MISC
Multi-System & Internet Security Cookbook
100 % SÉCURITÉ INFORMATIQUE
L 19018 - 81 - F: 8,90 € - RD
N° 81 SEPT./OCT. 2015
France METRO : 8,90 € - CH : 15 CHF - BE/PORT CONT : 9,90 € - DOM TOM : 9,50 € - CAN : 16 $ cad - Maroc : 110 MAD - Tunisie 19 TND

ÉDITO Summer Leak

L'été 2015 restera marqué par deux leaks massifs de données professionnelles et personnelles très largement relayés par la presse grand public. Deux événements encore relativement inhabituels, mais qui risquent de se reproduire et vont certainement changer profondément la perception des utilisateurs sur la sécurité de leurs données privées stockées en ligne.

La première divulgation ayant dérangé le petit monde de la sécurité en pleine torpeur estivale est celle du piratage de Hacking Team. Passés les quelques détails croustillants tels que l'usage de mots de passe assez faibles, la présence d'une backdoor dans leur backdoor ou des logiciels crackés dans leurs archives, d'autres informations relèvent moins de l'anecdote et méritent un peu plus que l'on s'y attarde.

Tout d'abord, il apparaît que l'informatique offensive se banalise et qu'en la matière, les agences gouvernementales ont massivement recours à des officines privées. La liste des clients d'Hacking Team est très éclairante et démontre que tous les continents et régimes politiques ont recours à ce nouveau mercenariat. La guerre s'est privatisée (lire l'excellent Pukhtu de DOA sur les sous-traitants de la CIA en Afghanistan aux éditions Gallimard) et la cyberguerre n'est pas en reste.

Le second point c'est que l'analyse des données d'Hacking Team montre que la société utilisait, entre autres, un zero day d'Adobe Flash Player comme vecteur d'infection. On pourrait arguer que même si HTML5 le remplace avantageusement, les développeurs web sont des fainéants et que sans le plugin honni, tous les sites, ou presque sont cassés. Eh bien non, de Netflix à Pornhub en passant par YouTube et Dailymotion, des sites de Libération, du Figaro ou du Monde, les Dropbox, Evernote, Twitter ou Facebook, tout fonctionne parfaitement sans Flash Player. À moins d'être un joueur en ligne invétéré ou <Troll> de devoir administrer VMWare vSphere via vCenter </Troll>, il est possible de naviguer pendant des semaines sans Flash sans même s'en apercevoir. Au regard du lourd passif en matière de vulnérabilités de Flash, étendre à ce point la surface d'attaque d'un terminal pour un outil qui ne sert à rien pour la plupart des utilisateurs est vraiment désespérant.

Le second leak très largement relayé par la presse grand public au point d'apparaître sur les unes des quotidiens est celui du site de rencontres Ashley Madison. Si la divulgation des données d'Hacking Team ne concerne que les documents internes de la société, celle d'Ashley Madison jette en pâture les données de 33 millions de « comptes » utilisateurs : boites mails, mots de passe, fiches de présentation, éléments de transactions bancaires… La nature du site quelque peu sulfureuse a vite attiré des hordes de curieux se répandant ad nauseam sur les réseaux sociaux sur le profil présumé des abonnés au site de rencontres. Évidemment, la présence supposée d'utilisateurs dans les administrations françaises ajoute au scandale et s'il s'agit de femmes, visiblement peu nombreuses sur ce site, c'est encore mieux.

Ce voyeurisme est déjà particulièrement malsain, mais s'ajoute à celui-ci la fiabilité sujette à caution des bases de données. Peut-être afin de gonfler sa base d'utilisateurs, Ashley Madison a dû considérer que vérifier la validité des boites mails était une sécurité superflue. Il est ainsi possible d'inscrire sur le site n'importe quelle boite mail et aucun lien de confirmation n'est envoyé pour vérifier que la demande vient du propriétaire de la boite. Pour peu que le message vous informant que vous avez été inscrit se retrouve dans votre répertoire spam vous pouvez être référencé sur ce site sans le savoir. RSSI à l'éducation nationale j'ai été plusieurs fois saisi par des collègues harcelées qui s'étaient faites inscrire à leur insu sur de multiples sites de rencontres (car visiblement Ashley Madison n'est pas le seul site où l'on ne vérifie pas l'authenticité des boites mails) avec fiches de présentations ordurières, leur adresse, des photos volées prises avec un smartphone. Le lendemain du leak, un courageux justicier anonyme de la moralité sur Twitter postait ceci lorsque je lui faisais remarquer la fiabilité toute relative des données: « En tous cas là j'ai le compte d'une prof, y'a sa date de naissance complète, ville + beaucoup d'autres infos pas faciles à avoir » [sic].

Je ne sais pas si finalement le plus désespérant est la présence de Flash sur la plupart des navigateurs...

Bonne lecture !

Cedric Foll / cedric@miscmag.com / @follc

SOMMAIRE

www.miscmag.com

MISC est édité par **Les Éditions Diamond**
10, Place de la Cathédrale
68000 Colmar, France
Tél. : 03 67 10 00 20 - **Fax :** 03 67 10 00 21
E-mail : cial@ed-diamond.com
Service commercial : abo@ed-diamond.com
Sites : www.miscmag.com
www.ed-diamond.com

IMPRIMÉ en Allemagne - PRINTED in Germany
Dépôt légal : A parution
N° ISSN : 1631-9036
Commission Paritaire : K 81190
Périodicité : Bimestrielle
Prix de vente : 8,90 Euros

Directeur de publication : Arnaud Metzler
Chef des rédactions : Denis Bodor
Rédacteur en chef : Cédric Foll
Secrétaire de rédaction : Aline Hof
Conception graphique : Kathrin Scali
Responsable publicité :
Black Mouse Communication Tél. : 03 67 10 00 27
Service abonnement : Tél. : 03 67 10 00 20
Illustrations : www.fotolia.com
Impression : pva, Druck und Medien-Dienstleistungen GmbH, Landau, Allemagne
Distribution France : (uniquement pour les dépositaires de presse)
MLP Réassort :
Plate-forme de Saint-Barthélemy-d'Anjou. Tél. : 02 41 27 53 12
Plate-forme de Saint-Quentin-Fallavier. Tél. : 04 74 82 63 04
Service des ventes : Abomarque : 09 53 15 21 77

Membre April www.april.org

LES ÉDITIONS DIAMOND

Charte de MISC

MISC est un magazine consacré à la sécurité informatique sous tous ses aspects (comme le système, le réseau ou encore la programmation) et où les perspectives techniques et scientifiques occupent une place prépondérante. Toutefois, les questions connexes (modalités juridiques, menaces informationnelles) sont également considérées, ce qui fait de MISC une revue capable d'appréhender la complexité croissante des systèmes d'information, et les problèmes de sécurité qui l'accompagnent. MISC vise un large public de personnes souhaitant élargir ses connaissances en se tenant informées des dernières techniques et des outils utilisés afin de mettre en place une défense adéquate.
MISC propose des articles complets et pédagogiques afin d'anticiper au mieux les risques liés au piratage et les solutions pour y remédier, présentant pour cela des techniques offensives autant que défensives, leurs avantages et leurs limites, des facettes indissociables pour considérer tous les enjeux de la sécurité informatique.

UN PETIT PLEIN DE VIE PRIVÉE POUR LA RENTRÉE S'IL VOUS PLAÎT !

Fini les vacances, retour aux choses sérieuses avec une surprise attendue : un dossier proposé par la Commission nationale de l'informatique et des libertés.

Au cœur du sujet, les traceurs. Les cookies restent certainement la technique la plus simple et la plus utilisée pour stocker des données spécifiques à l'utilisateur (authentification, session, panier d'achat...). Mais le sujet est plus vaste: il concerne l'ensemble des techniques qu'un site web peut utiliser pour tracer un utilisateur et lui créer profil. Ce stockage controversé empiète sur la vie privée de l'utilisateur : la CNIL a donc établi des règles.

Certains sites web se sont adaptés pour respecter cette paix belliqueuse et suivre les règles, d'autres passent outre grâce à des techniques plus avancées. En effet, bloquer complètement l'utilisation des traceurs remettrait profondément en question le modèle économique de l'internet que l'on connaît aujourd'hui, et personne ne le souhaite. D'ailleurs, vous avez désinstallé Adblock, n'est-ce pas ?

Stéphane et Benjamin nous donnent une première introduction aux principaux traceurs que l'on retrouve sur le web. Cinq méthodes sont présentées pour suivre l'utilisateur sur plusieurs sites web, qu'ils soient parcourus à partir d'un seul ou de plusieurs périphériques. Ils nous apprennent comment entrevoir l'invisible, dévoilent les outils et les techniques qu'ils utilisent, et développent plusieurs points permettant d'assurer notre vie privée.

Clémence étudie ensuite l'obscure clarté des règles qui encadrent ce genre de traceurs. Elle revient en détail sur les nouvelles dispositions mises en place pour prévenir leur utilisation. Plus qu'un texte juridique, des exemples concrets nous mettent dans le feu de l'action et rivalisent avec les meilleurs de Lencioni. On y trouve des réponses à des questions que l'on se pose souvent, comme la célèbre « mais ce bandeau sur tous les sites, il bloque les cookies ? ».

Et si jamais vous n'aviez pas ce bandeau sur votre site web, ou que vous n'étiez pas sûr de l'avoir bien pris en compte, Vincent explique comment faire dans l'article qui suit. Un remède explosif à base de CookieCuttr, de tarte au citron, de CSP, de Ghostery et de DNT. Plus d'excuses.

Le premier article nous expliquait les dangers liés aux espaces mis aux enchères par plusieurs régies publicitaires, augmentant le risque de contenus dangereux. Claude examine dans son article ce phénomène d'enchères en temps réel. Il nous montre à quel point nous sommes traqués lors de nos sessions web, aussi bien par des méthodes basiques que par des techniques complexes. Flippant.

Finalement, Benoît et Pierre se focalisent sur une des techniques de traçage introduite précédemment : la prise d'empreinte (en général, pour le digital voir [1]). C'est un procédé novateur qui consiste à trouver des canaux auxiliaires et ainsi permettre d'identifier de façon très précise un utilisateur ou un périphérique. On reconnaît tout de suite le parallèle avec l'utilisation des canaux cachés sur du matériel : utiliser des composants à l'apparence anodine pour corréler des données et extraire une information cachée.

On comprend alors beaucoup mieux les nombreuses problématiques que doit traiter la CNIL. En revenant aux cookies, on peut quand même se dire que c'est sympa de pouvoir directement « liker » une page et qu'on en a vraiment besoin. Allons au-delà ! Imaginons plus loin que la partie financière. Puisque les profils servent à prévoir les comportements des internautes, ce type d'information, utilisé à bon escient, pourrait permettre d'instancier le nombre idéal de machines virtuelles (VM) côté serveurs pour absorber la charge tout en diminuant l'impact environnemental. Grâce aux traceurs, nous pourrions alors collaborer avec l'ordonnanceur de ressources distribuées (DRS) pour optimiser l'algorithme de déploiement des VM...

Je vous laisse sur ce vide envahissant.

Aurelien Wailly

[1] https://www.blackhat.com/docs/us-15/materials/us-15-Zhang-Fingerprints-On-Mobile-Devices-Abusing-And-Leaking-wp.pdf

AU SOMMAIRE DE CE DOSSIER :

DÉTECTER ET ANALYSER LES COOKIES ET AUTRES TRACEURS

Stéphane Labarthe – slabarthe@cnil.fr – misc@labarthe.es
Benjamin Vialle – bvialle@cnil.fr – misc@vialle.io
Auditeurs des systèmes d'information à la CNIL *

mots-clés : CNIL / COOKIES / TRACEURS / PIXELS INVISIBLES / HTTP / COOKIE MATCHING / CROSS CANAL / CROSS DEVICE / HTTPONLY / HSTS

Le traçage de la navigation des internautes pour alimenter les bases de données du monde de la publicité ciblée s'opère par différentes techniques qui sont présentées dans la première partie de cet article : pixels invisibles, cookies HTTP, cookies flash ou stockage web local. La deuxième partie décrit au lecteur des outils simples et une méthodologie qui lui permettront de le détecter et de l'analyser lui-même sur ses sites internet préférés. Elle explique également comment les acteurs publicitaires parviennent aujourd'hui à rompre le présumé anonymat des internautes non connectés pour leur adresser des publicités personnalisées par voie électronique, téléphonique ou postale (« cross canal ») ou sur leurs autres terminaux (« cross device »). La partie 3 de cet article propose des moyens de protection contre ces techniques de traçage. Enfin, en guise d'appendice, quelques aspects liés à la sécurité des cookies sont abordés. Cet article s'appuie en partie sur l'expérience du service des contrôles de la CNIL sur ces sujets.

1 Principaux traceurs utilisés dans la publicité ciblée

Il existe plusieurs manières de tracer un internaute au cours de sa navigation sur le web. Les méthodes les plus utilisées ont été rencontrées par la CNIL lors de contrôles effectués ces dernières années auprès d'acteurs majeurs de l'Internet, que ce soit des sites éditeurs, annonceurs ou des sociétés spécialisées dans la publicité ciblée.

1.1 Pixel invisible

La technique dite du « pixel invisible » consiste à appeler une image, hébergée sur un serveur tiers, à la taille de 0 ou de 1 pixel, donc invisible de l'internaute. Lors de cet appel, des informations sont transmises en paramètre de la requête HTTP/GET correspondante.

Exemples de tag [portion de code HTML ou JavaScript] incluant un pixel invisible :

```
<img src="http://serveur-de-tracking.com?&img=d690a7357b.png"
style="position:absolute; visibility:hidden">

<img src="http://serveur-de-tracking.com?&img=d690a7357b.png"
style="display:none">

<img src="http://serveur-de-tracking.com?pixel.png?APN=818677924104
7110843&pid=CAESEAZY8yBmmTLHS_0bBfHimto width="0" height="0">
```

Dans ces exemples, en ouvrant la page web ou le courrier électronique qui contiendrait un de ces tags HTML, le navigateur de l'internaute effectue une requête de type GET vers le serveur-de-tracking.com. Il transmet alors un paramètre : un nom d'image (**d690a7357b.png**) ou un contenu de variable (pid=**CAESEAZY8yBmmTLHS_0bBfHimto**). **Ce paramètre unique**, permet alors de

« reconnaître » l'internaute et donc de savoir qu'il a consulté telle partie d'une page internet ou qu'il a ouvert tel e-mail publicitaire.

1.1.1 Pixel invisible et dépôt de cookie

En réponse, le serveur web peut déposer un ou plusieurs cookies via l'instruction (en-tête) « set-cookie » du protocole HTTP. Dès lors, à chaque fois que l'internaute se rendra sur un site web contenant un tag redirigeant vers le domaine **serveur-de-tracking.com**, il sera pisté. Nous verrons, dans la partie *cross-canal*, comment les sites visités obtiennent alors l'adresse électronique de l'internaute, à des fins de sollicitation commerciale.

Cette technique requiert d'utiliser des cookies.

1.2 Cookies HTTP

Le cookie HTTP reste la technique la plus utilisée pour tracer les internautes.

Un cookie est défini dans la RFC 2965 **[RFC6265]** comme étant une suite d'informations envoyée par un serveur web à un client HTTP. Ce dernier retourne le contenu du cookie lors de chaque interrogation d'un serveur web associé au même nom de domaine (sous certaines conditions).

Les cookies ont été initialement prévus et utilisés à des fins techniques ou fonctionnelles. Ils permettent par exemple au protocole HTTP (nativement « stateless ») d'avoir la possibilité de gérer les sessions et de permettre certaines fonctionnalités. Cependant, leur usage initial a progressivement été complété par une utilisation à des fins de traçage publicitaire, car les cookies peuvent aussi stocker des informations relatives à la navigation de l'internaute ou un identifiant de « tracking ».

Le cookie se présente aujourd'hui sous différentes formes en fonction des navigateurs (cf. article de Cyrille Aubergier dans le *MISC* précédent) :

- Internet Explorer enregistre chaque cookie dans un fichier texte différent ;
- Mozilla Firefox et Google Chrome enregistrent les cookies dans une base de données SQLite ;
- Opera enregistre tous ses cookies dans un seul fichier et le chiffre — ce que ne font pas les trois autres.

1.3 Local Shared Objects (Cookies Flash)

Les *local shared objects* **[LSO]**, connus également sous le nom de « cookies flash », sont des données enregistrées sur l'ordinateur de l'internaute, lors de l'exécution d'applicatifs Flash (Adobe Flash Player et Macromedia Flash MX Player).

Par défaut, l'applicatif Flash peut écrire sur le terminal de l'utilisateur sans avoir requis préalablement le consentement de l'utilisateur.

