nom rare tel que *Réseau de M.Cunche* est un bon indicateur de lien social. Au contraire, partager

un réseau commun tel que *NETGEAR* ou *Mc Donald FreeWifi* n'implique pas forcément l'existence d'un tel lien.

Pour formaliser cette approche, nous avons utilisé une métrique de similarité pour comparer les empreintes. Plus cette métrique est grande, plus les empreintes sont similaires, et inversement. Au-dessus d'un certain seuil de similarité (déterminé en avance par des essais sur des échantillons contrôlés) on considère qu'il existe un lien social entre les possesseurs des terminaux. Comme indiqué plus tôt, pour inférer l'existence d'un lien social, il faut prendre en compte le nombre de SSIDs en commun ainsi que leur rareté. En utilisant une métrique qui prend en compte ces deux caractéristiques, nous avons construit un détecteur de lien social performant qui détecte 80

En utilisant cette approche il est possible de reconstruire un réseau social rien qu'en écoutant les SSIDs diffusés par les terminaux Wi-Fi. Ceci est une menace supplémentaire pour la vie privée. Tout d'abord, car nos liens sociaux n'ont pas toujours vocation à être public et ensuite parce qu'ils peuvent être utilisé pour identifier les propriétaires des terminaux. En effet au sein d'un réseau social, tout individu est presque toujours identifiable par sa liste de contact. L'identification de liens sociaux n'est qu'un exemple d'information que l'on peut inférer à partir des SSIDs diffusés par nos terminaux mobiles, et il est probable que d'autres types d'informations puissent être déduits des empreintes Wi-Fi.

4 Tracer les individus via Wi-Fi

Le fait que nos terminaux Wi-Fi déclarent en permanence leur présence peut être exploité pour capturer et analyser les mouvements de leurs porteurs dans le monde physique. On assiste actuellement à l'émergence de systèmes de traçage Wi-Fi qui enregistrent à large échelle les déplacements des individus grâce aux signaux émis par leurs terminaux. Ces systèmes reposent sur une technologie de géolocalisation radio et sont utilisés pour fournir des statistiques sur les déplacements des individus dans des zones d'intérêt. Ils sont aussi parfois utilisés pour diffuser des messages publicitaires ciblés.

4.1 Géolocalisation radio

Avant de parler de système de traçage par Wi-Fi, il convient de parler de géolocalisation radio.

La géolocalisation consiste à positionner un objet dans l'espace en mesurant sa distance par rapport à des points de référence dont la position est préalablement connue. Pour déterminer une position sur un plan, la distance par rapport à 3 points est suffisante. L'exemple le plus célèbre de système de positionnement est le GPS (*Global Positionning System*). Dans le cas de ce système, les points de référence sont des satellites en orbite basse dont la position est connue en temps réel et la distance entre l'objet et les satellites est déduite à partir du temps de trajet des signaux radio émis par ces derniers.

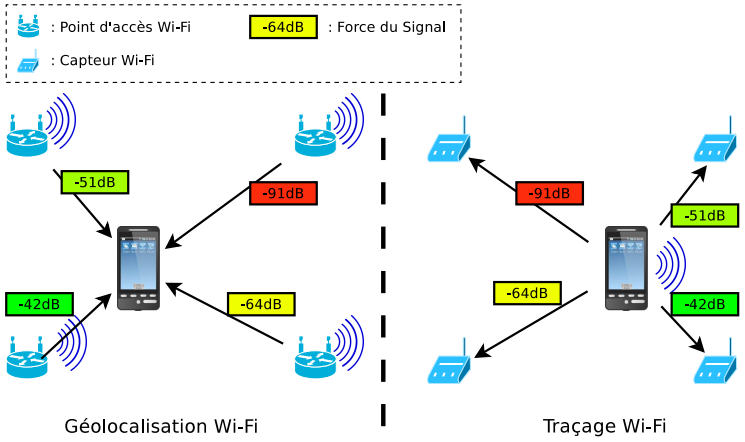

Figure 6: Les deux approche de géolocalisation par Wi-Fi.

En utilisant une approche similaire à celle du GPS il est possible de détourner la technologie Wi-Fi de son usage principal pour faire de la géolocalisation. En utilisant les points d'accès comme points de référence et en évaluant les distances grâce à la force des signaux reçus, on peut déterminer la position d'un terminal Wi-Fi.

