Les identifiants numériques et leur utilisation

FICHE INFO DU PPDT

INTRODUCTION

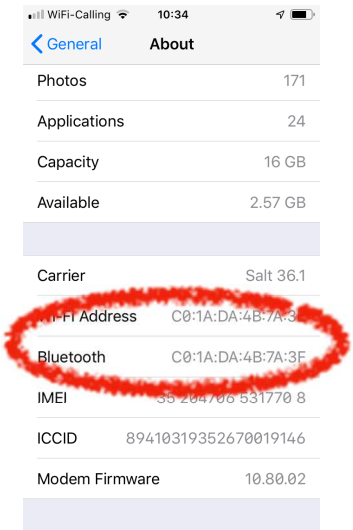
Notre vie quotidienne devient de plus en plus connectée grâce à des dizaines d'appareils (ordinateurs, téléphones, téléviseurs intelligents, etc.). Malgré la grande diversité de ces appareils, les méthodes sous-jacentes pour les connecter sont les mêmes et peu nombreuses. Chacune de ces techniques se sert d'un type d'identifiants pour les communications entre les appareils.

Dans cette fiche informative nous allons explorer les identifiants utilisés par la plupart des appareils connectés. Pour chaque type d'identifiants, nous expliquons:
Quelle est l'utilisation "normale"
Quelles sont les utilités "alternatives" telles que le traçage (de l'appareil, et de son utilisateur)
Quels sont les moyens d'éviter ces traçages.

Cette fiche est destinée aux lecteurs curieux, sans bagage technique nécessaire.

LES ADRESSES MAC

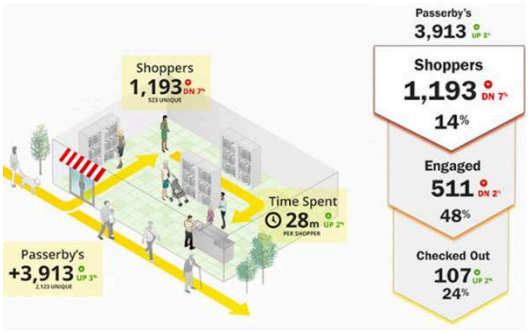
Sur un "réseau local" tel que l'ensemble des appareils dans une maison, connectés entre eux et avec le routeur Internet, les appareils communiquent selon un "protocole" qui donne à chacun son tour (sinon, comme quand plusieurs personnes dans une salle parlent en même temps, elles ne s'entendent et ne se comprennent pas bien). Ce protocole est le MAC, de "Medium Access Control". Quand deux appareils communiquent entre eux (ex. l'ordinateur au routeur Internet), ils s'identifient par leurs adresses MAC : "Je suis l'appareil A, et j'envoie ce message à l'appareil B". Le A et le B représentent les adresses MAC pour ces communications. A noter que c'est pour le "réseau local", formé par les appareils dans une petite région (à quelques dizaines de mètres de distance). Un réseau lointain n'entendrait pas ces communications, ni les A ni les B, et pourrait communiquer simultanément sans problème (comme pour d'autres personnes qui parlent dans une autre salle lointaine).



Ces adresses MAC ont un format spécifique tel que C0:1A:DA:4B:7A:3F. On peut les trouver dans les "paramètres" des téléphones mobiles. Dans ce cas-là, le WiFi a une adresse MAC (pour communiquer avec d'autres appareils comme le routeur WiFi), et le Bluetooth a une autre adresse MAC (pour communiquer avec d'autres appareils Bluetooth, tels que les écouteurs/casques).

Les utilisations « alternatives »

L'utilité des adresses MAC pourrait paraître purement technique, anodine. Cependant, leur traçage fournit une utilité analytique lucrative. Par exemple, en connectant son téléphone au WiFi d'un centre commercial pour accéder à Internet, les déplacements de l'utilisateur pourraient être suivis. C'est ainsi que le gestionnaire du réseau pourrait par exemple :