Où trouver les cookies Flash sur mon terminal :

```
Windows :
%APPDATA%\Macromedia\Flash Player\#SharedObjects\
%APPDATA%\Macromedia\Flash Player\macromedia.com\
Mac OS X :
~/Library/Preferences/Macromedia/Flash Player/#SharedObjects/
~/Library/Preferences/Macromedia/Flash Player/macromedia.com/
Linux/Unix :
~/.macromedia/Flash_Player/#SharedObjects/
~/.macromedia/Flash_Player/macromedia.com/
Linux/Unix (utilisant le logiciel libre Gnash, en remplacement de
Flash Player) :
~/.gnash/SharedObjects/
Dans le cas de chrome (Flash Player est intégré via Pepper Flash
(PPAPI) :
Windows : %localappdata%\Google\Chrome\User Data\Default\Pepper
Data\Shockwave Flash\WritableRoot\#SharedObjects
Mac OS X : ~/Library/Application Support/Google/Chrome/Default/
Pepper Data/Shockwave Flash/WritableRoot/#SharedObjects/
Linux/Unix : ~/.config/google-chrome/Default/Pepper Data/Shockwave
Flash/WritableRoot/#SharedObjects/
```

Les cookies flash peuvent aussi être affichés dans un navigateur via l'URL suivante (qui permet aussi de régler les paramètres du FlashPlayer) : **http://www.macromedia.com/support/documentation/fr/flashplayer/help/settings_manager07.html**.

Le processus d'une application Flash lit directement ces cookies dans les répertoires où ils sont stockés.

À noter : Adobe Flash propose un « mode privé » où les cookies Flash ne sont pas sauvegardés. Celui-ci s'active dans les paramètres Flash lorsque l'on se trouve sur une page contenant un applet Flash.

1.4 Local Storage/Stockage web local (HTML5)

HTML5 apporte une nouveauté par rapport à ses prédécesseurs : la possibilité de stocker des données dans le navigateur sans utiliser de cookies. Cette technique, appelée « stockage web local » (ou *local storage*), permet de sauvegarder des volumes de données plus importants qu'avec les cookies (de l'ordre de 5 Mo à 10 Mo pour l'instant contre quelques kilo-octets pour les cookies).

Il existe deux types de stockage web local : le stockage local et le stockage de session, équivalent respectivement aux cookies persistants et aux cookies de session.

Par défaut, les informations contenues dans la base locale ne sont pas systématiquement renvoyées au serveur web à chaque requête. Elles sont cependant récupérables sur demande (par exemple via des instructions JavaScript).

Dans le cas de Mozilla Firefox, ces données sont stockées dans une base de données de type SQLite, nommée **webappsstore.sqlite**. Avec un logiciel

permettant de lire une base de données SQLite, il est possible de consulter ces cookies (par exemple, avec l'extension Firefox SQLite Manager).

À ce jour, peu de sites utilisent cette technique pour identifier les internautes.

1.5 Fingerprinting

Le *fingerprinting* consiste à récupérer une « empreinte » du navigateur de l'internaute afin de l'identifier lors de connexions ultérieures.

Le fingerprinting fait l'objet d'un article dédié dans ce dossier.

2 Observer le tracking

La méthodologie décrite dans cette partie vise à observer et analyser les traceurs de type « pixels invisibles » et « cookies HTTP » qui demeurent encore de loin les plus utilisés.

Pour les besoins de l'article, nous avons réalisé un parcours de navigation sur Internet comprenant d'abord l'inscription sur quatre sites web d'éditeurs ou d'annonceurs ; puis la navigation sur deux sites de e-commerce avec la mise au panier d'articles ; enfin, la navigation sur un site d'annonceur avec affichage de publicités ciblées. L'objectif est d'illustrer le fait que certains aspects du traçage opéré par les publicitaires au moyen de cookies HTTP peuvent être simplement observés. La méthodologie étant reproductible et basée sur des extensions du navigateur Mozilla Firefox, le lecteur est invité à se livrer à la même expérience.

Les cookies déposés durant ce parcours, au nombre de 505, ont été examinés et sauvegardés avec l'extension *Cookie Manager +* **[CookieManager +]**. Par ailleurs, les en-têtes HTTP des échanges entre notre navigateur et les serveurs web distants ont été capturés au moyen de l'extension *Live HTTP Headers* **[LiveHTTPHeaders]**. Cette capture permet par exemple de constater d'autres transmissions de données, connexes à celles réalisées par les cookies :

- les transmissions opérées en cas d'appel de pixels invisibles ;
- le chaînage des appels de serveurs publicitaires et les dépôts/lectures de cookies qui en résultent dans le cadre d'une enchère publicitaire (RTB, voir article dédié dans ce dossier) ;
- les partages de cookies (*Cookie Matching*).

2.1 « Voir l'invisible » : la visualisation graphique avec Cookie Viz

La CNIL met à disposition de tous, sous licence libre (GPLv3), un outil de visualisation qui identifie en temps réel les requêtes HTTP vers des serveurs tiers et les cookies déposés par ces serveurs. Des alternatives à cet outil existent, comme Lightbeam. L'avantage majeur de CookieViz sur les autres outils similaires est qu'il est capable de visualiser les cookies de l'ensemble des navigateurs.

Concrètement, CookieViz analyse les interactions entre votre ordinateur, votre navigateur et les sites et les serveurs distants contactés au cours de la navigation.

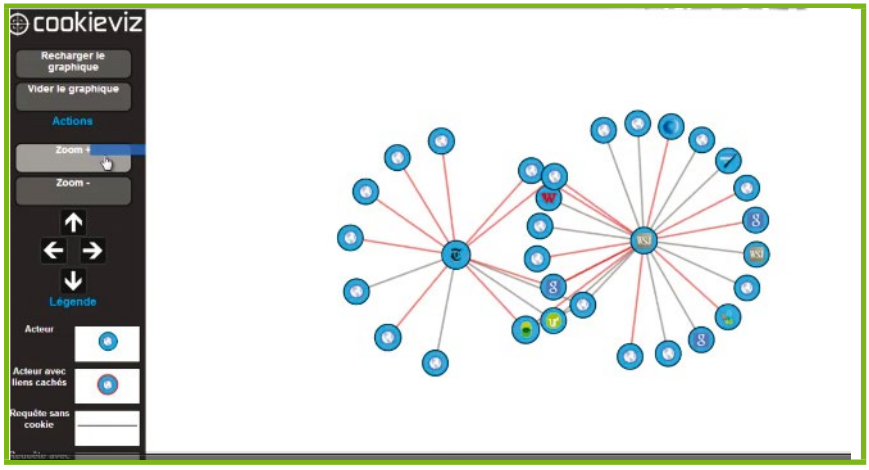

Fig 1 : CookieViz permet d'afficher les acteurs tiers contactés lors de la visite d'un site web.

2.2 L'analyse des caractéristiques des cookies avec Cookie Manager +

Pendant la vingtaine de minutes qu'a duré l'expérience, 505 cookies ont été déposés. La très grande majorité (378) sont des cookies tiers (« *third party* »), déposés par des serveurs web de domaines distincts du site web visité. Ils sont appelés suite à l'exécution de tags (images, code JavaScript) fournis par les prestataires et intégrés au code HTML de la page web. Une minorité (127) sont des cookies déposés et lus par les serveurs web des sites visités (« *first party* »). Cependant, certains de ces cookies peuvent avoir une finalité publicitaire (certains gros sites ont leur propre régie) ou être des « faux cookies internes », c'est-à-dire des cookies dont le domaine est celui du site, mais qui sont liés à des prestations réalisées par des tiers. C'est par exemple le cas des cookies de mesure d'audience « Google Analytics », présents sur la majorité des sites internet,

et dans notre expérience, sur les 7 sites visités. Ils ont pour noms : **_ga**, **_gat**, **_utma**, **_utmb**, **_utmc**, **_utmt**, **_utmv** et **_utmz**. Ces derniers sont lus par le serveur web du site, mais d'autres requêtes HTTP transmettent leur contenu (par passage en paramètre) aux serveurs de Google. Un blocage des cookies tiers est donc inefficace pour bloquer ce type de cookies.

Certains des cookies tiers qui ont été déposés ont des finalités liées aux boutons de partage sur les réseaux sociaux ou à la mesure d'audience. La majorité de ces cookies semble en revanche avoir des finalités publicitaires et certains contiennent des identifiants uniques du navigateur de l'internaute. Si les noms de domaine peuvent faire apparaître des noms d'acteurs publicitaires, les noms des cookies peuvent également être parlants.

Ainsi, une simple recherche des cookies dont le nom contient la chaîne « uid » (pour « *user id* ») en fait apparaître 53. En voici quelques-uns :

Name	Content
tuuid	c14added-4e0c-4d71-a6f4-9708cbddb5d8
uid	55959865148fd61d
uid	3473260044279437134
ADGRX_UID	9f771234-20f5-11e5-864d-1b9c040036e2
uids	@DDF>B@NKu@j`BDz{TMIQuFgH{|@DCF>>
uuid2	8186779241047110843
uid-bp-171	4527209332083766415
uid-bp-951	8186779241047110843
__cfduid	dfb309ba5b4dae0b7d673fe2120051a151435867904
uuid	67575228785204373
uid	7821592229675861147

Fig 2 : De nombreux cookies ont pour nom « ...uid... ».

Comme on peut le constater, ils contiennent en général des **numéros identifiants uniques** du navigateur de l'utilisateur en général au format décimal, hexadécimal ou chiffré. L'acteur publicitaire « reconnaît » ainsi le navigateur de l'utilisateur grâce à ce numéro, les données de navigation et d'analyse étant stockées en base, côté serveur.

Mais certaines régies peuvent faire le choix de stocker également une synthèse du profil publicitaire de l'internaute dans les cookies, côté client.

Ils sont en général dans des formats encodés :

```
Name:              evt
Path:              /
Content:           *1S%2fNl8ZLCh3bw%2fvduq5tHqo1uw%2biQjjzBVvt2
YCBDgp4%3d
```

D'une part, si les numéros d'identification de l'internaute sont propres à chaque acteur publicitaire, le cloisonnement des lectures et dépôts de cookies étant garanti par la « *same-origin-policy* » du navigateur, on peut remarquer que certains cookies de domaines différents utilisent le même identifiant. Comme évoqué dans l'article sur le RTB de ce dossier, c'est une conséquence du partage de cookies (« cookie matching » ou « cookie syncing ») sur laquelle nous reviendrons. À ce stade, ce point peut être facilement constaté en opérant un tri dans la colonne « Content » de *Cookie Manager+* :

Site	Name	Content
tradelab.fr	uuid2	8186779241047110843
adnxs.com	uuid2	8186779241047110843
ads.stickyadstv.com	uid-bp-951	8186779241047110843
rubiconproject.com	put_3876	8186779241047110843
rubiconproject.com	put_1986	8186779241047110843
sddan.com	map_nexus	8186779241047110843
burstnet.com	BI77335	8186779241047110843

Fig 3 : Illustration du «cookie matching».

D'autre part, l'examen **des durées de vie des cookies** peut être également un indice sur une finalité de traçage dans le temps. Dans notre expérience, 219 cookies (soit 43 %) ont une durée de vie supérieure ou égale à un an, certains pouvant avoir des durées très longues :

```
Name:              uids
Path:              /
Content:           @DDF>B@NKu@j`BDz{TMIQuFgH{|@DCF>>
Content raw:       @DDF>B@NKu@j`BDz{TMIQuFgH{|@DCF>>
Expires:           dim. 27 sept. 2037 02:00:13 CEST
```

2.3 Pour aller plus loin : l'analyse des en-têtes des requêtes HTTP

2.3.1 Dépôts et lectures de cookies de tracking dans les en-têtes HTTP

Les dépôts et lectures de cookies peuvent facilement être observés dans l'examen des en-têtes HTTP. Les dépôts et modifications de cookies apparaissent par le biais de l'instruction « Set-cookie » de la réponse du serveur web. Les cookies associés à un nom de domaine et déjà présents sur le navigateur sont systématiquement fournis (modulo des restrictions éventuelles sur les sous-domaines) dans le champ « Cookies » du navigateur. Par exemple, voici deux requêtes HTTP vers un serveur publicitaire :

```
http://ib.adnxs.com/getuidnb?http%3A%2F%2Fp.crm4d.
com%2Fsync%2Fappnexus%2Fs.gif%3Fuid%3D%24UID

GET /getuidnb?http%3A%2F%2Fp.crm4d.com%2Fsync%2Fappnexus%2Fs.
gif%3Fuid%3D%24UID HTTP/1.1
Host: ib.adnxs.com
User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:31.0) Gecko/20100101
Firefox/31.0 Iceweasel/31.6.0
[…]
Referer: http://www.monsite-ecommerce.com/
Connection: keep-alive


HTTP/1.1 302 Found
[…]
Expires: Sat, 15 Nov 2008 16:00:00 GMT
P3P: policyref="http://cdn.adnxs.com/w3c/policy/p3p.xml", CP="NOI
DSP COR ADM PSAo PSDo OURo SAMo UNRo OTRo BUS COM NAV DEM STA PRE"
X-XSS-Protection: 0
```

```
Location: http://ib.adnxs.com/bounce?%2Fgetuidnb%3Fhttp%253A%252F%2
52Fp.crm4d.com%252Fsync%252Fappnexus%252Fs.gif%253Fuid%253D%2524UID
Content-Type: text/html; charset=utf-8
Set-Cookie: sess=1; Path=/; Max-Age=86400; Expires=Fri, 03-Jul-2015
20:00:16 GMT; Domain=.adnxs.com; HttpOnly
Set-Cookie: uuid2=66106801265861589; Path=/; Max-Age=7776000;
Expires=Wed, 30-Sep-2015 20:00:16 GMT; Domain=.adnxs.com; HttpOnly
Date: Thu, 02 Jul 2015 20:00:16 GMT
Content-Length: 0
```

On peut observer qu'en réponse à cette requête, deux cookies sont déposés, l'un d'eux contenant un identifiant unique (uuid2). Lors de la requête suivante vers la régie, les cookies seront donc « servis » par le biais de l'en-tête « cookie » des requêtes HTTP :

```
http://ib.adnxs.com/getuid?http://mapping.nxtck.com/rtb/
um?n=msn&gid=$UID&uuid=e4647673-7aac-4a24-94f6-8103db8ada8d&cb=1123
713856&redir=http%3A%2F%2Fib.adnxs.com%2Fseg%3Fadd%3D209359%2526red
ir%253Dhttp%25253A%25252F%25252Fib.adnxs.com%25252Fsetuid%25253Fent
ity%25253D70%252526code%25253De4647673-7aac-4a24-94f6-8103db8ada8d

GET /getuid?http://mapping.nxtck.com/rtb/
um?n=msn&gid=$UID&uuid=e4647673-7aac-4a24-94f6-8103db8ada8d&cb=1123
713856&redir=http%3A%2F%2Fib.adnxs.com%2Fseg%3Fadd%3D209359%2526red
ir%253Dhttp%25253A%25252F%25252Fib.adnxs.com%25252Fsetuid%25253Fent
ity%25253D70%252526code%25253De4647673-7aac-4a24-94f6-8103db8ada8d
HTTP/1.1
Host: ib.adnxs.com
User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:31.0) Gecko/20100101
Firefox/31.0 Iceweasel/31.6.0
Accept: image/png,image/*;q=0.8,*/*;q=0.5
Accept-Language: fr,fr-fr;q=0.8,en-us;q=0.5,en;q=0.3
Accept-Encoding: gzip, deflate
Referer: http://www.monsite-ecommerce.com/
Cookie: sess=1; uuid2=66106801265861589
Connection: keep-alive
```

2.3.2 La mise en commun d'identifiants (le « cookie syncing » ou « cookie matching »)

Une recherche dans les captures HTTP de l'identifiant cookie observé précédemment et commun à plusieurs acteurs publicitaires, montre plusieurs requêtes. Par exemple celle-ci :

```
http://map.sddan.com/MAP.d?mn=nexus&mv=8186779241047110843

GET /MAP.d?mn=nexus&mv=8186779241047110843 HTTP/1.1
Host: map.sddan.com
User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:31.0) Gecko/20100101
Firefox/31.0 Iceweasel/31.6.0
Accept: image/png,image/*;q=0.8,*/*;q=0.5
Accept-Language: fr,fr-fr;q=0.8,en-us;q=0.5,en;q=0.3
Accept-Encoding: gzip, deflate
Referer: http://www.monsite-ecommerce.com/
Cookie: SDDAN=U9RISYRH8PSAAOH4IZXAXWA5SURLYBF7
Connection: keep-alive

HTTP/1.1 200 OK
Server: nginx/1.8.0
Date: Thu, 02 Jul 2015 20:00:17 GMT
Content-Type: image/gif
Transfer-Encoding: chunked
X-Powered-By: PHP/5.4.41-0+deb7u1
Set-Cookie: map_nexus=8186779241047110843; expires=Sun, 26-Jun-2016
20:00:17 GMT; domain=.sddan.com
Via: 1.1 google
Expires: Thu, 02 Jul 2015 20:00:17 GMT
Cache-Control: private
```

On peut voir que cette requête transmet, par passage en paramètre de l'URL, l'identifiant « 8186779241047110843 » et qu'en retour, le serveur web dépose un cookie avec le même identifiant, et cette fois associé à son nom de domaine.