Cette géolocalisation peut être effectué de deux manières. Soit le terminal détermine de manière autonome sa position à partir des trames beacon émises par les points d'accès alentour, soit un ensemble de capteurs Wi-Fi déterminent la position d'un terminal à partir des trames émises par ce dernier. Le premier cas correspond au système utilisé par les smartphones pour déterminer leur position sans faire appel au module GPS gourmand en énergie. C'est le cas de services tels que *Google Maps Geolocation API* qui utilise un système de requêtes avec une base de données contenant les positions des points d'accès Wi-Fi à l'échelle mondiale.

Le second cas de géolocalisation par Wi-Fi est celui qui va nous intéresser par la suite. Cette fois-ci, c'est un système qui écoute passivement les communications sur les canaux Wi-Fi afin de détecter et de géolocaliser les terminaux ayant leur interface Wi-Fi activée. Comme dans le cas précédent, la force du signal est utilisée pour évaluer la distance entre l'émetteur et le récepteur de la trame. Si les trames émises par un terminal sont reçues par plusieurs capteurs, la position du terminal pourra être déterminée avec une précision qui peut aller jusqu'au mètre. Il est important de noter que dans ce second cas, le système n'a pas besoin de la collaboration du terminal. En particulier, il n'est pas nécessaire d'installer une application sur le terminal pour que celui-ci puisse être détectable par le système.

4.2 Le traçage Wi-Fi

Nous venons d'expliquer comment un ensemble de capteurs pouvaient détecter et géolocaliser des terminaux Wi-Fi à partir des trames qu'ils émettent. Comme nous le savons déjà, les terminaux mobiles ayant leur interface Wi-Fi activée émettent régulièrement des trames même lorsqu'ils ne sont pas connectés à un réseau. Cette particularité a été exploitée pour créer des systèmes de détection et de géolocalisation des individus porteurs de terminaux Wi-Fi. Dans ces systèmes, chaque individu est identifié par l'adresse MAC de son terminal et les informations ainsi collectées sont utilisées pour des applications de traçage et d'analyse de foule que nous détaillerons plus loin.

Un système de traçage Wi-Fi est composé d'un ensemble de capteurs déployés dans la zone d'intérêt à surveiller. Il inclut également un serveur central qui se charge de collecter et de traiter les informations provenant des capteurs. Certains systèmes commerciaux ont fait le choix d'héberger le serveur sur des plates-formes de *Cloud Computing*. A titre d'exemple, *Euclid Analytics* utilise la plate-forme *Amazon Web Services*. L'utilisation de ces plates-formes peut poser certains problèmes de confidentialité des données.

Les capteurs utilisés dans les systèmes de traçage possèdent au minimum une interface Wi-Fi en mode monitor, une capacité de calcul modeste. Ce cahier des charges peut être rempli par des appareils tels que des routeurs Wi-Fi modifiés ou d'autres plateformes embarquées telles que le *Raspberry-Pi* [4]. Avec de tels appareils, il est possible de mettre en place un système de traçage à moindre coût.

Pour la transmission des données collectées par les capteurs, une interface filaire reliée à une infrastructure réseau est le plus souvent utilisée. Cependant, certains systèmes sont capables de créer une infrastructure sans fil autonome à partir des interfaces Wi-Fi des capteurs. C'est par exemple le cas du système *Navizon ITS* qui utilise des capteurs basés sur des points d'accès *Open-Mesh* capables de s'organiser en un réseau Wi-Fi maillé.

Pour réduire d'avantage les coûts de déploiement, les entreprises de traçage Wi-Fi ont établi des partenariats avec les gestionnaires des réseaux Wi-Fi déjà déployés dans les zones d'intérêt. Ainsi, afin d'éviter de déployer de nouveaux capteurs, les routeurs Wi-Fi déjà en place reçoivent une version modifiée du firmware qui leur permet de jouer le rôle de capteur Wi-Fi en plus des fonctions qui leur sont dévolues. Le coût de déploiement d'un système de traçage Wi-Fi est alors proche de zéro.

4.3 Champs d'application

Les informations collectées par les systèmes de traçage Wi-Fi sont utiles pour quiconque souhaite analyser les déplacements d'individus dans une zone d'intérêt.