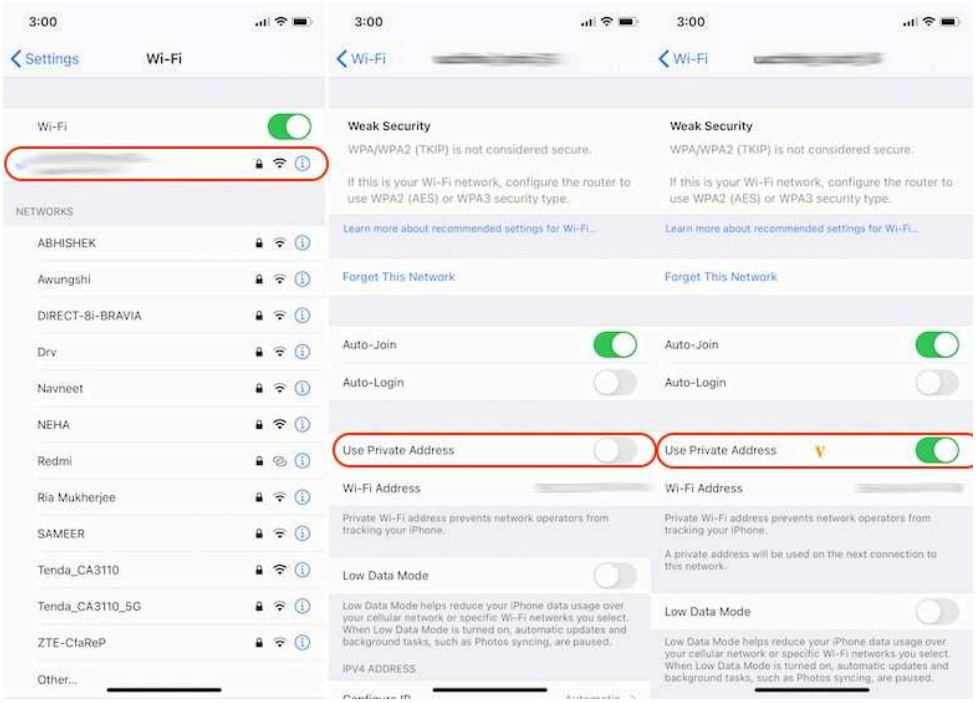


- Compter les clients
- Suivre les parcours des clients entre les rayons
- Créer des "heatmaps" (cartographie sur la concentration des clients dans le magasin)
- Identifier les coins les plus visités du centre commercial (et ainsi adapter leurs loyers ?)
- Mesurer les fréquences et les durées des visites
- Identifier si un client a été déjà vu dans une autre filiale

Et ce n'est pas que les déplacements des personnes qui sont valorisés. Les premiers caractères d'une adresse MAC identifient le fabriquant de l'appareil, les suivants identifient le modèle. C'est ainsi que le pouvoir d'achat d'un client pourrait être estimé à partir de son adresse MAC (ex. un téléphone dernière génération vs un vieux téléphone pas cher).

Comment éviter les traçages des adresses MAC ?

Pour pallier ce problème de traçage, certaines marques de téléphones donnent la possibilité de changer ces adresses MAC afin de l'éviter.



Quant aux ordinateurs, avec quelques connaissances informatiques c'est possible de changer les adresses MAC (les clients n'amènent pas souvent leurs ordinateurs dans les centres commerciaux... mais ça peut être le cas dans les hôtels)

Mais il faut faire attention : dès qu'un utilisateur est connecté avec un nom d'utilisateur, ou qu'un numéro de téléphone est fourni, le traçage pourrait se faire grâce à ces derniers, même quand l'adresse MAC de l'appareil change !

LES ADRESSES IP

La section précédente expliquait l'utilisation des adresses MAC pour que des appareils communiquent entre eux sur un réseau local : du téléphone aux écouteurs, de l'ordinateur jusqu'au routeur, etc. Mais comment communiquer à travers l'Internet d'un bout du monde à l'autre ? C'est là où les adresses IP sont utilisées (IP : Internet Protocol). Tout appareil connecté à Internet doit avoir une adresse IP. Avec une adresse IP d'un appareil-destination (ex. page web), les messages de communication sont "routés" jusqu'à celui-là, lequel pourrait répondre utilisant l'adresse de l'appareil-source (ex. un ordinateur utilisé pour visiter la page web)[1].

Une adresse IP (version 4) est formée de 4 numéros, comme 167.147.22.172. Chaque numéro va de 0 à 255, ce qui fait qu'au total il y a 256 x 256 x 256 x 256 = 4.29 milliards d'adresse IP (v4). Ce n'est pas assez si on considère tous les appareils du monde entier qu'on connecte à l'Internet.

Pour pallier ce problème, la pratique est que les fournisseurs Internet achètent des "plages d'adresses", qu'ils attribuent d'une manière dynamique aux utilisateurs[2]: l'adresse IP d'un utilisateur-A pourrait être attribuée à un utilisateur-B dès que l'utilisateur-A éteint son

---

[1] *Je vous laisse imaginer comment combiner adresses MAC et adresses IP sur plusieurs "tronçons locaux" pour qu'une communication aboutisse de bout-en-bout.*

[2] *Il y a aussi la possibilité d'acheter une adresse IP fixe.*

appareil par exemple. Ce qui signifie qu'une adresse IP à elle seule ne pourrait pas identifier un utilisateur. Il faut avoir recours à l'opérateur pour qu'il puisse identifier qui avait une certaine adresse à un certain moment (ex. si la justice cherche un criminel qui a utilisé une certaine adresse pour visiter des sites web illégaux).

IP(v4) date du début des années 80, quand les 4.29 milliards d'appareils paraissaient non-atteignables. Maintenant, on utilise IP version 6, où le format est :

2001:171b:227d:d170:dd10:e0a0:a4a7:8471

Ce qui fait 340'282'366'920'938'463'463'374'607'431'768'211'456 adresses au total. Ceci résout le problème de pénurie d'adresses IP(v4), sans pour autant garantir qu'une certaine adresse IP corresponde tout le temps à un seul appareil / une seule personne.

Cela dit, un site web de vente de livres, par exemple, ne va pas pouvoir recommander un livre à l'utilisateur derrière l'adresse IP-X juste en supposant que c'est la même personne avec la même adresse IP-X qui avait acheté un livre similaire une semaine auparavant (il va falloir que, comme pour l'exemple de la justice et le criminel, le site web demande au fournisseur Internet qui est la personne derrière cette adresse IP, ce qui n'est ni légal ni pratique à grande échelle). Pour ce type d'identifications / recommandations, les sites web utilisent les "cookies" qu'on explique dans la section suivante.