On peut également observer, dans le fichier de capture, que de nombreuses autres requêtes transmettent cet identifiant en paramètre, mais sans dépôt de cookie en retour.

Dans tous les cas, il s'agit de partage/synchronisation de cookies opérés dans le cadre d'enchères publicitaires — le « *matching* » étant réalisé côté client ou côté serveur. Dans ce dernier cas, des tables de correspondance entre les identifiants des différentes régies et acteurs sont construites au niveau des échangeurs et/ou des régies elles-mêmes (ou d'autres acteurs intermédiaires). Cela n'est pas sans implication sur l'ampleur du traçage et sur le lien qui peut désormais être fait entre les différentes données collectées par les différents acteurs publicitaires. En effet, le cloisonnement des identifiants par acteur publicitaire, qui prévalait avant l'avènement du RTB, n'a aujourd'hui plus cours (voir article sur le RTB dans ce dossier).

2.3.3 Mise en évidence du RTB par le chaînage des requêtes

L'observation du chaînage des requêtes, par le biais du « Referer » du protocole HTTP qui indique la requête HTTP précédente ayant appelé celle-ci, peut servir à mettre en évidence les appels générés lors des enchères publicitaires (RTB) décrites dans un article de ce dossier. Par exemple la requête suivante :

```
http://bid.g.doubleclick.net/xbbe/invitepixel/set_partner_uid?
partnerID=79&partnerUID=85eeb731c50701a62fae6f1bb96edebe&sscs_
active=1
GET /xbbe/invitepixel/set_partner_uid?partnerID=79&partnerUID=
85eeb731c50701a62fae6f1bb96edebe&sscs_active=1 HTTP/1.1
Host: bid.g.doubleclick.net
[...]
Referer: http://loadeu.exelator.com/load//net.php?n=eJyNUctuwjAQ%2
FBUUVbnVSSjPFhNVglbcWqGekWMvzkLiWPGSpn9fE1JEb9x2ZleehxdY6kEOqHPiQR
IMvlFR3k2uljzIiexzFDnQJRh6tNhCwdA0SFCCQsFkVfot7ayoyUC9O6FK%2B3mz4t
N52IMvj2ZjgGz6lMhxPI0TMRnuBUz2SZbNJ6Agg9A56XZCEjbAk2C5iLy55cLJGi0N
6McCDwhaig6iERf2v001HDsdGbQWahSmM%2Fd6UghGwubzZtxC3aAE5hoZvQO9YUFQ
g1q3diVIpB%2FXAEk8Cz9OWYEuv%2BLLNV%2F7KgTBFrQLBt7rxdBycVehmDGhTOs6
ixrI12aytF8W1VClT3F%2Bv6q7kzMVpZYP41GoeZyMwgz5gy81PPD42tRd0rJkmqnK
hwJZoDwyAxR1H5vqqtIF7AwqflGHUJT2padleYuc%2FJP9BQxvx8w%3D&h=3ba0f80
e64e874b2dd09ead22a236ef1&ver=2
Cookie: id=22d79de510020013||t=1435867217|et=730|cs=002213fd486980
abab9dadcba2
Connection: keep-alive
```

Il s'agit d'une requête à destination du sous-domaine « bid » (bid=enchère) de « g.doubleclick.net », domaine de la régie publicitaire de Google. Elle est initiée à la suite d'une requête vers un sous-domaine de « exelator.com » qui apparaît dans le « Referer » et qui appartient à la société Exelate. Celle-ci fournit une offre de place des marchés publicitaires (marketplace ou échangeur). L'examen des requêtes met en évidence d'autres requêtes avec le même « Referer » et à destination d'autres acteurs publicitaires.

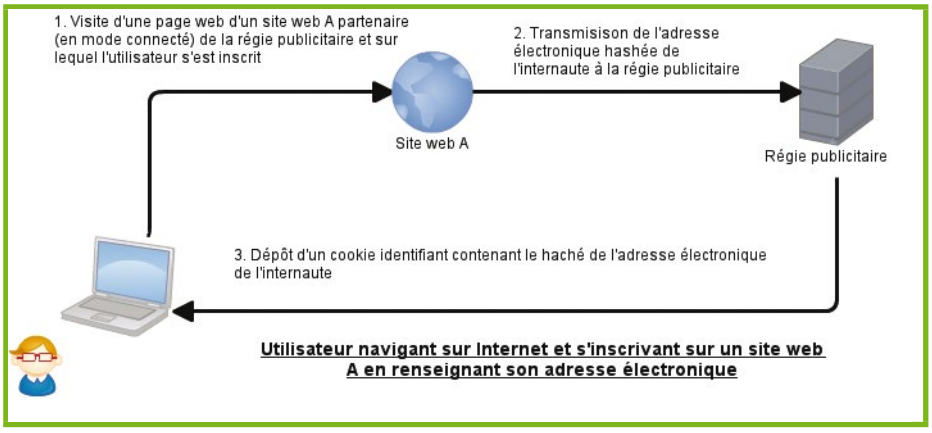

Figure 5

2.3.4 L'adresse électronique comme clé de voûte de l'identification : du « cross canal » au « cross device »

La navigation de l'internaute peut désormais être associée à son adresse électronique, rompant ainsi l'« anonymat » (en fait pseudonymat) annoncé dans la plupart des politiques « cookies » des sites web. D'ailleurs, même sans l'utilisation des techniques décrites dans ce paragraphe, cet « anonymat » relatif du traçage opéré par les cookies peut être rompu par d'autres manières (cf. par exemple **[MISC-Analyse]**).

Le *retargeting* ou le re-ciblage publicitaire sur des articles consultés par l'internaute sur un site sur lequel il n'est pas inscrit est désormais possible par courriel (« e-mail retargeting »). Il recevra alors des publicités par e-mail pour des produits similaires ou semblables à ceux qu'il a consultés sur des sites internet.

L'internaute pourra par ailleurs être contacté sur un autre canal (cross canal) si la régie publicitaire dispose d'une base plus complète à même d'associer à l'adresse e-mail, l'adresse postale de l'internaute ou son numéro de mobile, ou si elle paye une prestation « d'enrichissement de base » auprès d'une société spécialisée.

Cette identification de l'internaute est réalisée par des dépôts de cookies contenant un haché (en général SHA1 ou MD5) de son adresse électronique. Ce dépôt s'effectue soit à l'ouverture d'un courrier électronique, soit à la connexion sur un site partenaire (« matching partner », site web A des figures 5 et 6) qui vend cette possibilité d'identification de l'internaute. Ce second cas est expliqué dans les schémas suivants : voir Figures 4 à 6.

Les captures HTTP réalisées mettent en évidence des requêtes et dépôts de cookie du domaine « emailretargeting.com ».

```
Set-Cookie: etuix=FVO9KtSlVspzVohjEoHNrb71FLhCPd9WYSont0N8uUO1r7CPikBnEA--; expires=Tue, 29 Dec 2015 20:12:07 GMT; path=/; domain=.emailretargeting.com
```

Les hachés d'adresses électroniques peuvent aussi être utilisés pour relier les différents terminaux d'un même utilisateur à des fins de prospection publicitaire (PC personnel, professionnel, smartphone, tablette, etc.). C'est ce qu'on appelle le « cross device ».

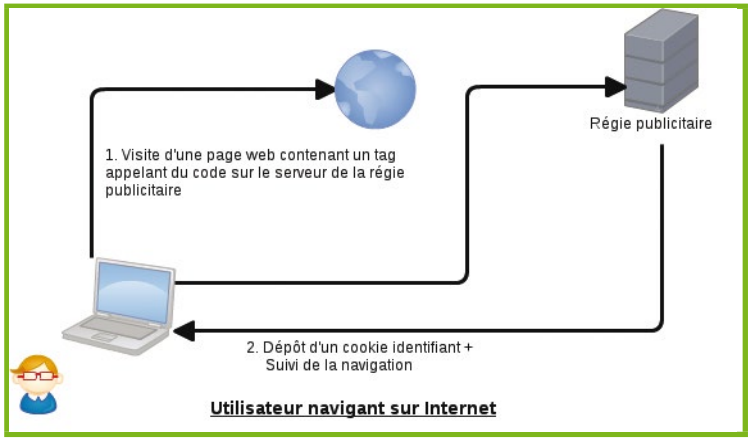

Figure 4

3 Quelques solutions simples pour arrêter (limiter) le traçage

On peut parfois (souvent) lire dans la rubrique cookies des sites internet que l'utilisateur peut s'opposer aux dépôts de cookies en les bloquant via les paramètres de son navigateur, mais qu'après cette action, certaines fonctionnalités du site (souvent essentielles comme la connexion au compte)

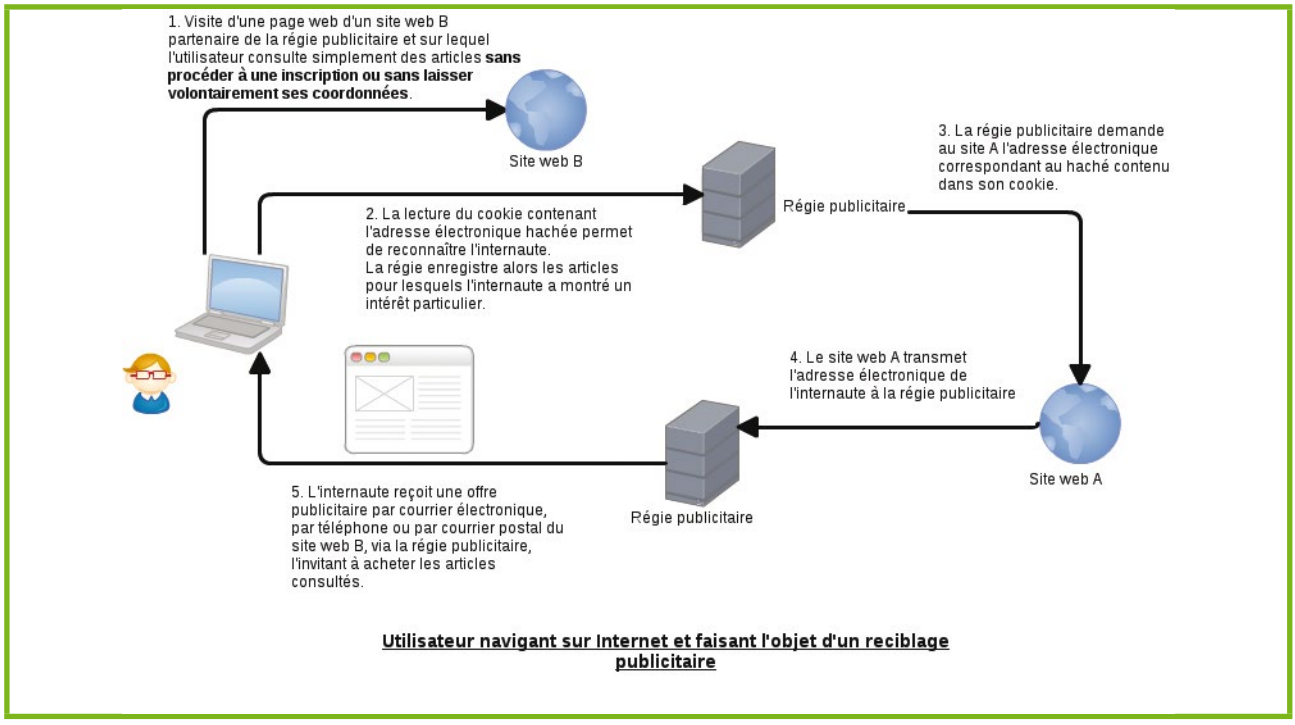

Figure 6

seront inopérantes. C'est une incitation habile à accepter tous les cookies…

Voici donc quelques solutions simples, non exclusives, qui limitent fortement le traçage sans entraver la navigation de l'internaute :

1. Utilisation du mode de navigation privée (« ne jamais conserver l'historique » sous Mozilla Firefox). Cela limite la « mémoire » du traçage à une session seulement.
2. Blocage des cookies tiers dans les paramètres du navigateur. Les cookies tiers étant dans la quasi-totalité des cas non techniques et non fonctionnels, cela n'entravera que très peu l'expérience de navigation. À noter que cette solution ne protège cependant pas des « faux cookies first party » comme ceux de Google Analytics par exemple.
3. Utilisation d'extensions limitant ou bloquant les traceurs, par exemple : DoNotTrackMe, Disconnect, Ghostery ou AdBlockEdge (qui bloque aussi l'affichage des publicités).
4. Limitation du traçage dans les paramètres d'Adobe Flash.

Pour plus de détails, le lecteur pourra se reporter à une rubrique dédiée sur le site Internet de la CNIL **[CNIL-CONSEILS]**.

Enfin, l'e-mail devenant la clé de voûte d'un traçage plus étendu, et même si cela est plus complexe, car cela nécessite de posséder ou d'acheter son propre nom de domaine, il est recommandé à l'internaute qui souhaite vraiment se protéger d'utiliser un alias de son adresse e-mail propre à chaque site. Par exemple, sur le site « monjournal », on pourra fournir l'alias (redirection e-mail) « monjournal@domain » où « domain » correspond à notre nom de domaine. Cela permet de plus de détecter l'origine d'une revente indue de l'e-mail à un spammeur et de couper la redirection si nécessaire…

4 Appendice : quelques remarques sur la sécurité et les cookies

4.1 Les cookies apprécient aussi l'HTTPS – de l'utilité de l'attribut « secure »

La RFC 6265 définit l'attribut **secure** pour les cookies : cet attribut peut-être associé indifféremment à chaque cookie. Lorsqu'il est activé pour un cookie, ce dernier est transmis uniquement à travers un protocole sécurisé.

EXTRAIT DE LA RFC 6265

If the cookie's secure-only-flag is true, then the request-uri's scheme must denote a "secure" protocol (as defined by the user agent).

Certains sites ne chiffrent pas les données (potentiellement personnelles ou sensibles) contenues dans des cookies qu'ils déposent sur le terminal de l'internaute. Voici, par exemple, l'un des (nombreux) cookies tels que déposés après identification (via le protocole HTTPS). Il contient nom, prénom et e-mail de l'internaute :

```
HTTP/1.1 200 OK
Content-Type: text/javascript; charset=utf-8
Date: Thu, 02 Jul 2015 20:35:07 GMT
Expires: now
Pragma: no-cache
Server: Apache
Set-Cookie: info_user_web=%7B%22jelec%22%3A%7B%22droit%22%3A0%7D%
2C%22nom%22%3A%22LABARTHE%22%2C%22prenom%22%3A%22%22%Stephane%22%
3A%22%22%2C%22email%22%3A%22slabarthe%40cnil.fr ; path=/; domain=.
monsite.fr
Set-Cookie: info_user_web_track=%3Fage%3D%26prof%3D%26abo%3D0%
26civ%3D%26cp%3D%26pays%3D%26pub%3D0%26quo%3D1%26web%3D%26pub_
lm%3D0%26pub_lmfr%3D0; path=/; domain=.monsite.fr
```

Or, de nombreuses pages de « monsite.fr » ne sont pas accessibles en HTTPS. Ainsi, l'internaute, au cours de sa navigation, oscille entre des pages en HTTPS et des pages en HTTP.

Il en résulte que, sur les pages accessibles en HTTP, ce cookie, contenant un certain nombre de données personnelles, est transmis en clair.

Pour contrer ce phénomène, il existe plusieurs solutions. Il est possible d'utiliser l'attribut « secure » sur ce cookie. Ainsi, il sera transmis par le navigateur uniquement lors d'appels de pages utilisant un protocole sécurisé (RFC 6265).

Une autre solution, plus radicale, est de forcer l'utilisation du protocole HTTPS pour l'intégralité du site, via le HSTS.

4.2 Le HSTS – politique de chiffrement pour l'intégralité d'un domaine

Le protocole *HTTP Strict Transport Security* **[HSTS]** est un mécanisme de politique de sécurité proposé pour HTTP permettant à un serveur web d'indiquer au navigateur, s'il est compatible, qu'il doit communiquer avec lui en utilisant une connexion sécurisée.

Lorsque la politique HSTS est active pour un site web, le navigateur de l'utilisateur remplace automatiquement tous les liens non sécurisés par des liens sécurisés (http://www.mon-site.fr devient https://www.mon-site.fr). Ainsi, toutes les données transmises, y compris les dépôts et lectures de cookies, le sont sur un canal chiffré.

En cas d'impossibilité d'établir une connexion sécurisée, un message d'alerte est affiché à l'utilisateur.

ATTENTION À L'UTILISATION DE L'HSTS

N.B. : l'HSTS, mécanisme de sécurité simple, mais relativement puissant, doit être utilisé avec prudence. En cas de difficulté avec les certificats SSL/TLS, un navigateur ayant préalablement activé la politique HSTS se verra dans l'impossibilité d'accéder au site web si l'accès en HTTPS est inopérant.

4.3 L'attribut HTTPOnly

Les cookies sont fournis à chaque requête HTTP au serveur web correspondant. Par ailleurs, ils sont accessibles aux fonctions JavaScript qui s'exécutent dans la page. Cela peut représenter un danger et permettre par exemple à des « partenaires » ayant fourni ce type de code d'accéder à d'autres cookies que les leurs ou à un pirate de mener une attaque de type XSS pour « voler » des cookies.