Ainsi, ces systèmes sont utilisés pour analyser les mouvements des clients dans les magasins et les centre commerciaux (voir Figure 7). Ils fournissent des informations telles que l'affluence en temps réel, la fréquence des visites, le temps passé dans le magasin ou devant la vitrine, les rayons visités, etc.

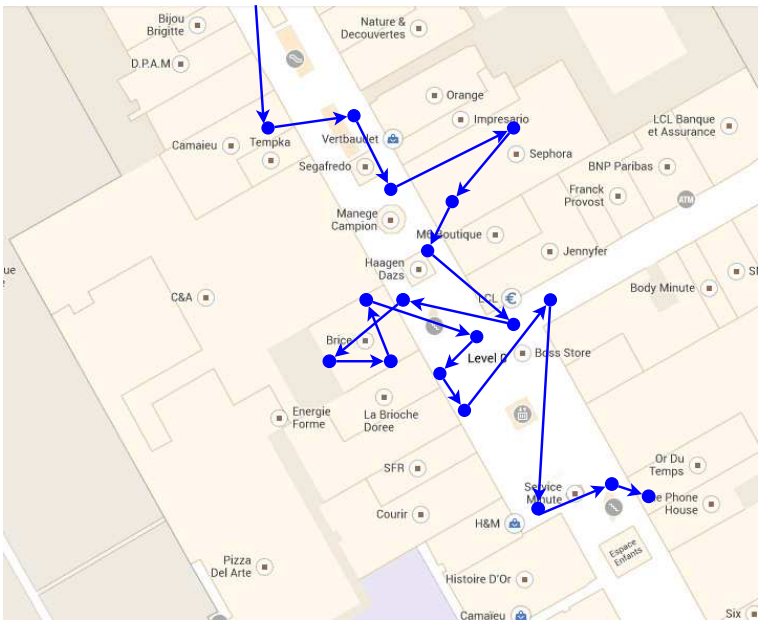

Figure 7: Exemple fictif de trace de mobilité obtenu à partir d'un système de traçage.

Ces outils que l'on appelle *Physical analytics* sont l'équivalent, dans le monde réel, des *Web analytics* qui existe sur le Web pour analyser l'audience des sites Web. Ils permettent aux gérants de ces espaces d'améliorer leur commerce en modifiant la configuration de leur points de vente et en optimisant l'allocation des ressources humaines.

Ces systèmes suscitent un grand engouement de la part des gestionnaires de commerce et ils sont déjà déployés dans un grand nombre de sites. D'après *Euclid* , son système a déjà tracé plus de 50 millions de terminaux en seulement quelques mois d'activité[3].

Les systèmes de traçage Wi-Fi sont également utilisés par les collectivités pour analyser les flux sur les axes de transports et les lieux publics. En plaçant des capteurs le long d'une route, il est possible de mesurer la densité du trafic en temps réel, de détecter les engorgements, ou encore de mesurer le temps de trajet entre deux points. Il a également été proposé de les utiliser pour mesurer les foules dans des événements de masse tels que des concerts ou des manifestations [4].

Une autre utilisation possible de ces systèmes, est la mise en place de publicités ciblées sur des panneaux d'affichage dit "intelligents". Le principe est de reconnaître les passants grâce au Wi-Fi et d'afficher des publicités spécialement sélectionnés pour correspondre à leurs profils. Un système de traçage Wi-Fi com-

[3]voir `http://www.franken.senate.gov/files/docs/130328_Euclid.pdf`

biné à un système d'affichage publicitaire intelligent a récemment été déployé à Londres[2]. Un ensemble de capteurs Wi-Fi intelligemment placés dans un bar collecte les adresses MAC des clients en même temps qu'il établit leur profil. Ce profil est constitué d'informations classiques telles que la durée et la fréquence des visites la durée, la nature de la visite (verre en terrasse ou repas à l'intérieur) et va jusqu'au sexe de l'individu grâce à des capteurs placés dans les WC du pub. Les panneaux publicitaires intelligents placés dans le quartier, détectent la présence des clients du bar parmi les passants et affiche des offres correspondant à leurs profils.