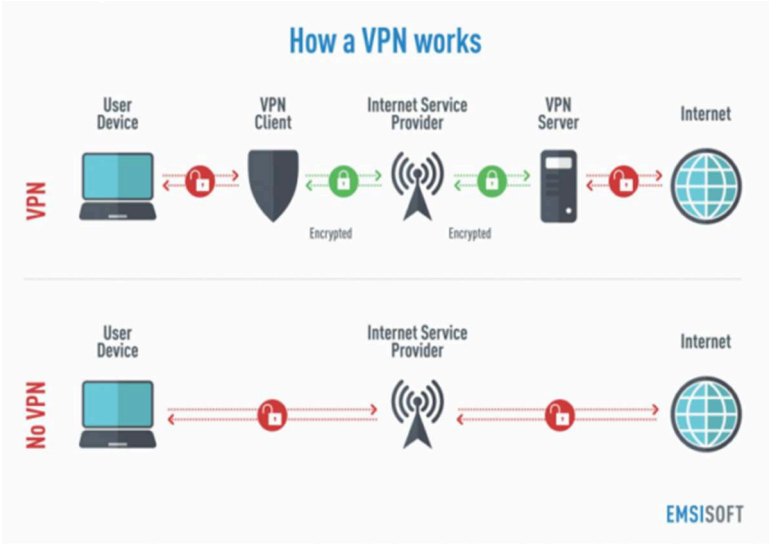
Les utilisations « alternatives »

Cependant, le site web qui voit l'adresse IP du visiteur pourrait reconnaître le fournisseur Internet (qui a réservé la "plage d'adresses"), ainsi que sa région, pour :

- Adapter la langue de la page web
- Faire des publicités adaptées à la région (mais pas à l'utilisateur)
- Choisir automatiquement la monnaie (CHF ou EUR etc.)
- Ou même adapter les prix, selon le pays de l'utilisateur.

Comment éviter l'identification par les adresses IP ?

Quand l'utilisateur aimerait cacher sa région, afin de profiter d'un prix plus bas offert dans une autre région par exemple, il pourrait utiliser un serveur VPN (Virtual Private Network) situé dans l'autre région. Le serveur web verra l'adresse IP du serveur VPN plutôt que l'adresse IP de l'utilisateur d'origine, et s'adapte selon la région du serveur VPN (langue, monnaie, prix etc.).



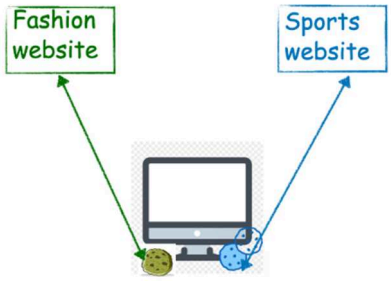
Quant à la possibilité de l'identification de l'utilisateur (pas seulement de sa région) quand la justice impose au fournisseur Internet la révélation de l'utilisateur d'une adresse IP, il y a des moyens de se cacher. Pour être plus correct, c'est aussi le moyen pour un journaliste de cacher ses connexions des surveillances d'un régime autoritaire. Le réseau TOR (The Onion Router) fonctionne comme une séquence de connexions VPN de l'exemple précédent.

Dans cette séquence, chaque "nœud" connaît son prédécesseur et son successeur sur le chemin, mais personne n'a une vue globale (ex. tel utilisateur s'est connecté à tel site web). TOR est un outil utilisé par des militants pour éviter les persécutions quand ils s'expriment sur certains site web. Mais TOR est aussi utilisé par des réseaux criminels sur tout ce qui se passe sur le "dark web".

Les identifiants numériques et leur utilisation

FICHE INFO DU PPDT

LES « COOKIES »



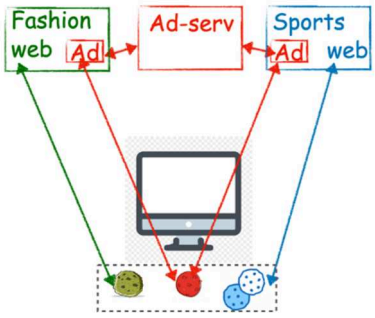
Dans la section précédente, nous avons vu qu'un serveur web ne pourrait pas supposer que, à deux instants différents, c'est toujours le même utilisateur derrière une même adresse IP, puisque l'attribution des adresses IP est souvent dynamique. Et puis comment reconnaître le même utilisateur derrière des adresses IP différentes ? Dans les années 1994-1995, ça a commencé à être un besoin pour pouvoir suivre les préférences des utilisateurs, les profiler, et leur faire des recommandations. La solution est que le serveur web enregistre un "cookie" sur l'ordinateur de l'utilisateur. C'est un petit fichier avec le nom du serveur web, un identifiant aléatoire de l'utilisateur, et une durée de vie. Dès que le serveur web revoit le même cookie, il pourra supposer que c'est le même utilisateur, peu importe si c'est la même adresse IP ou pas.

Chaque serveur web peut laisser ses propres cookies sur les ordinateurs des visiteurs et les relire par la suite. En revanche, un serveur web ne pourrait pas lire/changer les cookies laissés par d'autres serveurs. A chaque serveur web ses propres cookies.

Chaque serveur web va pouvoir suivre les activités/achats des utilisateurs sur ce site web, les profiler, et leur faire des recommandations, sans avoir aucune information sur leurs activités sur d'autres sites web.

L'intérêt économique se manifeste encore une fois et la nouvelle question est : comment pouvoir tracer les activités des utilisateurs à travers divers sites web, pour pouvoir améliorer le profilage et leur passer de meilleures recommandations ou des publicités personnalisées ?