Une bonne pratique, dès lors que le cookie stocke des données sensibles (cookie d'authentification, contenant des données personnelles, etc.), est d'utiliser l'attribut « HTTPOnly » qui indique que le cookie est uniquement accessible via HTTP(s). ■

Note

*** Les avis, opinions et positions exprimées dans le présent article n'engagent que leur(s) auteur(s) et en aucun cas l'institution à laquelle ils appartiennent.**

▪ Références

[LSO] Local Shared Objects (cookies flash) :
- https://fr.wikipedia.org/wiki/Objet_local_partag%C3%A9
- https://www.adobe.com/fr/special/products/flashplayer/articles/lso/
- https://www.adobe.com/fr/support/flashplayer/ts/documents/4c68e546.htm

[RFC6265] RFC 6265 (*HTTP State Management Mechanism*) : http://tools.ietf.org/html/rfc6265 et RFC 2965 https://tools.ietf.org/html/rfc2965

[HSTS] HSTS sur Wikipédia : https://fr.wikipedia.org/wiki/HTTP_Strict_Transport_Security

[CookieViz] Dépôt CookieViz sur GitHub : https://github.com/LaboCNIL/CookieViz

[CookieManager +] https://addons.mozilla.org/fr/firefox/addon/cookies-manager-plus/

[LiveHTTPHeaders] https://addons.mozilla.org/fr/firefox/addon/live-http-headers/

[MISC-Analyse] Article *MISC n°76* (nov./déc. 2014) : « Analyse d'une inscription en ligne : comment vos données fuitent sur Internet ».

[CNIL-CONSEILS] Cookies : conseils pour les maîtriser, rubrique sur le site Internet de la CNIL accessible à l'URL : http://www.cnil.fr/vos-droits/vos-traces/les-cookies/conseils-aux-internautes/

PROFESSIONNELS !

DÉCOUVREZ NOS NOUVELLES OFFRES D'ABONNEMENTS ...

PDF COLLECTIFS

OFFRE	ABONNEMENT	PROFESSIONNELS 1 - 5 lecteurs Réf	Tarif TTC	6 - 10 lecteurs Réf	Tarif TTC	11 - 25 lecteurs Réf	Tarif TTC
PROMC2	6n° MISC	PRO MC2/5	168,-	PRO MC2/10	336,-	PRO MC2/25	672,-
PROMC+2	6n° MISC + 2n° HS	PRO MC+2/5	216,-	PRO MC+2/10	432,-	PRO MC+2/25	864,-

Prix TTC en Euros / France Métropolitaine

PROFESSIONNELS : N'HÉSITEZ PAS À NOUS CONTACTER POUR UN DEVIS PERSONNALISÉ PAR E-MAIL : abopro@ed-diamond.com OU PAR TÉLÉPHONE : 03 67 10 00 20

ACCÈS COLLECTIFS BASE DOCU

OFFRE	ABONNEMENT	PROFESSIONNELS 1 - 5 connexion(s) Réf	Tarif TTC	6 - 10 connexions Réf	Tarif TTC	11 - 25 connexions Réf	Tarif TTC
PROMC+3	MISC + HS	PRO MC+3/5	177,-	PRO MC+3/10	354,-	PRO MC+3/25	708,-
PROH+3	GLMF + HS + LP + HS + MISC + HS + OS	PRO H+3/5	447,-	PRO H+3/10	894,-	PRO H+3/25	1788,-

Prix TTC en Euros / France Métropolitaine

Les abréviations des offres sont les suivantes : LM = GNU/Linux Magazine France HS = Hors-Série LP = Linux Pratique OS = Open Silicium

...EN VOUS CONNECTANT À L'ESPACE DÉDIÉ AUX PROFESSIONNELS SUR :

www.ed-diamond.com

COOKIES ET AUTRES TRACEURS : QUELLES RÈGLES ? QUELLE PROTECTION POUR LA VIE PRIVÉE ?

Clémence Scottez – cscottez@cnil.fr – Juriste du secteur économique, Commission Nationale de l'Informatique et des Libertés *

mots-clés : *COOKIES / TRACEURS / CIBLAGE / MARKETING / LOI INFORMATIQUE ET LIBERTÉS / VIE PRIVÉE*

Les techniques de traçage en ligne et le traitement des données qu'elles fournissent sont encadrés par des règles précises et complémentaires tendant à garantir aux internautes la maîtrise de leurs données, dans un environnement complexe où l'opacité tend à régner.

1 La loi encadre le traçage et le traitement ultérieur des données

1.1 L'encadrement de la technique de traçage par l'article 32 II de la loi « Informatique et Libertés »

En 2009, la réforme d'un corpus de règles européennes connues sous le nom de « Paquet Télécom », et en particulier de la directive dite « vie privée dans le secteur des communications électroniques » (directive 2009/136/CE), a renforcé la maîtrise des internautes sur leurs données en passant d'un principe de droit de « refus » (dit d'opposition) du traçage, au demeurant méconnu et mal appliqué, à un principe de consentement préalable conditionnant l'usage de ces traceurs. La directive prévoit que l'équipement terminal de l'utilisateur ainsi que toute information stockée sur cet équipement relèvent « *de la vie privée de l'utilisateur, qui doit être protégée au titre de la convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales* » et ce indépendamment du fait que les informations collectées puissent être qualifiées de données à caractère personnel. L'idée maîtresse est ici de protéger la confidentialité des communications électroniques.

Ces nouvelles règles ont été intégrées en 2011 dans l'article 32 II de la loi « Informatique et Libertés » française. Désormais, « *tout abonné ou utilisateur d'un service de communications électroniques doit être informé de manière claire et complète, sauf s'il l'a été au préalable, par le responsable du traitement ou son représentant :*

- *de la finalité de toute action tendant à accéder, par voie de transmission électronique, à des informations déjà stockées dans son équipement terminal de communications électroniques, ou à inscrire des informations dans cet équipement ;*
- *des moyens dont il dispose pour s'y opposer.*

Ces accès ou inscriptions ne peuvent avoir lieu qu'à condition que l'abonné ou la personne utilisatrice ait exprimé, après avoir reçu cette information, son accord qui peut résulter de paramètres appropriés de son dispositif de connexion ou de tout autre dispositif placé sous son contrôle.

Ces dispositions ne sont pas applicables si l'accès aux informations stockées dans l'équipement terminal de l'utilisateur ou l'inscription d'informations dans l'équipement terminal de l'utilisateur :

- *soit a pour finalité exclusive de permettre ou faciliter la communication par voie électronique ;*
- *soit est strictement nécessaire à la fourniture d'un service de communication en ligne à la demande expresse de l'utilisateur* ».

Souvent présentées, à tort, comme ne ciblant que la technique des cookies (ou témoins de connexion), ces dispositions régissent en réalité **toutes les actions consistant à « accéder (...) ou à inscrire des informations » (lecture et écriture) sur le terminal d'un utilisateur d'internet**, et répondant à un objectif autre que la fourniture de la connexion ou du service demandé. Par exemple, l'utilisation de « web bugs », d'« etag », du « fingerprinting », de pixels invisibles, permet la transmission d'un paramètre unique permettant de reconnaître l'internaute et constitue donc une opération soumise à accord préalable. De même les cookies http, les « local shared object » (cookie flash), le stockage web local, qui conduisent à inscrire des informations dans le terminal et agissent comme une « balise » permettant

d'individualiser l'internaute, nécessitent également le consentement de l'internaute concerné.

Afin de préciser la portée et les conséquences pratiques de l'article 32.II, la CNIL a publié une recommandation le 5 décembre 2013, dont il résulte que l'accord de l'internaute doit « se manifester par le biais d'une action positive de la personne préalablement informée des conséquences de son choix et disposant des moyens de l'exercer ». En d'autres termes, pour être valable l'accord doit être exprimé librement et en connaissance de la finalité des cookies déposés, et ceci, préalablement au dépôt de cookie. Le consentement étant révocable à tout moment, un moyen simple doit être proposé aux utilisateurs pour, d'une part, supprimer les cookies déjà déposés et, d'autre part, bloquer la lecture et le dépôt de nouveaux cookies.

Dans ces conditions, on comprendra aisément que le paramétrage d'un navigateur, souvent configuré par défaut pour accepter sans distinction tous les cookies, ne puisse pas toujours être considéré comme l'expression d'un choix préalable, libre et avisé de l'internaute. L'abstention de l'internaute ne peut pas non plus s'interpréter comme une action positive et éclairée quant à la finalité des traceurs utilisés par chaque site internet visité, mais comme un consentement « à l'aveugle » conduisant à accueillir indifféremment la lecture et l'écriture ultérieures d'informations sur leur terminal.

1.2 L'encadrement des différents traitements des données de suivi de navigation

Le cas le plus classique reste l'usage de traceurs dans le cadre du marketing digital, afin de diffuser de la publicité personnalisée au regard de la navigation de l'internaute. Alors qu'afficher une publicité contextuelle conduit à adapter le contenu présenté à la nature du site visité (par exemple une publicité pour des baskets sur le site d'un magasin de sport), servir une publicité comportementale repose sur l'analyse des actions du visiteur (visites successives de sites, interactions, mots clés, production de contenu en ligne, etc.) pour établir un profil spécifique et afficher lors de ses visites des publicités personnalisées. Ce suivi peut être riche d'enseignements, par exemple pour définir le sexe et l'âge approximatif de la personne, déduire des pages visitées une classification socio démographique, déterminer des centres d'intérêt et ceci par analogie avec les schémas comportementaux identiques constatés chez d'autres personnes.

Ces profils dits prédictifs, car déduits du comportement, peuvent être croisés avec des profils explicites construits à partir des informations fournies par l'internaute lors de la création et de l'utilisation d'un compte client. À titre d'exemple, la CNIL a relevé dans une délibération n°2013-420 du 3 janvier 2014 prononçant une sanction pécuniaire à l'encontre de Google Inc que cette société « *traitera l'ensemble des données de navigation issues de sites tiers intégrant l'outil DoucleClick pour créer des profils utilisateurs à des fins de ciblage publicitaire, que les personnes concernées soient ou non des utilisateurs authentifiés. Les données collectées par ce biais seront associées aux données figurant dans les comptes utilisateurs Google quand les personnes concernées accéderont ultérieurement à des services requérant leur authentification préalable, et ce alors même que leur consentement spécifique n'aura pas été recueilli en amont* ».

En outre, le croisement des données de comptes utilisateurs et des données de navigation permet, par exemple, d'effectuer des envois de courriers électroniques, postaux, d'identifier les différents terminaux utilisés par une même personne (cf. pratique du « cross canal » et du « cross device » explicitée par **l'article « Détecter et analyser les cookies et autres traceurs »**), ou d'affiner le ciblage et la segmentation applicable à la personne.

De manière plus générale, les techniques de traçage rétablissent, voire renforcent le lien entre un terminal et son utilisateur. Elles sont à ce titre de plus en plus exploitées dans le cadre de traitements de lutte contre la fraude et l'usurpation d'identité, voire pour consolider des techniques d'authentification plus classiques. Ici la reconnaissance de l'individu, grâce aux informations inscrites sur son terminal ou aux paramètres émis par ce dernier, vient confirmer ou infirmer un risque d'usurpation d'identité. Ce sera par exemple le cas lors de la connexion à un compte bancaire, ou d'un paiement par carte bancaire à l'occasion d'un achat en ligne (cf. délibération de la CNIL n°2013-367 du 28 novembre 2013 autorisant la société ONEY TECH à mettre en œuvre un traitement automatisé de données à caractère personnel ayant pour finalité la lutte contre les risques de fraude au paiement sur internet, intégrant, après recueil du consentement préalable, la technique du fingerprinting).

Si la finalité de ces traitements n'est clairement pas « marketing », le traçage des personnes réalisé pour atteindre cette finalité n'est pas pour autant nécessaire à la navigation de l'internaute. La condition du recueil de l'accord préalable et informé de l'internaute inscrite à l'article 32.II de la loi leur est donc applicable.

Certains cookies servent également une finalité très générale de mesure d'audience des sites, pour produire des statistiques. Les fonctionnalités de ces outils sont variables, allant du simple « comptage » du nombre d'internautes entrés sur une page Web (en distinguant les visiteurs uniques revenant à plusieurs reprises sur un même site ou nombre de visites) et n'ayant pas poursuivi leur navigation (taux de rebond), à la réalisation d'opérations dites d'AB testing (par exemple, deux groupes d'internautes sont redirigés vers deux versions distinctes d'un même site web pour évaluer leur succès respectif), ou encore à l'évaluation d'un parcours client pour en améliorer l'ergonomie. Sous réserve du respect d'une série de conditions (telles que la seule production de statistiques, l'interdiction de croiser les données avec des données extérieures, l'interdiction de suivi entre différents sites et la possibilité de refuser le suivi à tout moment, etc.) limitant les risques de ces outils pour la vie privée des internautes, la CNIL permet leur mise en œuvre sans accord préalable de l'internaute.

En résumé, les données issues du traçage concernant des personnes individualisées peuvent être exploitées dans des finalités très diverses. Il faut retenir que chaque traitement opéré sur ces données (profilage publicitaire, lutte contre la fraude, etc.) sera régi par l'ensemble des principes de la « loi Informatique et Libertés ». En effet, la loi couvre toute information rattachable à une personne

susceptible d'être identifiée par différents moyens, que ceux-ci soient à disposition du détenteur des données ou non. Par exemple, les régies traitant les informations relatives aux connexions réalisées depuis un navigateur donné ne sont théoriquement pas en mesure d'affirmer que ledit navigateur est utilisé par Jean Dupont à partir de l'adresse IP collectée ou des éléments que ce dernier a consulté. Néanmoins, quelques recoupements d'informations, par exemple entre l'adresse IP de connexion et les données détenues par le fournisseur d'accès à internet de Jean Dupont, ou encore l'analyse de la localisation de chaque connexion, ou la simple analyse des commentaires laissés par des utilisateurs sous pseudonyme sur certains sites, ou l'identifiant unique propre au cookie déposé permettent de remonter jusqu'à la personne.

Conformément à la loi « Informatique et Libertés », chaque acteur décidant d'exploiter les données de suivi pour son compte doit définir précisément l'objectif de son traitement, s'assurer de sa légitimité, vérifier que les données ne sont pas conservées au-delà de la durée nécessaire pour atteindre cet objectif (par exemple, si les données comportementales peuvent être intéressantes sur un historique de 13 mois pour cibler des offres commerciales, leur conservation au-delà de cette durée ne présente pas beaucoup d'intérêt). De même, les données traitées doivent être pertinentes au regard de l'objectif fixé et ne doivent pas présenter un caractère sensible au sens de la loi (informations portant sur la vie sexuelle, les opinions politiques ou syndicales, à la santé, etc.), sauf si le consentement exprès des intéressés est recueilli.

La protection apportée par la loi « Informatique et Libertés » repose aussi sur la reconnaissance de droits aux personnes concernées, notamment la possibilité de demander l'effacement ou la rectification de leurs informations ou d'en obtenir une copie complète, et d'obtenir la liste des organismes qui en ont été rendus destinataires. Enfin, cela signifie que la collecte des données et leur traitement doivent se faire de manière sécurisée et confidentielle.

2 Mise en application : une brève séquence de la vie en ligne des époux Martin

Après avoir consulté différents sites de voyage pour planifier ses vacances d'hiver au soleil, Madame Martin se rend sur deux sites de presse gratuite ; sur chacun de ces sites, une bannière publicitaire s'affiche lui proposant une offre spéciale pour des vacances à la Réunion et en Guadeloupe.

Entre-temps, un courriel lui a été adressé par une enseigne de cosmétiques vantant les mérites d'une crème solaire. Séduite, elle décide de se rendre immédiatement sur le site de la marque, de se créer un compte client et d'acheter ladite crème en renseignant son adresse pour être livrée à domicile. Toujours à la recherche de bonnes affaires, Mme Martin accepte, comme à son habitude, que ses données soient transmises aux partenaires commerciaux de la marque.

De son côté, Monsieur Martin s'inquiète des conséquences du changement de régime alimentaire inhérent à tout voyage pour un problème de cholestérol récemment diagnostiqué par son médecin. Il tente de trouver des réponses sur différents forums spécialisés accessibles gratuitement en ligne. Une semaine plus tard, un courrier vantant les mérites d'un médicament censé lutter contre le mauvais cholestérol atterrit dans la boîte aux lettres du domicile des époux Martin.

À la suite de chacune de leur visite, ces personnes ont donc reçu de la publicité personnalisée au regard des pages auxquelles elles se sont intéressées. Si ces pratiques ne sont pas interdites, elles résultent de plusieurs opérations techniques devant se succéder et s'opérer dans le respect des principes rappelés plus haut.