Au delà de ces applications *utilitaires* ces systèmes peuvent être utilisés pour des taches d'espionnage et d'intelligence. Par exemple le projet Open Source *Snoopy* [8] fournit les bases logicielles pour mettre en place un système de traçage Wi-Fi avec du matériel informatique commun. Plus récemment, un nouveau système de traçage Wi-Fi a été présenté à la conférence Black-Hat. Ce système, dénommé *CreepyDOL* [12], a la particularité d'intégrer des mécanismes de sécurisation des données et des communications qui empêche l'identification du contrôleur du système dans le cas où des capteurs seraient découverts. Ces systèmes artisanaux ne constituent probablement que la partie émergée de l'iceberg, et il y a fort à parier que des agences de renseignement possèdent et utilisent déjà des systèmes de traçage Wi-Fi.

4.4 Traçage Wi-Fi et vie privée

Quelle que soit la technologie employée, le traçage des déplacements d'individus représente une menace sérieuse pour la vie privée. Ceci est d'autant plus vrai lorsque ce traçage se fait à l'insu de l'individu. En effet, ce traçage étant passif, les individus tracés n'ont aucun moyen de se rendre compte que leur présence et leurs déplacements sont enregistrés. En guise d'indication, des écriteaux indiquant en termes vagues la présence d'un système de traçage sont mis en place aux entrées des zones sous surveillance. L'ampleur de ces systèmes de traçage et leur caractère invisible a fait apparaître des inquiétudes concernant leur impact sur la vie privée. A ce sujet on peut se référer à la déclaration du sénateur de l'état du Minnesota sur les systèmes de traçage Wi-Fi [1].

L'étude de l'impact sur la vie privée comprend toujours l'analyse de la durée de conservation des données. Pour calculer les statistiques sur les visiteurs d'une zone d'intérêt, les systèmes de traçage doivent conserver des données pendant une période de durée variable. Si compter le nombre de visiteurs quotidiens ne nécessite qu'une conservation de donnée à l'échelle de la journée, calculer la fréquence de visite nécessite la conservation des données sur des durées beaucoup plus longues. Les données sur les individus sont donc conservées longtemps après leur visite de la zone surveillée.

Les concepteurs de ces systèmes objectent qu'ils ne stockent que l'adresse MAC du terminal à partir de laquelle on ne peut pas retrouver l'identité du propriétaire. Malheureusement, il existe de nombreux moyens pour obtenir l'association entre une adresse MAC et l'identité d'une personne. On peut par exemple y accéder directement sur le terminal ou à distance grâce à des tech-

niques de rejeux des communications Wi-Fi [5]. L'adresse MAC peut également être collectée de manière plus systématique par des applications mobiles. C'est le cas par exemple de l'application mobile RATP qui accède à l'adresse MAC et au nom du terminal et qui transmet ensuite ces informations à des serveurs contrôlés par une tierce partie [3]. L'adresse MAC ne peut donc pas être considérée comme un identifiant anonyme.

Pour pallier aux potentiels problèmes de vie privée, certains systèmes de traçage utilisent un mécanisme d'anonymisation de l'adresse MAC. Au lieu de conserver l'adresse MAC telle quelle, une fonction de hachage cryptographique est utilisée pour la transformer en un identifiant "anonyme". Malheureusement cette mesure n'est pas suffisante pour protéger l'anonymat des individus [10].

5 Stopper la fuite d'information

Les problèmes de vie privée dont nous venons de discuter n'ont pas été causés par un changement dans la technologie Wi-Fi. C'est plutôt un changement dans son mode d'utilisation qui en est à la source. En effet, le Wi-Fi a été initialement conçu pour remplacer les infrastructures filaires entre ordinateurs fixes ou portables. Aujourd'hui, nous trouvons de la connectivité Wi-Fi un peu partout : à notre domicile, chez des amis, dans les halls de gare, dans les hôtels, etc. Nos smartphones, qui ne nous quittent jamais, ont bien souvent leur interface Wi-Fi active, cherchant en permanence des réseaux et se connectant automatiquement à ceux qu'ils connaissent. Les concepteurs du Wi-Fi n'avaient probablement pas prévu ces changements et si les problèmes de sécurité avaient été clairement identifiés, les problématiques de vie privée n'avaient pas à l'époque l'importance qu'elles ont aujourd'hui.

Malgré le tableau sombre que nous venons de dresser, il existe des solutions aux problèmes de vie privée posés par la technologie Wi-Fi.