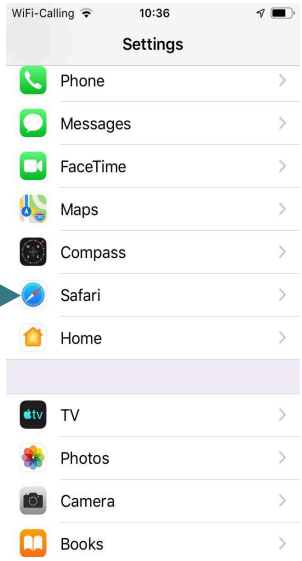
Les cookies tiers

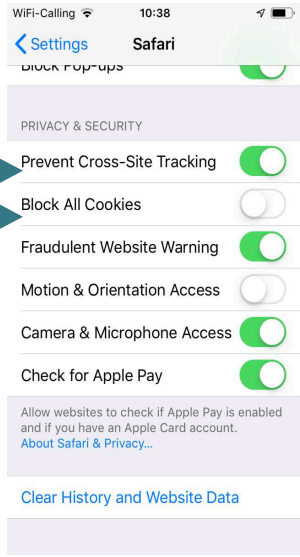


En allouant une partie de leurs pages web à un serveur tiers (de publicité, par exemple) contre une petite rémunération, les serveurs web primaires vont permettre au serveur tiers d'installer ses propres cookies sur l'ordinateur du visiteur, et de pouvoir tracer ce dernier à travers divers sites.

C'est ainsi que les serveurs de publicités (tel que Google par exemple) forment des profils accrus des utilisateurs, à travers divers sites web, lancent des enchères pour les publicitaires, et offrent des publicités personnalisées sur le site visité, en l'espace de quelques millisecondes.

Comment contrôler les cookies ?



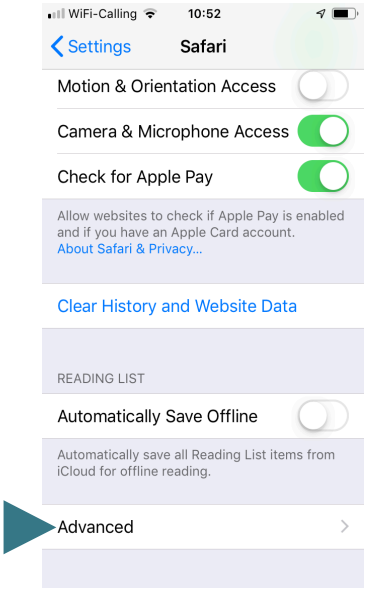


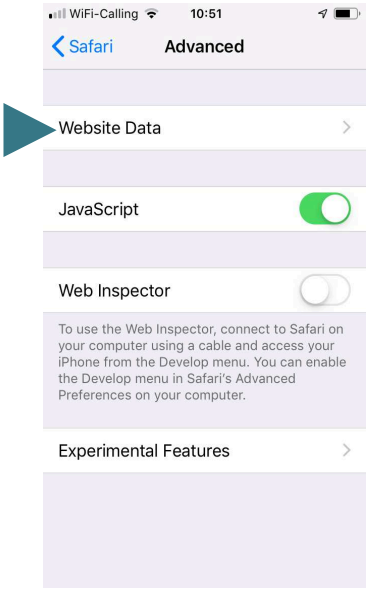
Il est possible de configurer le navigateur (Safari, Chrome, Firefox, Internet-Explorer etc.) pour effacer ou bloquer tous les cookies (au prix d'une dégradation de praticabilité, puisque les cookies primaires seraient bloqués aussi) ou juste les cookies tiers. Dans ce dernier cas, l'utilisateur continuerait à recevoir des publicités, mais celles-ci ne seraient plus personnalisées.

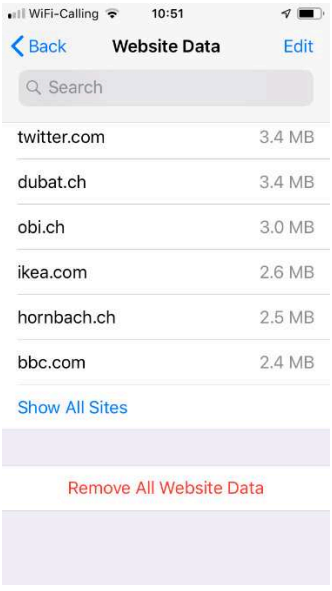
Il existe aussi des "add-ons" pour les navigateurs, des outils ajoutés qui aident à gérer les publicités, les cookies, ou plus généralement les "trackers".

On entend parler depuis quelque temps d'un changement de paradigme, où les navigateurs n'utiliseraient plus des cookies pour la personnalisation et les publicités, mais plutôt des FLoC (Federated Learning of Cohorts), censés améliorer la préservation de la vie privée des utilisateurs. En bref, il n'y aurait plus de cookies que les serveurs web lisent pour identifier les utilisateurs, les profiler, et estimer leurs préférences. C'est plutôt le navigateur de l'utilisateur qui fait son profilage et l'envoie aux serveurs visités, pour pouvoir recevoir des publicités personnalisées. En principe, les serveurs n'auraient plus de cookies pour identifier et tracer les utilisateurs.

Cependant, utilisant les FLoC, les préférences envoyées par les navigateurs pourraient être assez riches, formant ainsi une combinaison unique qui servirait aussi à identifier l'utilisateur d'une manière univoque. Cela va dépendre alors du suivi des règles par les serveurs web.