2.1 Affichage de bannières publicitaires personnalisées sur les sites de presse

Les sites de voyages consultés par Mme Martin intégraient un tag permettant d'appeler le serveur de la régie publicitaire déclenchant ainsi le dépôt d'un cookie sur son terminal. Ce cookie a été ensuite « reconnu » par la même régie lorsque Mme Martin s'est rendue sur les sites de presse en ligne, grâce à une redirection automatique (et invisible pour Mme Martin) du navigateur vers le serveur de la régie. Cet appel enclenche la diffusion d'une publicité personnalisée (offres de voyage), en temps réel, sur l'espace prévu à cet effet.

Lors de son « arrivée » sur les sites de voyages et les sites de presse, Mme Martin doit avoir été mise en mesure de consentir préalablement au dépôt ou à la lecture de ces témoins de connexion, de manière éclairée. En pratique, cela signifie que sur chaque site visité, un message lisible lui expliquant en des termes simples la finalité des traceurs utilisés (constituer des profils aux fins de publicité ciblée par exemple) et la manière d'y consentir doit s'afficher. Aucun de ces traceurs ne doit être activé tant que Mme Martin n'a pas manifesté son accord, que ce soit en poursuivant sa navigation en toute connaissance de cause ou en cliquant sur une case prévue à cet effet. Par ailleurs, si Mme Martin souhaite visiter le site sans accepter les cookies, un moyen simple et aisément accessible doit lui être proposé, par exemple, via un lien figurant dans le bandeau. Ce lien peut renvoyer vers une page expliquant plus en détail la finalité des cookies et la manière de les bloquer. Le refus qu'exprimerait Mme Martin, que ce soit lors de son arrivée sur le site ou ultérieurement, ne doit pas avoir pour conséquence de la priver de l'accès au site ou à certaines de ses fonctionnalités importantes (ex : achat, accès à l'espace connecté), à défaut de quoi, le choix qu'elle exprime ne serait pas considéré comme libre. En effet, conditionner l'accès à un contenu à l'acceptation des traceurs aurait inévitablement pour effet d'orienter la décision de Mme Martin.

Notons que les éditeurs des sites internet consultés par Mme Martin maîtrisent le code intégré dans leurs pages et les tags appelant les serveurs de régies tierces. Dès lors, il leur appartient d'adapter ces appels en fonction des choix exprimés par Mme Martin, au besoin en recourant à des outils de gestion de tag prévenant le déclenchement des éléments qui vont déposer ou lire des cookies (cf. **article « Comment mettre son site en conformité »**).

La prise en compte de l'opposition au dépôt de cookies peut passer par le dépôt d'un cookie dit « d'opt-out », l'emploi des solutions de tag management précitées, ou, le paramétrage du navigateur, sous certaines réserves, cette dernière solution n'étant aujourd'hui pas adaptée à la majorité des situations. En effet, la plupart des navigateurs actuels distinguent deux groupes de cookies en fonction de leur provenance, à savoir les cookies tiers (renvoyant à des serveurs tiers au site) et les cookies first (renvoyant au serveur du site visité). Le paramétrage permet de bloquer ces cookies par groupe, sans spécifier la finalité et l'exploitant de chaque cookie. Par conséquent :

- si les cookies publicitaires concernés par le refus proviennent de serveur tiers, les éditeurs de site peuvent utilement expliquer, sur une page dédiée, la manière de bloquer le dépôt et la lecture de cookies venant de serveurs tiers, sans distinction, par un paramétrage adapté de son navigateur ; opter pour ce paramétrage ne fera pas obstacle à l'accès au site ;
- à l'inverse si les cookies publicitaires concernés par le refus renvoient au serveur du site consulté (cookie « first »), la seule option consiste à bloquer l'ensemble des cookies first en paramétrant le navigateur. Dans cette configuration, tous les cookies déposés par le serveur du site, y compris ceux nécessaires au confort de navigation (et donc ne nécessitant pas d'accord préalable de l'internaute) seront bloqués, compromettant ainsi l'accès de l'internaute au site qui l'intéresse. Pour les mêmes raisons, le paramétrage du navigateur ne sera pas forcément une solution satisfaisante si le site utilise des cookies « analytics » renvoyant à son domaine (en « first » donc) nécessitant le consentement préalable. Dans ces deux hypothèses, une solution d'opposition ad hoc devra être proposée.

Des extensions de navigateur à destination du grand public permettent également aux internautes de filtrer les traceurs et d'exprimer par ce biais leur choix. En ce sens, Mme Martin pourrait utilement intégrer à son navigateur des extensions afin de limiter sa traçabilité. Le choix de recourir à ces outils doit être pleinement pris en compte par les acteurs de la publicité ciblée, comme l'expression d'un droit de refuser ou de consentir appartenant à la personne concernée. La pratique du « respawning » (ou résurrection) développée par certains fournisseurs de réseau publicitaire pourrait en ce sens être contraire à l'article 32.II dès lors qu'elle vise à remplacer les cookies traceurs traditionnels par de nouvelles techniques de traçage indifférentes aux paramétrages traditionnels des navigateurs tels que les « flash cookies » ou le fingerprinting. Si le navigateur est en mesure de bloquer la technique des cookies, les outils n'ont pas encore été développés pour détecter et interdire l'usage d'une technique de fingerpinting ou d'autres méthodes de traçage (cf. article du dossier relatif au fingerprinting). Dans la même logique, les solutions permettant de contourner le filtrage paramétré par les internautes au moyen de ces extensions nient le choix ainsi exprimé par l'internaute.

2.2 Envoi de courriels personnalisés à Mme Martin

Dans le scénario précité, la proximité entre vacances au soleil et l'objet du courriel publicitaire reçu par Mme Martin est frappante. Cet envoi ciblé peut s'expliquer par la propension de Mme Martin à accepter que ses coordonnées soient transmises à des partenaires commerciaux lorsqu'elle crée des comptes en ligne. Il est probable que l'adresse e-mail fournie par Mme Martin lors de son inscription sur l'un des sites de voyages ait été transmise par ce dernier, en version hachée, à sa régie afin qu'un cookie comportant le haché de l'adresse électronique soit déposé sur le terminal de Mme Martin (tel que décrit par **l'article « Détecter et analyser les cookies et autres traceurs »**). L'enseigne de cosmétique recourt aux services d'emailing de la même régie et la mandate pour cibler les prospects présents en base, susceptibles d'être intéressés par de la crème solaire. Après avoir identifié les personnes entrant dans ce segment (dont Mme Martin), la régie demande au site de voyages l'adresse électronique correspondant au haché contenu dans le cookie présent sur le terminal de Mme Martin afin de lui adresser une offre spéciale « crème solaire ».

Plusieurs « couches » réglementaires sont ici susceptibles de s'appliquer.

En premier lieu, Mme Martin doit avoir accepté le dépôt des cookies, après avoir été informée de leur finalité, notamment du croisement de ses données de navigation avec son e-mail (cf. 2.2).

En second lieu, Mme Martin doit avoir consenti expressément à la transmission de son adresse électronique à des fins de prospection pour le compte de tiers, conformément à l'article L.34-5 du code des postes et des communications électroniques (réglementation anti-spam). Selon cette même disposition, l'émetteur du message de prospection est également tenu de faire figurer sur chaque message un moyen d'opposition ainsi que l'identité de l'organisme pour le compte duquel la prospection est effectuée (en l'espèce l'enseigne de cosmétique).

En troisième lieu, au regard de la loi « Informatique et Libertés » Mme Martin doit avoir été informée de la nature des traitements effectués sur ses données (y compris par la régie), notamment de leur finalité (la loi impose tant au site internet qu'à la régie de s'assurer que leur objectif est explicite), de la manière d'exercer ses droits d'accès de rectification et d'opposition, auprès de chaque acteur (responsable du site et régie en l'espèce), et de connaître, en toute logique, l'identité des destinataires de ses données. Une liste précise des commerçants susceptibles d'utiliser ses coordonnées doit être mise à sa disposition, sur simple demande. Les différents accords exprimés par Mme Martin ne doivent pas être absorbés dans une acceptation générale des CGU et être suffisamment éclairés pour que Mme Martin sache à quoi s'attendre.

2.3 Envoi de courriers postaux personnalisés à M. Martin

L'origine de l'offre relative à l'anti-cholestérol adressée par voie postale peut être expliquée de la même manière que pour le courriel publicitaire relatif à la crème solaire. Monsieur M a utilisé le même navigateur que son épouse lors de la consultation des forums en ligne. Ces forums appellent également le serveur de la régie proposant d'effectuer pour le compte d'annonceurs des envois publicitaires ciblés. La différence ici est que l'adresse e-mail hachée de Mme Martin permettra à la régie, non

pas de récupérer la version en clair de cette donnée, mais de demander au site en possession de ces informations l'adresse postale indiquée lors de la création de compte.

Une autre hypothèse tout aussi plausible est que le e-commerçant à l'origine de la collecte de toutes ces coordonnées les ait transmises dans leur ensemble à la régie, laquelle se trouve parfaitement en mesure de faire le lien avec l'adresse postale associée. Les principes applicables à l'envoi de la publicité pour la crème solaire sont donc applicables en l'espèce, avec toutefois une contrainte additionnelle liée à la nature des données exploitées : en l'espèce l'élément révélateur pour cibler les besoins potentiels de M. Martin est son état de santé, information qualifiée de « sensible » au sens de la loi.

Or le traitement de ce type d'information, notamment à des fins commerciales, est par principe prohibé par la loi « Informatique et Liberté » sauf si la personne concernée a expressément consenti à une telle exploitation. En d'autres termes, M. M aurait du cocher une case ou signer un écrit par lequel il accepte spécifiquement que les informations relatives à son diabète naissant servent à des commerçants pour lui adresser de la publicité ciblée. Un consentement « fort » est d'autant plus nécessaire dans ce cas que la divulgation de ces informations peut s'avérer gênante. En pratique, la pratique exposée dans cet exemple ne répond pas aux exigences de la loi dès lors que la collecte des données nécessaires au ciblage publicitaire et à l'envoi d'un courrier postal a été réalisée à l'insu de M. M et, a fortiori, en l'absence de tout consentement.

Les 3 hypothèses étudiées ci-dessus peuvent se décliner indéfiniment Ces situations doivent être étudiées, au cas par cas, en tenant compte de chaque acteur impliqué, de leur rôle respectif, du type d'informations et de croisements traités ainsi que de l'objectif de chaque traitement effectué. Indépendamment de la complexité de l'exercice, les acteurs soumis à ces différentes règles doivent garder à l'esprit l'objectif de transparence et de maîtrise des internautes sur leurs propres données.

3 Le traçage : vecteur de risque pour la sécurité des internautes

L'article 34 de la loi impose aux responsables d'un traitement de prendre « *toutes précautions utiles, au regard de la nature des données et des risques présentés par le traitement, pour préserver la sécurité des données et notamment, empêcher qu'elles soient déformées, endommagées ou que des tiers non autorisés y aient accès* ». Il s'agit d'une obligation dite de « moyen renforcée », à savoir qu'en cas de faille ou de violation des données, il appartiendra au responsable du traitement de prouver qu'il a pris toutes les mesures permettant de prévenir le risque. Si la faille concerne le traçage ou le traitement subséquent des données, la responsabilité reposera selon les cas, sur l'éditeur du site ayant permis le dépôt, sur l'annonceur, sur le fournisseur de réseau publicitaire, etc.

La transmission en clair des informations parfois directement identifiantes contenues dans les cookies pourrait être reprochée à l'éditeur du site internet et au fournisseur de réseau publicitaire qui l'a généré (sur ce sujet cf. **article « Détecter et analyser les cookies et autres traceurs »**).

Le risque peut naître du fait que l'éditeur abandonne à des tiers, dont il connaît peu ou mal l'activité et le sérieux, voire qui lui sont inconnus, la suite du processus permettant d'occuper un espace visuel sur son site pour afficher une annonce. Cette problématique est d'autant plus prégnante que la collaboration entre régies s'étoffe progressivement pour rentabiliser des mécanismes de mise aux enchères (cf. **article relatif aux plateformes RTB**). Le risque peut également résulter d'une potentielle faille de sécurité propre à la technologie de ciblage ou touchant les bases de données des régies.

Un récent rapport du Sénat américain du 15 mai 2014 nommé « *Online Advertising and Hidden Hazards to Consumer Security and Data Privacy* » (http://www.hsgac.senate.gov/hearings/online-advertising-and-hidden-hazards-to-consumer-security-and-data-privacy) souligne, expérience à l'appui, les vulnérabilités intrinsèques à l'écosystème du marketing ciblé en ligne pour conclure à l'urgence d'une prise de conscience des acteurs. Ce rapport souligne que le consommateur consultant un site est contraint de faire confiance, par défaut, à des tiers dont il ignore souvent l'identité. Les responsables de sites ne sont pas non plus en mesure d'identifier tous les acteurs du réseau publicitaire interagissant avec les internautes depuis leur site. La qualité et la sécurité des transmissions sont finalement confiées aux intermédiaires de types « ad network », « supply side platform » et « demand side platform », sans véritable garde-fou. Le rapport souligne à cet égard que l'affichage d'une publicité passe classiquement par 5 ou 6 intermédiaires, chaque point de contact constituant une source de risque potentielle. De même, des codes malveillants pourraient aisément être intégrés dans les publicités affichées sur le site par des serveurs tiers, code exécutable à l'insu de l'utilisateur et de l'éditeur de site peu précautionneux.

Plusieurs cas, encore peu médiatisés en Europe ont été recensés aux États-Unis. Par exemple, en 2012, les visiteurs du site de la Major League Baseball (MLB) ont été exposés à une publicité (pour des montres de luxe) diffusant un virus à la suite d'un simple clic, à la suite de la probable compromission de leur « ad network ». Les experts ont à l'époque attribué le problème aux risques créés par la multiplication des couches de « syndication » (dans le rapport, ce terme anglais renverrait à la pratique qui consiste à acheter un espace publicitaire aux enchères puis à le remettre en vente), rendant quasiment impossible l'identification de la source du malware.

La source immédiatement visible pour l'utilisateur victime de ces pratiques sera en revanche le site qui a permis sa diffusion. Au-delà des problématiques de responsabilités qui se poseront inévitablement, la sécurisation des mécanismes de traçage est un enjeu pour l'image et la réputation des sites qui y font appel. ■

Note

*** Les avis, opinions et positions exprimées dans le présent article n'engagent que leur(s) auteur(s) et en aucun cas l'institution à laquelle ils appartiennent.**

METTRE SON SITE WEB EN CONFORMITÉ AVEC LA RECOMMANDATION « COOKIES »

Vincent Toubiana – vtoubiana@cnil.fr
Ingénieur au Service de l'Expertise Technologique de la CNIL *

mots-clés : *COOKIES / OUTILS / GESTION DE TAGS / MESURE D'AUDIENCE/ CONTENT SECURITY POLICY / DONOTTRACK*

Depuis près d'un an, les internautes ont vu fleurir des « bandeaux cookies » sur les sites web. Mais si informer les internautes lors d'un dépôt de cookie est effectivement nécessaire, ce n'est pas suffisant. De nombreux sites ne sont aujourd'hui pas conformes. Dans cet article, nous expliquons comment peut être mis en place un bandeau conforme et les choix qu'il doit offrir à l'utilisateur.

La CNIL a publié sa recommandation en décembre 2013 [http://www.cnil.fr/documentation/deliberations/deliberation/delib/300/], puis annoncé en octobre 2014 le lancement de contrôles de son application. Depuis, un nombre grandissant de « bandeaux cookies » est constaté sur le web français, afin d'informer les utilisateurs que des cookies sont déposés, à tel point que de nombreux internautes en font une indigestion ! Pourtant, la CNIL ne recommande pas une information constante des internautes en maintenant le bandeau d'information au-delà de la première visite du site. En effet, dès lors que l'information proposée est visible, mise en évidence et complète lors de la première visite du site, et à partir du moment où elle permet aux internautes d'exprimer leur accord par une action concrète (par exemple en poursuivant leur navigation par le déroulement de la page, en cliquant sur un lien, etc.), ou leur refus (en utilisant les mécanismes d'opposition proposés par le site), avant tout dépôt ou lecture de cookies, il n'est pas nécessaire de redemander l'accord des mêmes internautes en affichant systématiquement le bandeau d'information sur toutes les pages du site ou lors de visites ultérieures. En résumé, le bandeau ne doit être affiché que tant que l'utilisateur n'a pas poursuivi sa navigation et surtout il doit permettre de faire un vrai choix.

1 La mesure d'audience

Le recueil du consentement, et donc l'affichage d'un bandeau *cookie*, n'est pas systématiquement requis. Ainsi, les cookies nécessaires à la fourniture du service demandé par les internautes (par exemple l'accès au site), ne nécessitent ni le recueil de leur consentement, ni une information préalable. De fait, tous les cookies techniques (entendu comme absolument nécessaires pour le confort de navigation) ne sont pas soumis à la réglementation.

En plus des cookies techniques, certains cookies de mesure d'audience (i.e. *analytics*) peuvent être déposés dès l'arrivée de l'internaute, et donc sans recueillir préalablement son accord, à condition de respecter les conditions prévues dans la recommandation « cookies », permettant de bénéficier de l'exemption du recueil du consentement.