La première consiste à modifier le protocole afin de ne plus laisser transparaître en clair des données personnelles et des identifiants uniques. Plusieurs protocoles reposant sur la cryptographie pour cacher ces données sensibles ont déjà été proposés [11]. Ces protocoles ne sont pas rétro-compatibles avec les protocoles utilisés actuellement. Leur déploiement parait donc difficilement envisageable car ils nécessiteraient une modification en profondeur des stations et points d'accès existants. Il faudra probablement attendre une nouvelle version du Wi-Fi ou l'émergence d'un protocole concurrent pour voir une intégration de ces mécanismes.

Une autre solution consisterait à utiliser une information de *géolocalisation* pour la découverte de service [7]. En effet un point d'accès ne couvre qu'une petite zone géographique. En gardant en mémoire les coordonnées géographiques de nos réseaux Wi-Fi favoris, on pourrait tirer parti de cette information pour ne chercher que des réseaux qui sont susceptibles d'être à portée. Par exemple, ne chercher son réseau Wi-Fi personnel que quand on s'approche de notre domicile et ne pas faire de découverte de service active lorsque l'on est éloigné de tous les points d'accès connus. Cette modification du mécanisme de découverte de

service actif nécessite des modifications mineures au système d'exploitation et permet de limiter de manière significative la fuite d'informations personnelles.

Pour finir, la solution ultime semble être de désactiver l'interface Wi-Fi de notre terminal mobile, ainsi que son interface Bluetooth qui expose l'utilisateur à des risques similaires. Malgré ces précautions, il pourrait subsister des problèmes de vie privée liés à l'interface radio avec laquelle notre smartphone se connecte au réseau cellulaire. En effet, au sein des réseaux GSM et UMTS, les terminaux sont également identifiés par des numéros uniques (TMSI et IMSI). Heureusement, des précautions ont été prises pour dissimuler ces identifiants et ainsi empêcher le traçage des utilisateurs. Il faut donc espérer que ces mesures se révéleront efficaces.

References

[1] Lettre du sénateur Al Franken à Euclid inc, 13 mars 2013,. http://www.franken.senate.gov/?p=hot_topic&id=2325.

[2] This recycling bin is following you. Quartz.com. 8 août 2013. http://qz.com/112873/this-recycling-bin-is-following-you/.

[3] Jagdish Achara, Prasad, James-Douglass Lefruit, Vincent Roca, and Claude Castelluccia. Detecting Privacy Leaks in the RATP App: how we proceeded and what we found. In *GREHACK 2013*, Grenoble, France, November 2013. Guillaume Jeanne. to appear also in Springer Journal of Computer Virology and Hacking Techniques (JCVHT).

[4] Bram Bonne, Arno Barzan, Peter Quax, and Wim Lamotte. Wifipi: Involuntary tracking of visitors at mass events. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*, pages 1–6, 2013.

[5] Mathieu Cunche. I know your MAC Address: Targeted tracking of individual using Wi-Fi. In *International Symposium on Research in Grey-Hat Hacking - GreHack*, Grenoble, France, November 2013.

[6] Mathieu Cunche, Mohamed Ali Kaafar, and Roksana Boreli. I know who you will meet this evening! Linking wireless devices using Wi-Fi probe requests. In *WoWMoM - 13th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks - 2012*, San Francisco, United States, 2012.

[7] Mathieu Cunche, Mohamed-Ali Kaafar, and Roksana Boreli. Linking wireless devices using information contained in Wi-Fi probe requests. *Pervasive and Mobile Computing*, (0):–, 2013.

[8] Cuthbert Daniel and Wilkinson Glenn. Snoopy: Distributed tracking and profiling framework. In *44Con 2012*, 2012.

[9] Philippe Golle and Kurt Partridge. On the anonymity of home/work location pairs. In *Proceedings of the 7th International Conference on Pervasive Computing*, Pervasive '09, pages 390–397, Berlin, Heidelberg, 2009. Springer-Verlag.

[10] Cédric Lauradoux and Levent Demir. Guesswork, October 2013. Multisystem & Internet Security Cookbook (MISC) Hors-série, 8.

[11] Janne Lindqvist, Tuomas Aura, George Danezis, Teemu Koponen, Annu Myllyniemi, Jussi Mäki, and Michael Roe. Privacy-preserving 802.11 access-point discovery. In *Proceedings of the second ACM conference on Wireless network security*, WiSec '09, pages 123–130, New York, NY, USA, 2009. ACM.

[12] Brendan O'Connor. CreepyDOL: Cheap, Distributed Stalking. In *BlackHat*, 2013.