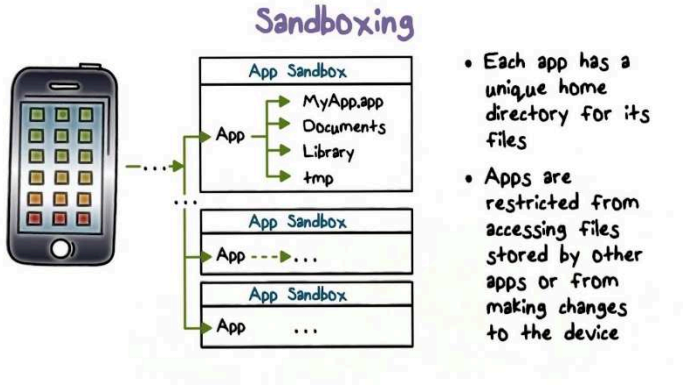




LES Ad-ID

Dans la section précédente nous avons vu comment, dans un navigateur web, plusieurs sites peuvent identifier un utilisateur, grâce aux cookies tiers. Dans les smartphones nous trouvons une problématique similaire, et une solution similaire aussi.

Dans les smartphones les applications sont mises dans des silos ("sandboxes") et ne sont pas supposées communiquer entre elles.



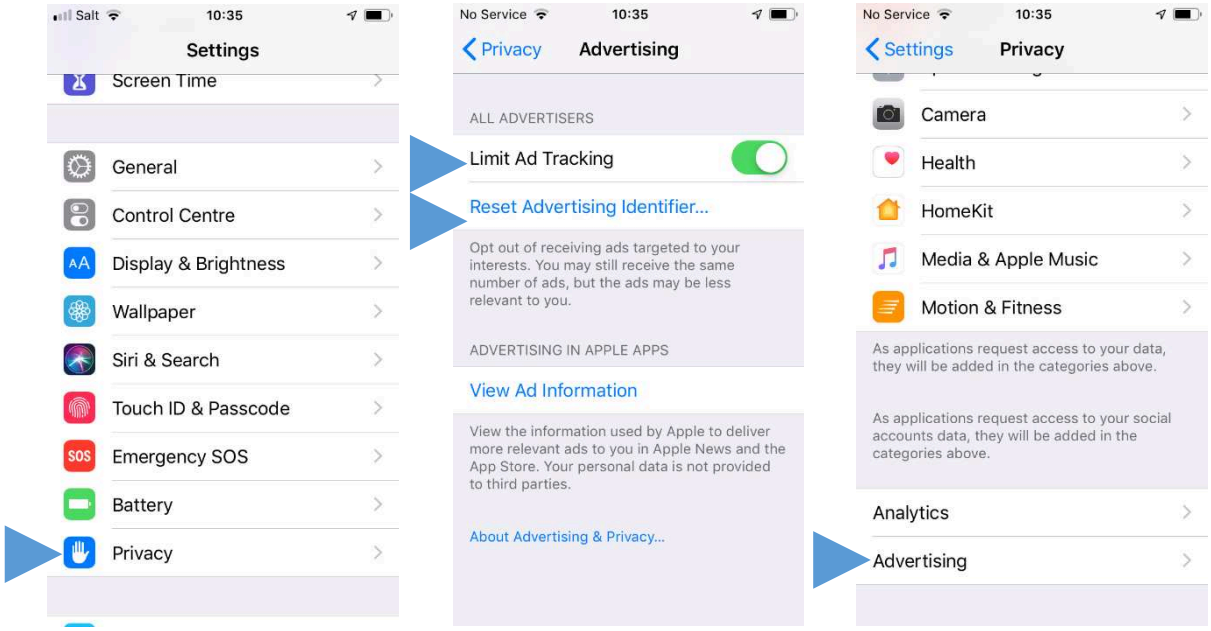
Par exemple, l'application "MonJournalPréféré" ne peut pas communiquer avec "MonTransportPréféré" pour optimiser le contenu à montrer à l'utilisateur. Les applications, et leurs serveurs correspondants, n'ont a priori pas de moyen de savoir qu'ils servent le même utilisateur. "MonJournalPréféré" et son serveur peuvent optimiser et profiler l'utilisateurX, pendant que "MonTransportPréféré" et son serveur profilent et optimisent le contenu pour l'utilisateurY. Si un jour l'entreprise derrière "MonJournalPréféré" veut collaborer avec celle de "MonTransportPréféré", pour optimiser leurs contenus ou pour servir des publicités personnalisées par exemple, elles n'ont pas moyen de savoir que l'utilisateurX est l'utilisateurY (ce qui était la problématique des sites web et le navigateur de l'utilisateur, s'il n'y avait pas les cookies tiers).

C'est là que le Ad-ID (Advertising ID) entre en jeu. Les smartphones mettent un identifiant unique par appareil, à disposition des différentes applications. Ça pouvait être le numéro de téléphone de l'utilisateur ? Ou pourquoi pas l'adresse MAC ? Ils sont uniques par smartphone, mais ils ne sont pas rendus accessibles aux applications. C'est seulement l'Ad-ID qui l'est, et c'est grâce à cet Ad-ID qu'un serveur de publicité utilisé par "MonJournalPréféré" et "MonTransportPréféré" peut identifier que l'utilisateurX est le même utilisateurY, et ainsi le profiler avec plus de précision. C'était exactement la même utilité des cookies tiers dans les navigateurs.

Comment contrôler les Ad-ID ?

L'utilisateur peut soit empêcher l'utilisation des Ad-ID, soit les réinitialiser. Dans le premier cas un serveur de publicité aurait du mal à identifier que l'utilisateurX est aussi l'utilisateurY. Les publicités vont toujours être affichées dans les applications, mais la personnalisation serait moins réussie. A noter que l'utilité des Ad-ID n'est pas limitée aux publicités. Les Ad-ID peuvent être utilisés pour l'identification ou le profilage en arrière-plan.

Quand l'utilisateur réinitialise son Ad-ID, c'est un nouvel identifiant qui est généré (utilisateurZ), qui va être utilisé par la suite.