1.1 Les solutions exemptées

À ce jour, Google Analytics ne peut pas être configuré pour bénéficier de l'exemption du recueil du consentement, seuls les outils de la société At-Internet (ex Xiti) et le logiciel libre Piwik peuvent l'être.

En effet, les critères d'exemptions imposent notamment :

- des cookies dont la date d'expiration n'est pas prorogée au cours de la navigation ;
- une rétention limitée des données collectées ;
- la possibilité de s'opposer au suivi de la navigation.

Les outils mentionnés ci-dessus offrent de telles options. Les éditeurs n'ont qu'à intégrer les mentions d'information dans leur politique de gestion des cookies et à fournir un lien vers le mécanisme d'opposition (*opt-out*) de la solution retenue, sans la nécessité d'informer les utilisateurs par le biais d'un bandeau.

Dans le cas de Piwik, l'*opt-out* par défaut se configure en se rendant sur un lien, qui permet aux internautes de s'opposer au suivi de leur navigation : **URL_DE_VOTRE_SITRE/index.php?module=CoreAdminHome&action=optOut&language=fr**.

Sur AT-Internet, si vous avez opté pour la solution en hébergement tiers, le lien est le suivant : http://www.xiti.com/fr/optout.aspx.

Piwik est une solution qui est majoritairement installée sur le serveur de l'éditeur du site, il faut donc être vigilant lors de sa configuration afin de ne pas ouvrir de brèche dans votre système. Il est vivement recommandé de suivre les conseils de sécurisation de Piwik [http://piwik.org/docs/how-to-secure-piwik/].

1.2 Le tag Google Analytics

Google Analytics est l'outil de mesure d'audience (*analytics*) le plus utilisé sur le web actuellement [https://hal.inria.fr/hal-00915249/PDF/SellingOffPrivacyAtAuction.pdf]. Il ne peut pas entrer dans le cadre de l'exemption, car les données collectées sont conservées pour une durée indéfinie par Google [http://www.google.com/intl/en/policies/technologies/ads/].

Toutefois, Google permet de désactiver dynamiquement Google Analytics sur une page en définissant la variable `'window['ga-disable-UA-ID_DU_SITE']'` à `'true'` [https://developers.google.com/analytics/devguides/collection/analyticsjs/advanced#optout]. C'est en se basant sur cette solution que les services de la CNIL ont élaboré un code fourni ici [https://github.com/CNILlab/Cookie-consent_Google-Analytics] permettant d'afficher un bandeau de demande de consentement et de retenir le dépôt des cookies tant que l'utilisateur n'interagit pas avec la page.

Le code est certes plus long que le code fourni par Google, mais il permet d'avoir un bandeau conforme à la recommandation « cookie » de la CNIL. À noter que, si vous utilisez les fonctionnalités publicitaires fournies par Google Analytics, il faudra modifier le texte d'information figurant dans le bandeau pour faire mention de cette finalité et permettre de s'y opposer.

2 Les boutons sociaux

L'utilisation de boutons sur les sites web permet de mettre en avant certains contenus via les plateformes sociales (Facebook, Twitter, LikedIn, Tumblr, etc.) en bénéficiant d'un « effet viral ». De nombreux éditeurs disposent de ces « boutons sociaux » sur la totalité de leurs pages. Ils incitent ainsi au partage. Toutefois, la plupart de ces boutons sociaux sont déposés directement depuis les domaines des plateformes sociales tierces. Leur simple présence sur le site visité provoque donc des dépôts et des lectures de cookies, même si l'utilisateur ne souhaite pas partager de contenu. Les plateformes utilisent les informations collectées via ces modules d'extension (*plugins*) pour personnaliser les publicités qu'elles affichent.

Par ailleurs, les plugins sociaux ne s'appuient pas uniquement sur les cookies pour suivre les internautes. Ils font souvent appel aux empreintes de matériels et logiciels (*fingerprinting*). Ainsi, AddThis a été pris « la main dans le sac » en 2014 [https://securehomes.esat.kuleuven.be/~gacar/persistent/#results] et les politiques de confidentialité de Facebook [https://www.facebook.com/help/cookies/update] et Google [https://www.google.fr/intl/fr/policies/privacy/] mentionnent l'utilisation de « cookies et de technologies similaires », ce qui signifie que ces sociétés pourraient avoir recours au fingerprinting, auquel cas recourir au blocage des cookies tiers deviendrait une solution d'opt-out inefficace.

2.1 Social Share Privacy

Social Share Privacy est une solution qui remplace le plugin social « classique » par un bouton disposant d'un interrupteur. Tant que le visiteur n'appuie pas sur l'interrupteur, le bouton est inactif et les cookies ne sont pas déposés. Le plugin Social Share Privacy se présente sous la forme d'un tag qui fait appel à une librairie JQuery. Il est donc intégrable sur la plupart des sites. Pour les utilisateurs d'outils de gestion de contenus (*Content Management System*, CMS), des plugins existent, mais ils ne sont pas nécessairement maintenus.

L'intégration de Social Share Privacy se fait en ajoutant la ligne suivante dans l'entête (*header*) afin de désactiver les services non nécessaires. La ligne est adaptée en fonction des outils que vous souhaitez proposer sur votre site :

```
<script type="application/x-social-share-privacy-settings">
{"path_prefix":"https://panzi.github.io/SocialSharePrivacy/","layout":
"line","services":{"options":{"status":false},"buffer":{"status":false},
"delicious":{"status":false},"disqus":{"status":false},"fbshare":{"status":
false},"flattr":{"status":false},"gplus":{"status":false},"hackernews":
{"status":false},"linkedin":{"status":false},"mail":{"status":false},
"pinterest":{"status":false},"reddit":{"status":false},"stumbleupon":
{"status":false},"tumblr":{"status":false},"xing":{"status":false}}}
</script>
```

À l'endroit où vous souhaiterez disposer les boutons, vous n'aurez qu'à ajouter la ligne suivante :

```
<div data-social-share-privacy='true' width=140></div>
```

Social Share Privacy permet un choix fin de l'utilisateur : celui-ci consent pour une page donnée. Il n'impacte pas la charte graphique des sites sur lesquels il est intégré. En effet, le bouton est assez similaire au bouton « *like* » de Facebook. Malheureusement, seules les fonctions « *like* » et « *tweet* » sont incluses. Les *widgets* un peu plus évolués comme les « *embeds* » et les « *timelines* » ne sont pas supportés.

2.2 Le remplacement par des images

Une autre solution mise en avant consiste à simplement remplacer les boutons sociaux par des images pointant vers la page du site sur la plateforme sociale. Ainsi, lorsqu'un visiteur clique sur un bouton social, il est redirigé sur la page dédiée au site visité sur le réseau social et peut dès lors partager le contenu comme il le souhaite. Cette solution présente l'avantage d'être adaptable à toute forme de site.

Certains sites ont mis en pratique des solutions différentes. C'est notamment le cas du site Mashable, qui

Boutons sociaux de Mashable avant que le visiteur ne passe son pointeur sur le bouton « Facebook ».

affiche par défaut une image « like » et qui ne télécharge le bouton Facebook « original » que si l'utilisateur passe sa souris sur cette image. Si l'intention est bonne, il faut noter que cette seule action ne peut être interprétée comme un consentement.

Boutons sociaux de Mashable après le téléchargement du bouton Facebook.

2.3 L'affichage du nombre de « likes »

L'affichage du nombre de partages d'une page fournit souvent un signal au visiteur des contenus les plus populaires et impacte sa navigation sur le site. Malheureusement, ni Social Share Privacy, ni les boutons sous forme d'image ne permettent d'afficher le nombre de personnes ayant « aimé » une page avant que l'utilisateur ne clique sur le bouton.

Il peut être donc intéressant d'avoir soit une estimation de ce chiffre, soit sa valeur exacte. Dans les deux cas, il sera nécessaire de passer par une requête XML-HTTP afin de récupérer cette valeur. Dans le cas de Facebook, cette valeur est accessible même si l'utilisateur n'envoie aucun cookie, ce qui permet d'effectuer la requête en temps réel et d'obtenir une valeur à jour du nombre de « likes » d'une page.

3 Les outils de gestion de tags

Les outils présentés dans les sections précédentes sont adaptés à des fonctionnalités bien précises. Si vous souhaitez fournir des fonctionnalités autres que le partage de contenu, il vous sera nécessaire de recourir à une solution de gestion de tags plus polyvalente, mais qui peut sembler plus compliquée à intégrer.

Les solutions de « *Tag Management* » permettent de contrôler l'activation d'une balise JavaScript. Dans le cadre des cookies, ce contrôle permet de prévenir le déclenchement des éléments qui vont déposer ou lire des cookies tant que l'utilisateur n'a pas donné son consentement ou s'il ne consent pas.

Les tags qui peuvent être ainsi gérés sont de toutes les natures : publicités, vidéos, boutons sociaux, widgets, etc. Tant que l'appel JavaScript déclencheur peut être circonscrit, cette approche est applicable.

3.1 Le marché des solutions de gestion de tags

De nombreuses solutions de gestion de tags existent. Une grande partie d'entre elles s'intègre dans le cadre de prestations fournies par des entreprises spécialisées. Le présent article n'a pas pour objectif de les comparer, puisque nous allons avant tout nous focaliser sur les alternatives gratuites et libres. Néanmoins, si vous choisissez de passer par une solution payante, assurez-vous bien que celle-ci est conforme et qu'elle ne dépose pas de cookies non nécessaires avant que l'utilisateur ait poursuivi sa navigation. Il est par ailleurs primordial de vérifier qu'aucun cookie non nécessaire n'est déposé lorsque l'utilisateur s'oppose au dépôt de cookies.

En effet, certaines solutions déposent juste des cookies d'opt-out à la publicité ciblée lorsque l'utilisateur s'oppose au dépôt de cookies. La plupart de ces cookies ne sont pas nécessaires et contiennent des identifiants qui seront utilisés pour tracer l'utilisateur. De plus, le dépôt de ces cookies ralentit considérablement la visite de l'internaute lors de sa première visite. **Il faut donc vérifier les moyens d'opposition déposés par les *Tag Managers* lorsque l'utilisateur s'oppose.**

3.2 Cookie Cuttr

Cette solution a été développée suite à l'adoption de la loi anglaise sur les cookies. Il s'agit d'une solution fonctionnant en JavaScript, qui consiste à encapsuler les tags dans les scripts qui ne seront exécutés que si l'utilisateur a effectivement consenti. Si le consentement n'a pas été obtenu, le tag ne sera tout simplement pas appelé. Initialement, un seul consentement était disponible. Mais l'outil a été adapté pour fournir un consentement par famille de cookies. Il est donc désormais possible d'exprimer un choix pour chacune des grandes familles (mesure d'audience, réseaux sociaux et publicité).

Pour mettre en place l'outil, il faut télécharger les fichiers disponibles ici [https://github.com/CNILlab/cookieCuttr]. Il faut ensuite modifier les tags des différents scripts en les faisant précéder des balises conditionnelles. Ces balises permettront de s'assurer que les tags ne sont appelés que si les conditions sont remplies.

3.3 Tarte au citron

« Tarte au citron », dont le nom est assez peu conventionnel, est une solution de gestion de tags efficace et flexible. Initialement, cette solution ne concernait qu'un petit nombre de services. Le catalogue de services supportés a considérablement grandi au cours des derniers mois et l'outil permet désormais de gérer les cookies liés aux vidéos, les widgets associés à différents services, ainsi que plusieurs régies publicitaires.

L'outil est disponible dans une version payante, qui peut être intégrée à certains CMS et qui dispose de facilités de configuration. Nous n'étudierons ici que la version gratuite, qui nécessite de disposer des bibliothèques JavaScript sur votre site et de les mettre à jour régulièrement.

L'installation requiert de copier les bibliothèques dans le répertoire de votre choix. Une fois les bibliothèques copiées, il vous faudra éventuellement intégrer la mise en forme de l'outil avant de configurer chacune des fonctionnalités en les intégrant dans vos différentes pages.

```
<head>
 <script type="text/javascript" src="/tarteaucitron/tarteaucitron.js"></script>
        <script type="text/javascript">
        tarteaucitron.init({
                "hashtag": "#tarteaucitron", /* Ouverture automatique du panel avec le
hashtag */
                "highPrivacy": false, /* mettre à true désactive le consentement
implicite */
                "orientation": "top", /* le bandeau doit être en haut (top) ou en bas
(bottom) ? */
                "adblocker": false, /* Afficher un message si un adblocker est détecté */
                 "showAlertSmall": true, /* afficher le petit bandeau en bas à droite ? */
                "cookieslist": true, /* Afficher la liste des cookies installés ? */
                "removeCredit": false /* supprimer le lien vers la source ? */
      }) ;
 </script>
</head>
```

Il faut ensuite déclarer tous les plugins que vous allez utiliser sur votre site et les intégrer dans le corps des pages. Pour la plupart d'entre eux, il suffit juste de les déclarer dans le tableau des « *job* » de Tarte au citron. Dans de rares cas, il faut ajouter des paramètres spécifiques à votre site :

```
<script type="text/javascript">
tarteaucitron.user.productId = 'YOUR-ID';  // Si vous avez un product ID
tarteaucitron.user.functionmore = function () { /* Parfois il est possible
d'ajouter des paramétres*/ };
(tarteaucitron.job = tarteaucitron.job || []).push('product_name_1');
...
tarteaucitron.job.push('product_name_X');
</script>
```

Vous n'avez pas nécessairement besoin de mettre ces éléments sur toutes les pages. Il suffit qu'ils soient présents sur les pages où les plugins sont utilisés. Toutefois, il est plus simple de les intégrer sur toutes les pages et cela permet d'obtenir un consentement de l'internaute en amont. Chaque type de plugin se configure différemment. Certains nécessitent d'être légèrement modifiés pour être configurés. Nous prendrons ici l'exemple d'une vidéo :

```
<iframe width="640" height="480" src="http://www.dailymotion.com/embed/video/
x16lt53_qu-est-ce-qu-un-cookie_tech " frameborder="0" allowfullscreen></iframe>
```

Si vous souhaitez insérer la même vidéo avec Tarte au citron, il faut tout d'abord adapter et intégrer dans le corps de la page le bout de script précédent pour que Tarte au citron intègre Dailymotion à la liste des outils pour lesquels il va demander un consentement.

Reste ensuite à intégrer la vidéo, il suffit de remplacer l'appel habituel de la vidéo par le « **div** » suivant :

```
<div class="dailymotion_player" videoID=" x16lt53_qu-est-ce-qu-un-
cookie_tech " width="640" height="480" showinfo=" 1" autoplay="0"></div>
```

Au moment de l'affichage de la page, **tarteaucitron.js** va être appelé et la balise « div » ne sera remplacée par la vidéo que si le consentement a été obtenu. Dans le cas contraire, une image demandant le consentement sera affichée à la place de la vidéo.

L'outil Tarte au citron est open source (licence MIT) et peut être modifié si vous avez des fonctionnalités particulières à apporter, soit en faisant la demande [https://opt-out.ferank.eu/fr/contact/], soit en contribuant [https://github.com/AmauriC/tarteaucitron.js].

4 Introduction aux « Content Security Policies »

Une « *Content Security Policy* » (CSP) est une politique, définie par l'auteur d'une application ou d'une page web, qui informe le client des sources qui seront autorisées à charger du contenu sur le site (« **declarative policy that lets the authors (or server administrators) of a web application inform the client about the sources from which the application expects to load resources** »). Cette politique peut être vue comme une liste de contenus qui peuvent être chargés par le navigateur lors de l'accès à la page. Ces politiques peuvent être déclarées de deux façons : soit via le header HTTP, soit par le champ **http-equiv** présent dans l'entête HTML d'un document.

4.1 Déclaration d'une politique

Pour se mettre en conformité et empêcher que des contenus tiers ne soient déposés avant l'obtention du consentement, un site web peut déclarer une politique qui bloque le contenu provenant de sites tiers. Notez que cette approche n'est pas « *cookie-centric* », puisqu'elle permet de bloquer tous types de contenus et que, par conséquent, elle empêche aussi le chargement des services de calcul d'empreinte (fingerprinting).

Une solution rapide de « cookie consent » consiste à vérifier au moment du traitement d'une requête si le consentement a été obtenu et d'adapter la CSP en fonction. Si le consentement n'a pas été obtenu, la CSP n'autorisera que les ressources provenant de l'éditeur et bloquera les ressources provenant des tiers. Si le consentement a été obtenu, la CSP usuelle du site sera chargée.