Il faut préciser que l'Ad-ID n'est pas le seul moyen pour que deux applications (et leurs serveurs respectifs) identifient un même utilisateur. Les smartphones ne mettent que l'Ad-ID à disposition des applications pour l'identification. Cependant, l'utilisateur lui-même pourrait s'identifier auprès de plusieurs applications/services d'une manière univoque, par exemple en utilisant le même goldorak1971@gmail.com pour plusieurs comptes, ou encore plus précisément en révélant son vrai nom et prénom (ex. jean-luc.panchaud@bluewin.ch).

LES "NOMS D'UTILISATEURS", LES ADRESSES MAILS, ET LES CERTIFICATS

Lorsque j'ouvre un compte auprès d'un fournisseur de service, ou pour utiliser son application, je pourrais choisir un nom d'utilisateur tel que goldorak1971. Selon le type de service, goldorak1971 pourrait être un nom acceptable, peu importe la vraie identité de l'utilisateur.

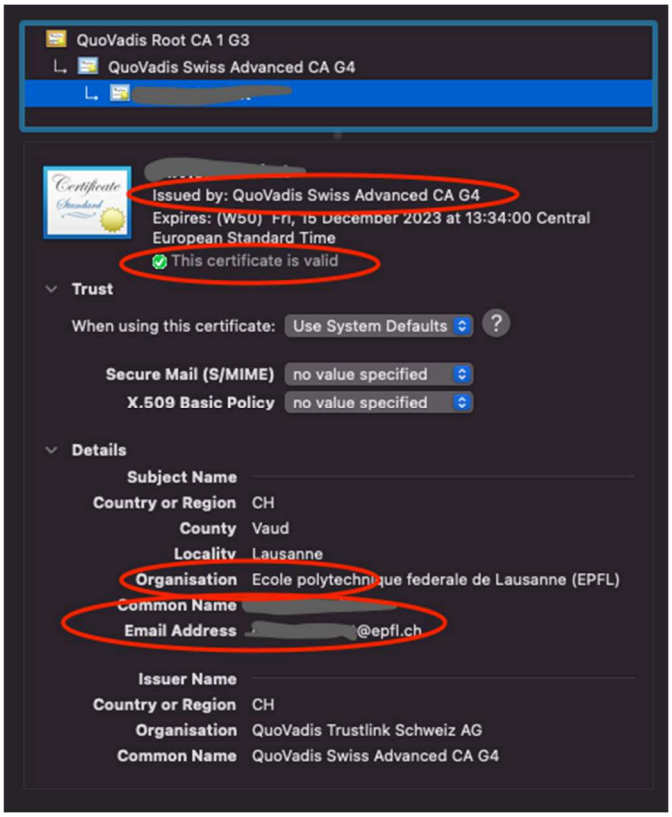
Dans d'autres situations, il convient de vérifier si celui qui crée le compte sur un site sensible pour jean-luc.panchaud@supermail.biz est vraiment le propriétaire du compte correspondant. Un mail de vérification est alors envoyé à partir duquel il faut confirmer la création du compte. Cependant, ceci ne confirme d'aucune manière que jean-luc.panchaud@supermail.biz appartient au "vrai" Jean-Luc Panchaud. Ça pourrait bien être à sa voisine, Danielle Duvoisin, que l'adresse mail appartient. C'est pour cela que certains services exigent encore un autre niveau d'authentification, tel que l'envoi d'une copie de la pièce d'identité, une vidéo avec la personne montrant la pièce d'identité, ou une authentification au guichet pour être plus sûr, etc.

Cependant, ces moyens d'identification sont encombrants, et exigent souvent des interventions humaines. C'est pour pallier ce problème que nous avons les "identités numériques" : je crée une identité, et je la fais certifier par un organisme "de confiance". Par la suite, je pourrais utiliser cette identité, et son authenticité pourrait être vérifiée automatiquement. Nous abordons le sujet de cette certification un peu plus tard.

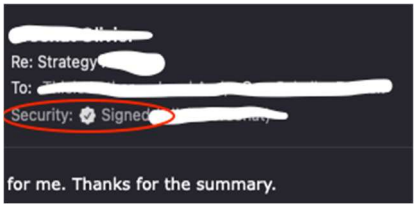

"On the Internet, nobody knows you're a dog."

Ce besoin "de confiance" est aussi nécessaire dans les échanges d'emails : Si je reçois un mail du premier ministre nigérien me proposant un transfert de quelques millions de dollars, rien ne me garantit que ce soit vraiment le premier ministre nigérien. N'importe qui peut m'envoyer un mail de la part de president@whitehouse.us, au nom de mon chef d'entreprise ou au nom du support technique qui m'envoie un certain lien pour mettre à jour mon mot de passe. Cette problématique n'était pas courante durant les années 70-80, mais le besoin de "certification" est devenu de plus en plus important plus tard.

La pratique de certification des mails est courante dans les environnements professionnels (et moins fréquente chez les particuliers) : L'adresse mail d'un employé A est "signée" numériquement par l'entreprise E. A son tour, l'identité de l'entreprise E est signée par une autre entreprise de certification CA (Certification Authority), mondialement connue.



A la réception d'un mail de l'employé A, le destinataire pourrait ne pas connaître l'employé A, ni son entreprise E, et de toute façon un courriel affiché de president@whitehouse.us ne veut rien dire. C'est alors l'application des mails (ex. Outlook) qui s'occupe des vérifications : si la CA a bien signé l'identité de l'Entreprise E, et par suite si celle-ci a bien signé l'identité de l'employé A. Selon le résultat des vérifications, l'identité de l'employé A s'affiche comme certifiée, ou qu'il y a un problème de vérification.