Le plus simple pour un éditeur est d'utiliser une simple balise JavaScript pour insérer un champ **http-equiv** dans l'entête HTML du document. Cela n'est pas forcément recommandé, car cela atténue la séparation entre la CSP et le contenu qu'elle protège [https://bugzilla.mozilla.org/show_bug.cgi?id=663570#c4]. Les quelques lignes ci-dessous devraient faire l'affaire :

```
if ( document.cookie.indexOf('hasConsent') < 0 ) {
 var hostname = window.location.hostname;
 if (hostname.indexOf("www.") ===0) hostname = hostname.substring(5);
 var meta = document.createElement('meta');
 meta.httpEquiv = "content-security-policy";
 meta.content = "script-src 'self' 'unsafe-inline' *." + hostname + ";
img-src *." + hostname + "";
 document.getElementsByTagName('head')[0].appendChild(meta);
 }
```

Une solution légèrement plus compliquée consiste à utiliser le header HTTP. Le code pour accomplir cela reste assez semblable, mais l'approche est plus spécifique au serveur que vous utilisez. Par exemple, si vous fonctionnez sur un environnement PHP, le code devrait ressembler à :

```
if(!isset($_COOKIE["hasConsent"])) {
 $allowed_hosts = "*.unsearcher.org";
 header("Content-Security-Policy: script-src 'self' 'unsafe-inline' " .
$allowed_hosts . "; img-src 'self' " . $allowed_hosts);
 }
```

4.2 Compatibilité

Le standard CSP est en cours de validation, mais n'est toujours pas accepté par tous les navigateurs. À l'heure actuelle, Firefox, Chrome et Safari supportent toutes les fonctionnalités requises, dont le support du tag `http-equiv`. Enfin, si Edge (le nouveau navigateur de Microsoft) supporte toutes ces fonctionnalités, actuellement ni l'entête HTTP ni l'entête HTML ne sont pris en compte par Internet Explorer.

4.3 La conformité en utilisant les CSP

Une fois qu'elles seront plus largement supportées, les CSP permettront de bloquer les contenus tiers tant que le consentement n'a pas été obtenu. Néanmoins, cette solution est principalement viable pour le premier affichage. En effet, bloquer de façon permanente les contenus tiers n'est pas toujours concevable. La solution consiste alors à indiquer durant le premier affichage comment bloquer les cookies tiers. C'est la page d'information que de nombreux sites fournissent. Toutefois, cette solution présente deux inconvénients majeurs. D'une part, elle suppose que les cookies sont la seule technique de traçage utilisée sur le site. Or, on voit de plus en plus apparaître d'outils s'appuyant sur des prises d'empreintes du terminal. D'autre part, dans l'hypothèse ou vous utilisez des *cookies first party*, par exemple si vous utilisez Google Analytics, il sera nécessaire de faire appel à un script tel que celui qui est décrit dans le premier chapitre.

DNT LA FIN DE L'ATTENTE ?

Un autre paramètre qu'il est important de considérer est l'entête « DoNotTrack » (DNT). La recommandation cookie indique que ce paramètre peut être utilisé pour exprimer le consentement ou l'opposition au dépôt de cookies. Par défaut, DNT n'est pas actif sur les navigateurs, c'est donc à l'internaute d'indiquer sa préférence. La seule exception reste Internet Explorer qui active l'envoi du signal DNT par défaut, choix sur lequel Microsoft est récemment revenu.

Tenir compte du header DNT, à l'exemple de Piwik, ainsi que du tag GA proposé sur le site de la CNIL est déjà largement encouragé. Mais cela n'a pas encore de caractère obligatoire, car la norme DNT n'est pas finalisée. Néanmoins, le standard devrait être publié cette année [http://www.w3.org/blog/news/archives/4814] et le support du standard défini par le W3C (qui va bien au-delà du simple envoi du signal DNT) par les différents navigateurs devrait vite arriver.

Il est important de noter que DNT permettra d'exprimer tant son opposition que son consentement, et que les éditeurs pourront toujours recourir à un mécanisme de demande de dérogation (*User Granted Exception*).

DNT permettra non seulement de ne plus avoir à recourir systématiquement au bandeau d'informations, mais aussi de déléguer une partie de la responsabilité de l'éditeur aux tiers qui déposent des cookies sur sa page. En effet, ces derniers seront aussi destinataires du header et seront en mesure de savoir quand ils ne sont pas autorisés à déposer des cookies.

COMMENT OBTENIR UN CONSENTEMENT VALABLE ?

Il n'est pas nécessaire que l'internaute clique sur « OK » ou ferme le bandeau pour que celui-ci disparaisse. En effet, le consentement se manifeste par une simple « action positive » de l'utilisateur. Comment traduire cette notion juridique en contrainte technique ? Dès la publication de la recommandation, il a été explicitement fait mention qu'un clic sur n'importe quel élément de la page (sauf le lien « En savoir plus » ou « Paramétrer vos cookies ») vaut consentement. Le défilement (scroll) fait aussi partie des actions qui valent une poursuite de la navigation. Mais celui-ci doit être assez significatif et ne doit pas être dû à une erreur de manipulation.

5 Comment tester que votre site est conforme ?

Une fois tout votre site configuré, il est nécessaire d'effectuer des tests pour vérifier que les différents scripts sont fonctionnels. Les utilisateurs se réfèrent souvent à des plugins comme Ghostery, afin de vérifier qu'aucun cookie n'est déposé avant que l'utilisateur ne consente. Malheureusement, si Ghostery est très efficace pour obtenir une estimation du nombre d'acteurs présents sur une page, il ne permet pas d'avoir une vue précise des cookies déposés et le nom d'un fournisseur de service comme Facebook sera susceptible d'apparaître même si aucun cookie n'est déposé dès lors qu'un appel vers la plateforme sera détecté.

Conclusion

Mettre un site en conformité est une tâche plus ou moins ardue en fonction des éléments tiers auxquels vos pages font appel. Pour les pages n'incluant que des outils de mesures d'audience ou des plugins sociaux, des solutions clés en main existent. Par contre, l'intégration de publicités, de vidéos ou de widgets va nécessiter de recourir à des outils généralement plus complexes. Heureusement, la démocratisation des outils de gestion de tags simplifie désormais grandement cette tâche et les récents standards du *World Wide Web Consortium* (W3C), que sont les *Content Security Policies* et *DoNotTrack*, devraient faciliter davantage le travail des éditeurs. ■

Note

*** Les avis, opinions et positions exprimées dans le présent article n'engagent que leur(s) auteur(s) et en aucun cas l'institution à laquelle ils appartiennent.**

DÉCOUVREZ NOS NOUVELLES OFFRES D'ABONNEMENTS !

PRO OU PARTICULIER = CONNECTEZ-VOUS SUR :

www.ed-diamond.com

LES COUPLAGES PAR SUPPORT :

VERSION
PAPIER
Retrouvez votre magazine favori en papier dans votre boîte à lettres !

VERSION
PDF
Envie de lire votre magazine sur votre tablette ou votre ordinateur ?

ACCÈS À LA
BASE DOCUMENTAIRE
Effectuez des recherches dans la majorité des articles parus, qui seront disponibles avec un décalage de 6 mois après leur parution en magazine.

Sélectionnez votre offre dans la grille au verso et renvoyez ce document complet à l'adresse ci-dessous !

Voici mes coordonnées postales :	
Société :	
Nom :	
Prénom :	
Adresse :	
Code Postal :	
Ville :	
Pays :	
Téléphone :	
E-mail :	

❑ Je souhaite recevoir les offres promotionnelles et newsletters des Éditions Diamond.
❑ Je souhaite recevoir les offres promotionnelles des partenaires des Éditions Diamond.

En envoyant ce bon de commande, je reconnais avoir pris connaissance des conditions générales de vente des Éditions Diamond à l'adresse internet suivante : **boutique.ed-diamond.com/content/3-conditions-generales-de-ventes** et reconnais que ces conditions de vente me sont opposables.

MISC
Multi-System & Internet Security Cookbook

Édité par Les Éditions Diamond
Service des Abonnements
B.P. 20142 - 67603 Sélestat Cedex
Tél. : + 33 (0) 3 67 10 00 20
Fax : + 33 (0) 3 67 10 00 21

Vos remarques :

.....

.....

VOICI TOUTES LES OFFRES COUPLÉES AVEC MISC !

POUR LE PARTICULIER ET LE PROFESSIONNEL ...

Prix TTC en Euros / France Métropolitaine

CHOISISSEZ VOTRE OFFRE !

Prix en Euros / France Métropolitaine

Offre	SUPPORT / ABONNEMENT	PAPIER Réf	PAPIER Tarif TTC	PAPIER + PDF (PDF 1 lecteur) Réf	PAPIER + PDF Tarif TTC	PAPIER + BASE DOCUMENTAIRE (1 connexion BD) Réf	PAPIER + BASE DOCUMENTAIRE Tarif TTC	PAPIER + PDF + BASE DOCUMENTAIRE (PDF 1 lecteur + 1 connexion BD) Réf	PAPIER + PDF + BASE DOCUMENTAIRE Tarif TTC
MC	6n° MISC	☐ MC1	42,-	☐ MC12	62,-	☐ MC13	99,-	☐ MC123	111,-
MC+	6n° MISC + 2n° HS	☐ MC+1	54,-	☐ MC+12	81,-	☐ MC+13	103,-	☐ MC+123	130,-
LES COUPLAGES « LINUX »									
B	6n° MISC + 11n° GLMF	☐ B1	100,-	☐ B12	147,-	☐ B13	233,-	☐ B123	280,-
B+	6n° MISC + 2n° HS + 11n° GLMF + 6n° HS	☐ B+1	172,-	☐ B+12	248,-	☐ B+13	300,-	☐ B+123	381,-
C	6n° MISC + 6n° LP + 11n° GLMF	☐ C1	135,-	☐ C12	197,-	☐ C13	312,-	☐ C123	374,-
C+	6n° MISC + 2n° HS + 6n° LP + 3n° HS + 11n° GLMF + 6n° HS	☐ C+1	236,-	☐ C+12	339,-	☐ C+13	403,-	☐ C+123	516,-
LES COUPLAGES « EMBARQUÉ »									
E	6n° MISC + 6n° HK* + 4n° OS	☐ E1	105,-	☐ E12	158,-	☐ E13	179,-*	☐ E123	232,-*
E+	6n° MISC + 2n° HS + 6n° HK* + 4n° OS	☐ E+1	119,-	☐ E+12	179,-	☐ E+13	193,-*	☐ E+123	253,-*
LES COUPLAGES « GÉNÉRAUX »									
H	6n° MISC + 6n° HK* + 6n° LP + 11n° GLMF + 4n° OS	☐ H1	200,-	☐ H12	300,-	☐ H13	402,-*	☐ H123	499,-*
H+	6n° MISC + 2n° HS + 6n° HK* + 6n° LP + 3n° HS + 11n° GLMF + 6n° HS + 4n° OS	☐ H+1	301,-	☐ H+12	452,-	☐ H+13	493,-*	☐ H+123	639,-*

Les abréviations des offres sont les suivantes : LM = GNU/Linux Magazine France | HS = Hors-Série | LP = Linux Pratique | OS = Open Silicium | HC = Hackable

* HK : Attention : La base Documentaire de Hackable n'est pas incluse dans l'offre.

N'hésitez pas à consulter les détails des offres ci-dessus sur : www.ed-diamond.com !

LE REAL TIME BIDDING (RTB) OU COMMENT VENDRE LES ESPACES PUBLICITAIRES ET LES PROFILS AUX ENCHÈRES

Claude Castelluccia, Inria

mots-clés : *PUBLICITÉ EN LIGNE / ÉCONOMIE / DONNÉES PERSONNELLES / RTB / PROFILAGE / COOKIE MATCHING*

Les systèmes d'enchères en Temps réel (Real-Time Bidding ou RTB en anglais) sont en pleine croissance et devraient représenter plus de 25% des ventes totales d'affichage publicitaire aux États-Unis en 2015 contre 10% en 2011. Ils permettent de vendre aux enchères, et en temps réel, les espaces publicitaires des sites Web aux plus offrants. Les publicitaires peuvent ainsi acheter les espaces publicitaires en fonction de la taille des bannières, du contenu des pages web ou encore des profils des visiteurs. Cet article explique le fonctionnement et l'économie des systèmes de RTB. Il montre également comment ils favorisent les échanges des données personnelles et le profilage des internautes.

1 L'évolution de la publicité en ligne

La publicité en ligne est très répandue sur le Web et apporte des revenus substantiels à une majorité d'entreprises de l'Internet. Elle constitue un marché de plusieurs milliards d'euros par an, en constante progression. Par conséquent, des méthodes de plus en plus sophistiquées, souvent basées sur une analyse complexe des données des utilisateurs, ont été développées pour améliorer leur efficacité (meilleur ciblage et donc rentabilité).

1.1 La publicité ciblée

Contrairement à la publicité «classique», dans les magazines ou à la télévision, la publicité en ligne peut être plus facilement ciblée sur les centres d'intérêts, caractéristiques (âge, sexe...), comportements ou localisation des internautes. Depuis son introduction dans les années 1990, la publicité ciblée a évolué rapidement et représente aujourd'hui une part importante de la publicité en ligne. Pour les publicitaires, les avantages de la publicité ciblée sont multiples : une annonce personnalisée a plus de chances de susciter l'intérêt de l'internaute, et de provoquer un acte d'achat, augmentant ainsi les profits. De plus, les internautes reçoivent des annonces plus pertinentes et probablement moins ennuyantes.

Malheureusement, ce ciblage implique souvent le suivi des sites visités par les internautes et leur profilage par de nombreuses entités tierces, constituant ainsi un système de surveillance très élaboré. Le développement de « Data brokers », qui collectent les données personnelles des internautes pour les revendre, est une conséquence directe du développement de la publicité ciblée sur Internet.

1.2 La publicité aux enchères

Un autre développement important de ces dernières années de la publicité en ligne est l'apparition des systèmes d'enchères en temps réel.

Les enchères en Temps réel (*Real-Time Bidding* ou RTB en anglais) permettent d'améliorer la liquidité du marché de la publicité en ligne. Ces systèmes consistent

à vendre aux enchères, et en temps réel, les espaces publicitaires des sites Web aux plus offrants. Les publicitaires peuvent acheter les espaces publicitaires en fonction de la taille des bannières, du contenu des pages web ou encore des profils des visiteurs.

Cependant, comme nous allons le montrer dans la suite de cet article, les systèmes de RTB favorisent également les échanges des données personnelles et augmentent le profilage des internautes.

2 La publicité aux enchères temps réel

Nous décrivons ici le mécanisme de *DoubleClick* **[1]** qui est probablement le plus représentatif. D'autres systèmes, comme celui de *Pulse Point* **[2]**, ou l'initiative *OpenRTB* **[3]** qui vise à normaliser le RTB sont très similaires.

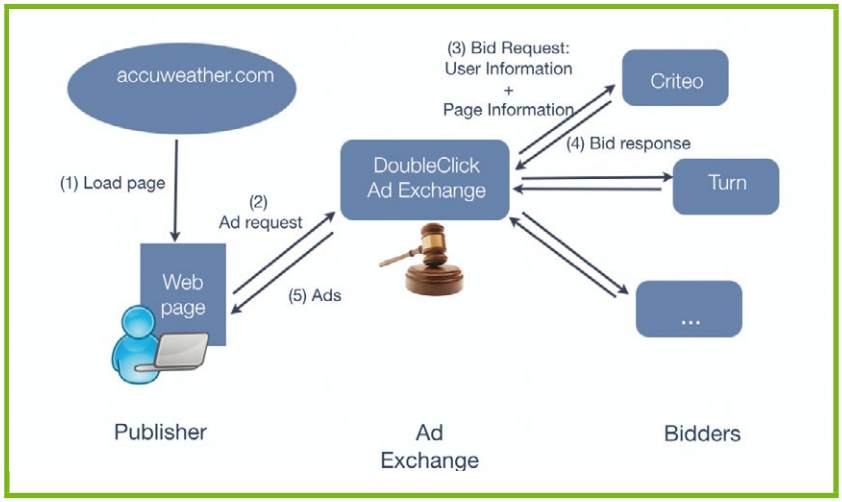

Figure 1 : Enchère avec le système RTB de DoubleClick. Dans cet exemple, un utilisateur se connecte sur le site www.accuweather.com qui contient un espace publicitaire appartenant à DoubleClick. Le navigateur de l'utilisateur envoie alors une « requête publicitaire » (message2) à DoubleClick. DoubleClick génère alors une « requête d'enchère » et l'envoie à des publicitaires qui répondent avec une « enchère » (message4).

Les architectures des systèmes RTB sont très complexes. Dans leur forme la plus simple, ils sont constitués de quatre acteurs principaux : 1) **les éditeurs de contenus** (par exemple www.accuweather.com), possédant des sites Web qui affichent des annonces, 2) **les échangeurs de publicités**, *Ad Exchange* en anglais (par exemple DoubleClick), intermédiaires qui permettent des transactions publicitaires entre les éditeurs et les publicitaires, 3) **les publicitaires** (par exemple *Criteo*, *Turn*), qui sont de grandes agences de publicité en ligne représentant des annonceurs, et enfin 4) **les annonceurs** (par exemple hotels.com), qui veulent faire de la publicité et vendre leurs produits en ligne. Une publicité affichée sur un site Web visité par un utilisateur, suite à un RTB, est dénommée « impression d'annonce ».