Attention : si je reçois un mail authentifié de mon chef, ça pourrait bien être envoyé depuis de son vrai compte mail, et que son identité est vérifiée/certifiée. Reste à voir si son compte ou le serveur mail ont été compromis ("hackés"). C'est le même problème quand je reçois un appel d'un numéro que je connais bien (ex. conseiller de la banque), mais qui me demande des informations susceptibles d'être mal utilisées. Il faut suivre la règle "Hang Up, Look Up & Call Back" (raccrocher, chercher, et rappeler) pour éviter que quelqu'un ait compromis le compte mail du chef, ou même le numéro de téléphone du conseiller de la banque (oui, c'est possible).

Identités des pages web

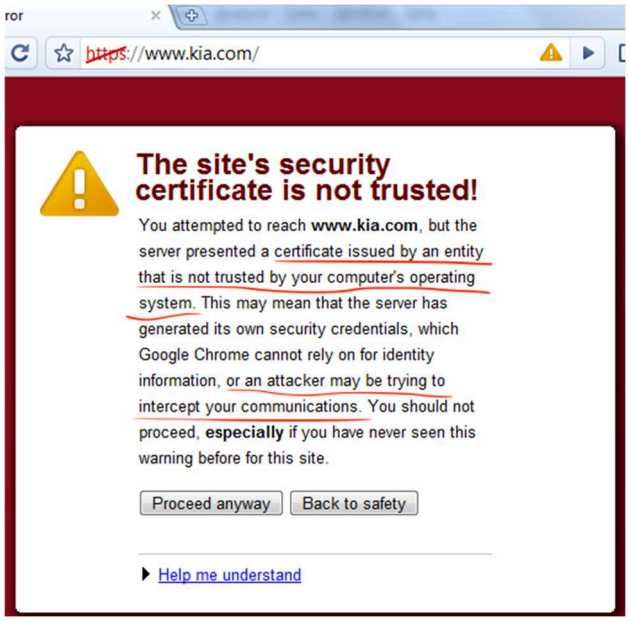
Le problème d'authentification des identités n'est pas limité aux personnes et à leurs adresses mail. Le même problème se manifeste lorsqu'un utilisateur visite une page web. L'identité de cette page web pourrait être compromise par n'importe qui se trouve entre l'utilisateur et le serveur de la page web ("man-in-the-middle"). Ainsi, le voisin propriétaire du WiFi où je me connecte, ou le restaurant qui m'offre la connexion WiFi, le fournisseur d'accès Internet, ou n'importe quel opérateur dans le monde qui pourrait se trouver sur le chemin entre mon ordinateur et la page web visitée, pourraient falsifier la page web que j'aimerais visiter pour intercepter la connexion, écouter la communication et/ou voler les mots de passe. Je pourrais faire une courte consultation de mon compte bancaire (sur la page affichée par le "man-in-the-middle"), pendant qu'il fait des transactions depuis mon compte sur la vraie page web de la banque.

Heureusement que ce problème se résout facilement à l'aide des certificats (quand j'utilise https://... Au lieu de http://...) :

- Mon navigateur vérifie la validité des signatures qui authentifient la page web visitée
- La communication est cryptée depuis mon navigateur jusqu'au serveur de la page web visitée, et non pas jusqu'au man-in-the-middle.

C'est ce que m'indique le petit cadenas à côté de l'adresse de la page web visitée

La plupart des “grandes” pages web utilisent cette sécurisation. Cependant, c’est à l’utilisateur de vérifier le cadenas et ne pas négliger les alertes que le navigateur lui montre.



LES IMEI ET IMSI

Les opérateurs téléphoniques se basent sur d’autres identifiants pour pouvoir identifier l’appareil mobile et/ou l’utilisateur. Chaque téléphone (smartphone ou pas) a son IMEI (“International Mobile Equipment Identity”) et chaque carte SIM son IMSI (“International Mobile Subscriber Identity”). Indépendamment des applications sur le smartphone, de l’Ad-ID ou des adresses MAC ou IP, l’opérateur téléphonique doit pouvoir identifier :

- L’appareil, pour la communication entre l’antenne cellulaire et le téléphone, grâce à l’IMEI
- Et son utilisateur à des fins de facturation, ou de traçage des communications qui est une obligation légale, grâce à l’IMSI.

Bien entendu, l’utilisateur n’a pas de contrôle sur ces identifiants, ni pour les changer ni pour les cacher. Même si l’utilisateur n’utilise pas son téléphone pour communiquer, son appareil se connecte et s’identifie avec l’opérateur, pour pouvoir recevoir les appels entrants par exemple.

Les utilisations « alternatives »

Grâce à ces identifiants, les opérateurs téléphoniques connectent les utilisateurs et leur font d'éventuelles facturations. Ces identifiants leur permettent aussi de localiser les utilisateurs et de les tracer pour diverses autres finalités : statistiques, planifications des villes, ou pour aider la police à localiser des personnes perdues, etc.

Il faut noter que la précision de ces localisations dépend de plusieurs facteurs. Un appareil téléphonique se connecte à une antenne de l’opérateur. Dans les zones rurales, une antenne sert à servir une région large de plusieurs kilomètres autour de l’antenne. En ville, une antenne sert une région large de quelques dizaines ou centaines de mètres autour de l’antenne. En combinant les informations reçues par plusieurs antennes (technique appelée « trilatération »), les forces des signaux, etc., l’opérateur pourrait affiner la localisation d’un téléphone mobile avec une précision allant de quelques dizaines de mètres à quelques kilomètres.

L'IDENTIFICATION AU "NIVEAU PHYSIQUE"

Quand deux appareils sans-fil communiquent entre eux, ils utilisent des fréquences radio, et s'ils sont digitaux alors ils codent la communication utilisant des 0 et des 1, des niveaux hauts ou bas qui s'alternent des millions de fois par seconde (les bits/s).

FICHE INFO DU PPDT

Les identifiants numériques et leur utilisation

Deux appareils de la même marque et du même modèle pourraient avoir des composants provenant de différents fournisseurs. Ces composants pourraient avoir des caractéristiques différentes (telle qu'une légère différence des fréquences, ou de la vitesse de passage de 0 à 1, ou de 1 à 0), qu'un écouteur bien équipé pourrait distinguer. Même pour les composants d'un même fournisseur, il est possible d'identifier certaines différences au niveau radio.

Comme chez les humains où un écouteur pourrait distinguer ou reconnaitre les personnes à partir de leurs voix ou leurs accents, les caractéristiques radio des appareils sans-fil pourraient être utilisées comme un niveau supplémentaire de sécurité/authentification. Cependant, ceci implique le déploiement de matériels d'écoute de haute précision, relativement coûteux, et qui se situent à portée radio des appareils communiquant, mais pas à travers l'Internet.

LES "COMBINAISONS IDENTIFIANTES"

Quand un utilisateur visite un site web, il n'y a pas que l'adresse IP et les cookies ou le nom d'utilisateur qui sont transmis. Plusieurs autres paramètres de son appareil et navigateur sont transmis au serveur, tels que :

- La marque d'appareil utilisé (ex. Apple, HP etc.) et le modèle
- Le système d'exploitation (ex. Windows, macOS, Linux etc.) et sa version
- Le navigateur utilisé (ex. Safari, Firefox, Chrome) et sa version
- La langue utilisée
- La taille de l'écran.

Ces détails sont transmis par les navigateurs pour que le serveur visité puisse adapter son contenu et format aux préférences de l'utilisateur. Cependant, la combinaison de ceux-ci pourrait être utilisée par les serveurs web comme identifiant unique de l'utilisateur, afin de pouvoir le profiler.

Comme dans le monde physique, une combinaison de critères même très génériques devient identifiante quand le nombre de critères augmente (ex. le brun, cheveux frisés, avec des lunettes, qui a une VW Polo, habite à Lausanne, et travaille chez Nestlé... il n'y en a pas plusieurs !).

Il y a des centaines de millions d'appareils HP dans le monde, des centaines de millions de Windows, autant de navigateurs Firefox utilisés, d'utilisateurs francophones, etc. Mais le nombre de machines qui utilisent une combinaison spécifique (ex. HP Pavillion, Windows 7, Firefox 6.4, Français, 1400x1050) est drastiquement réduit, potentiellement pouvant identifier certains utilisateurs d'une manière univoque. C'est-ce qu'on appelle "device fingerprinting" (empreinte digitale de l'appareil), une technique utilisée par les serveurs :

- Pour tracer les utilisateurs, en plus qu'avec les autres identifiants (ex. cookies)
- Comme mesure de sécurité supplémentaire, quand le serveur constate un changement de "profil" (en envoyant à l'utilisateur un mail "Alerte de sécurité : Nouvelle connexion d'un nouvel appareil").

L'utilisateur souhaitant éviter ce type de traçage pourrait utiliser certains navigateurs ou des "plug-ins" qui partagent des préférences dont la combinaison reste générique, ou ne transmettent pas de préférence du tout.

CONCLUSIONS

Afin de pouvoir communiquer, les ordinateurs, smartphones et serveurs utilisent divers identifiants dans diverses étapes de la communication (ex. locale, globale etc.). Ces identifiants sont aussi utilisés à des fins de traçage des utilisateurs. D'autres identifiants supplémentaires ont été introduits pour permettre à différents serveurs d'identifier un même utilisateur, pour affiner leurs profilages, souvent à des fins publicitaires et économiques.

Une identification sans lien avec la personne physique pourrait suffire pour une grande partie des cas, principalement pour les serveurs qui visent à affiner les profils des utilisateurs et leur mettre des publicités ou des contenus personnalisés, peu importe les vraies identités des personnes physiques. Nous avons décrit ces méthodes d'identification, les moyens d'éviter le traçage disponibles pour les utilisateurs, ainsi que les contre-mesures utilisées par les serveurs afin de garder le traçage possible. Le résultat dépend des moyens, des efforts, et de l'expertise que l'un ou l'autre est prêt à investir pour tracer, ou se cacher.

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

FICHE INFO DU PPDT

Les identifiants numériques et leur utilisation

Le passage des identifiants aux vraies identités des personnes physiques est une possibilité qui dépend de la richesse des identifiants et des données collectées, de la disponibilité d'autres données externes, de la motivation et des profits pour les opérateurs pour s'y investir, ainsi que des restrictions légales.

Dans les cas où la confiance est indispensable, il est nécessaire de pouvoir prouver son identité en utilisant des certificats digitaux, que ce soit pour vérifier l'identité de l'utilisateur ou celle du serveur. Nous avons brièvement décrit ces techniques, ainsi que les bonnes pratiques à suivre.

QUELQUES LECTURES SUPPLEMENTAIRES

"Examples of Data Points Used in Profiling", Privacy International

"Data and Goliath", Bruce Schneier

AUTEUR

Dr. Imad Aad, Centre pour la Confiance Numérique (https://www.c4dt.org/), EPFL.

PPDT – 21.12.2021