Les systèmes de RTB se complexifient avec l'apparition d'acteurs supplémentaires, tels que les DSP (*Demand Side Platforms*) qui permettent aux publicitaires d'enchérir sur plusieurs échangeurs, des SSP (*Supply Side Platforms*) qui permettent aux éditeurs de vendre leurs espaces à plusieurs échangeurs, et les DX (*Data Exchanges*) qui fournissent des données sur les utilisateurs. Par souci de clarté, cet article ne considère que le modèle simplifié décrit précédemment.

Les systèmes RTB opèrent comme suit : imaginons un site Web, par exemple www.lemonde.fr, qui possède un espace publicitaire appartenant à un échangeur, par exemple DoubleClick. Lorsqu'un utilisateur visite www.lemonde.fr, une requête HTTP est envoyée à DoubleClick. Cet échangeur va alors vendre l'espace publicitaire aux enchères en envoyant à ses partenaires publicitaires une requête d'enchère (*bid request*). Comme nous le verrons plus loin, cette requête contient un certain nombre d'informations sur la page visitée et sur l'utilisateur. Chaque publicitaire étudie alors la requête et répond éventuellement avec une offre contenant un prix. Le publicitaire le plus offrant gagne l'enchère, et peut alors publier sa publicité sur le site visité. Le publicitaire paye le montant de cette transaction à l'échangeur, qui en reverse une partie à l'éditeur.

L'ensemble de ce processus se produit généralement en moins de 100 ms (Figure 2).

Dans ces systèmes, un publicitaire (par exemple Criteo) achète les impressions publicitaires à l'échangeur au prix de l'enchère, mais facture l'annonceur uniquement si l'utilisateur clique sur la publicité. Il est donc essentiel pour le publicitaire de choisir la publicité qui va maximiser la probabilité que l'utilisateur clique ! D'où l'importance du profilage...

Il faut noter que le profilage des utilisateurs est profitable à tous les acteurs du RTB. En effet, plus l'utilisateur est profilé plus les enchères vont être nombreuses et élevées, et plus l'échangeur est gagnant. De même, plus l'utilisateur est profilé plus la probabilité que l'utilisateur clique est grande, ce qui est bénéfique pour le publicitaire. Finalement, plus la publicité est intéressante plus la probabilité que l'utilisateur achète le produit est grand ! C'est donc un système « gagnant-gagnant », sauf pour les utilisateurs qui se voient épiés, tracés et profilés en permanence ! En effet, comme décrit dans la section qui suit, le RTB met en place des systèmes de traçage renforcés et augmente la surveillance des internautes.

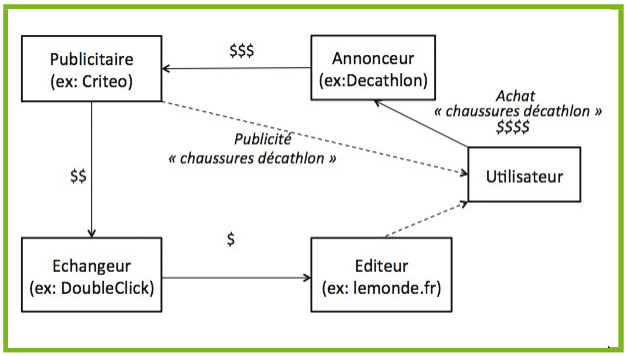

Figure 2 : Modèle économique simplifié des systèmes RTB. Dans cet exemple, l'utilisateur se connecte sur le site www.lemonde.fr. L'échangeur DoubleClick, qui possède un espace publicitaire sur www.lemonde.fr émet une enchère qui est gagnée par Criteo. Criteo qui a été payé par l'annonceur Décathlon pour une campagne pour ses chaussures, envoie la publicité à l'utilisateur qui, d'après son profil apprécie la marche en montagne. Criteo paye le prix de la publicité gagnée à DoubleClick, qui en garde une partie, et reverse le reste à l'éditeur, www.lemonde.fr. L'utilisateur achète éventuellement les chaussures chez Décathlon.

3 Le profilage des utilisateurs

Les systèmes de publicités en ligne « standards » reposent sur des techniques de profilage de bout en bout, comme les cookies, les scripts, les empreintes digitales (*fingerprinting*), qui sont bien connus et ont été largement étudiés **[7]**. En plus de ces techniques, les systèmes d'enchères en temps réel mettent en œuvre des mécanismes dans l'infrastructure, donc moins visibles, qui facilitent et amplifient le profilage des utilisateurs. Ils permettent souvent aux publicitaires de construire des profils des utilisateurs sans avoir à installer de « traceurs ».

En effet, comme illustré par la figure 3, les requêtes d'enchères, qui sont envoyées aux partenaires de l'échangeur à chaque enchère, contiennent généralement des informations telles que le cookie de l'utilisateur de l'échangeur, par exemple « UID= aaa », et le contexte de visite de l'utilisateur, y compris les informations suivantes : l'adresse URL du site Web visité, les catégories du site, les trois premiers octets de l'adresse IP de l'utilisateur, des informations diverses concernant le navigateur et l'utilisateur **[4, 5]**.

Chaque publicitaire partenaire peut donc savoir, juste en écoutant les requêtes consécutives accumulées pour chaque utilisateur identifié par son cookie « UID= aaa», la liste des sites qu'il a visité et ainsi construire son profil. Nos travaux de recherche ont montré que les échangeurs (par exemple DoubleClick) diffusent les sites visités par l'utilisateur à un nombre considérable de publicitaires, et que certains d'entre eux peuvent découvrir jusqu'à 27% des historiques des utilisateurs grâce à ce mécanisme, sans « traceur » **[6]**.

De plus, les systèmes de RTB utilisent le mécanisme de « Cookie Matching (CM) » qui permet à l'échangeur de synchroniser son cookie avec celui des publicitaires, et ainsi de contourner le « same-origin-policy ». En synchronisant leurs cookies, ils peuvent améliorer le profilage des utilisateurs en « combinant » leurs profils respectifs. En effet, un publicitaire peut avoir deux profils pour un même utilisateur : celui qu'il a construit lui-même, à partir de divers traceurs qu'il a installé sur divers sites, et indexé par son cookie « UID=bbb », et celui qu'il a construit grâce au RTB, comme décrit précédemment, indexé par « UID=aaa ». Grâce au CM, il pourra apprendre que les cookies « UID=aaa » et « UID=bbb » font référence au même utilisateur (ou plutôt au même navigateur) et ainsi « combiner » les deux profils en un profil plus précis.

```
id: "Mv\2005\000\017.\001\n\345\177\307X\200M8"
ip: "\314j\310"
user_agent: "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US) AppleWebKit/534.13 (KHTML, like Gecko)
Chrome/9.0.597.107 Safari/534.13,gzip"
url: "http://www.example.com/"
detected_language: "en"
detected_vertical {
  id: 22
  weight: 0.67789277
}
detected_vertical {
  id: 355
  weight: 0.32210726
}
cookie_version: 1
google_user_id: "CAESEIcS1pC2TBvb-4SLDjMqsY9"
seller_network_id: 1
publisher_settings_list_id: "\357\237V\206)\231\3125%|$\032\"
vertical_dictionary_version: 2
timezone_offset: -300
cookie_age_seconds: 7685804
```

Figure 3 : Exemple de requête RTB envoyée par un échangeur aux publicitaires. Ces requêtes contiennent les informations du navigateur, l'URL du site visité, la taille de l'espace publicitaire, le cookie Google.

La figure 4, page suivante, montre un exemple d'échange de « Cookie Matching » qui a lieu après une vente aux enchères, effectuée par **Adexchange.com** et gagnée par **Bidder.com**. On observe que l'échangeur de publicités, **Adexchange.com**, envoie un script ou une instruction de redirection au navigateur de l'utilisateur indiquant l'adresse url de la publicité à télécharger, **cm.bidder.com**. Cette redirection contient aussi le cookie de l'utilisateur « UID=aaa ». Le navigateur de l'utilisateur se connecte

alors sur cm.bidder.com et envoie son cookie, « UID=bbb ». Lorsque le publicitaire, bidder.com, reçoit ce message, il délivre la publicité et apprend au passage que les deux cookies « aaa » et « bbb » appartiennent au même utilisateur.

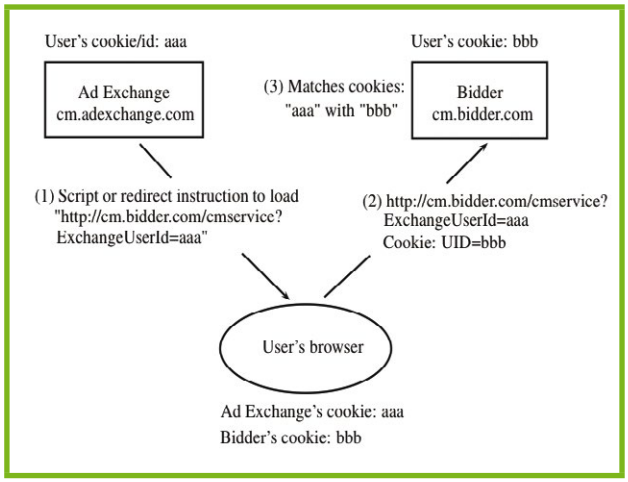

Figure 4 : Exemple de Cookie Matching.

Nos travaux ont montré que les CM arrivent très fréquemment et sont réalisés par un grand nombre d'entités; certains d'entre eux synchronisent leurs cookies sur une grande proportion des profils, jusqu'à 91% des profils selon notre étude **[6]**. De plus, nous avons mesuré que Facebook et AppNexus sont présents sur de nombreux sites Internet et peuvent reconstituer respectivement, en moyenne, 31,55% et 17,4% des historiques Web des internautes. En fusionnant leurs profils, leurs couvertures moyennes peuvent passer à 39,35% !

4 Le prix des publicités ciblées

Les systèmes de RTB sont en pleine croissance et devraient représenter plus de 25% des ventes totales d'affichage publicitaire aux États-Unis en 2015 contre 10% en 2011 **[10]**.

Comme nous l'avons décrit précédemment, les systèmes de RTB permettent aux publicitaires d'acheter des espaces publicitaires aux enchères. Lorsqu'un publicitaire gagne une enchère, il obtient non seulement un espace publicitaire, mais également la possibilité de faire un « Cookie Matching », et donc d'améliorer le profil de l'utilisateur. Mais quel prix un publicitaire est-il prêt à payer pour gagner aux enchères, et ce prix varie-t-il d'un internaute à un autre ? En d'autres termes, est-ce que certains internautes sont plus « bankables » que d'autres?

Comme expliqué précédemment, les publicitaires doivent payer l'impression des publicités qu'ils ont gagnées. Mais comme RTB utilise le système d'enchère

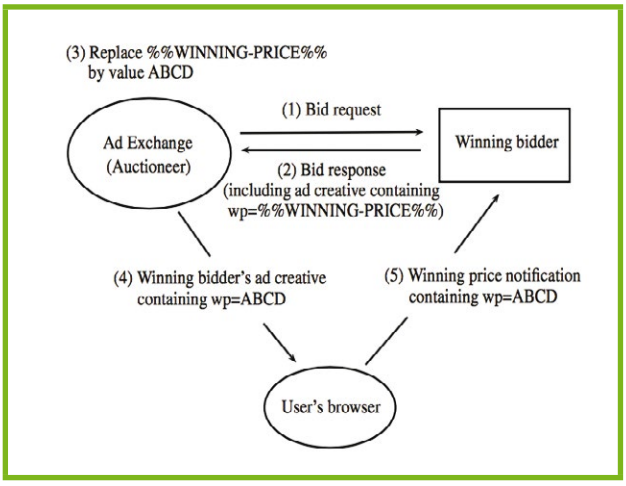

Figure 5 : Exemple de fuite du prix d'une publicité par le mécanisme de Cookie Matching. Le prix de l'enchère, « ABCD », est communiqué par l'échangeur au publicitaire par l'intermédiaire du navigateur de l'utilisateur (messages 4 et 5).

Vickrey **[11]**, le gagnant de l'enchère doit payer les 2eme prix le plus élevé et non celui qu'il a proposé. Par conséquent, lorsqu'un publicitaire gagne une enchère, l'échangeur lui communique le prix à payer par l'intermédiaire d'une macro `WINNING_PRICE`, qui comme le montre la figure 5, passe par le navigateur de l'utilisateur. Or, il s'avère que certains publicitaires, peu regardants, transmettent ces prix en clair! Ils sont donc accessibles !

Nous avons donc développé une extension Firefox et Chrome qui analyse les échanges entre les échangeurs et publicitaires et récupère le prix des publicités vendues aux enchères. Cette extension calcule alors le prix moyen de chaque utilisateur et envoie cette information, anonymisée, à l'un de nos serveurs. Un utilisateur qui a installé notre extension, peut découvrir le prix moyen de ses publicités, ainsi que son classement parmi tous les utilisateurs de notre extension en se connectant sur notre site yourvalue.inrialpes.fr (figure 6). Il peut donc voir si son profil est plus ou moins « bankable » que celui des autres internautes. Notre système utilise également le prix moyen ainsi calculé et le nombre de publicités qu'un utilisateur rencontre quotidiennement pour calculer le prix journalier déboursé par les publicitaires pour chaque utilisateur.

Nous avons distribué notre extension sur l'Internet, et recueilli les prix moyens de plusieurs volontaires avec différents profils. Le détail de nos travaux et résultats est disponible dans notre article scientifique : **[6]**.

Nos résultats montrent que le prix moyen par espace publicitaire est très bas : en moyenne 0,0005 dollars ! Les prix moyens changent également en fonction du pays du destinataire. En effet, les prix aux États-Unis (en moyenne $0,00069) sont visiblement plus élevés que ceux en France ($0,00036) et au Japon ($0,00024).

Figure 6 : Interface du service « How Much Are You Worth? ». Cet exemple montre que le prix moyen des publicités ciblées de l'internaute est de $0.001035 et que sur les 257 utilisateurs du service, cet utilisateur est classé 122eme en terme de valeur de son profil.

Les résultats confirment également que certains internautes sont plus « bankables » que d'autres et que les prix moyens de leurs publicités sont plus élevés. En effet, sur plus de 200 utilisateurs de notre extension, les prix moyens varient entre $0.01 et $0.000038 (http://yourvalue.inrialpes.fr).

Ces prix dépendent à la fois de la précision des profils, mais aussi des catégories les constituants. En effet, les catégories « Restaurant » ou « Shopping » semblent deux fois plus prisées, avec un prix moyen de $0.00068 et $0.00059, que les catégories « Humor » ou « Sports ».

Enfin, les publicitaires semblent avoir des stratégies d'enchère différentes : certains semblent cibler les profils les plus chers, d'autres au contraire les profils à coût moyen ou faible. L'économie des publicités en ligne est très complexe et dépend d'une multitude de paramètres.

Conclusions

L'impact sur la vie privée de la publicité ciblée a longtemps été une source de controverses et de débats. La publicité ciblée apporte des avantages économiques énormes : elle fournit un moyen pour les annonceurs de mieux atteindre leurs segments de marché et elle augmente les revenus des éditeurs, et par conséquent, des publicitaires. Cependant, elle envahit la vie privée des internautes qui se retrouvent tracés et profilés en permanence. Comme nous l'avons montré, les systèmes RTB permettent une surveillance encore plus avancée des internautes. Cette surveillance est omniprésente, « invisible » et difficilement contrôlable, car elle est mise en œuvre par l'infrastructure.

Les publicitaires en ligne se défendent souvent contre cette accusation en argumentant que les données qu'ils collectent, souvent les listes des sites visités par chaque internaute, sont anonymes et jamais associées à des noms. Ces arguments sont très faibles, car il a été montré que la liste des sites visités par un internaute est, dans la majorité des cas, unique et est donc « identifiante » **[8]**. Par ailleurs, il existe plusieurs solutions pour réidentifier ces données. Par exemple, un publicitaire peut récupérer le nom d'un utilisateur à partir de sites qui demandent aux utilisateurs de s'authentifier, comme les réseaux sociaux **[7]**. De plus, les techniques de « Cookie Matching », décrites dans cet article, qui facilitent les échanges des données personnelles entre entités tierces, favorisent grandement cette réidentification.

Finalement, même si ces données étaient réellement anonymes, le fait de connaître le profil d'un utilisateur, même sans connaître son identité, permet plus facilement de le manipuler en lui envoyant des messages très ciblés à des moments précis. Cette manipulation « psychologique » peut éventuellement être utilisée pour influencer les internautes sur leurs actes d'achat ou même sur leurs votes **[9]** !

À un moment où les débats sur la nouvelle loi sur le renseignement se terminent, il paraît urgent d'ouvrir un débat public sur l'éthique et la légalité de la surveillance effectuée par les publicitaires en ligne. Cette surveillance est aussi dangereuse, voire plus dangereuse, que la surveillance étatique, car elle est effectuée par des entreprises privées, souvent étrangères, sans aucun contrôle ni recours.

Des outils, comme Ghostery **[13]** ou AdBlock **[14]**, permettent de bloquer le téléchargement et/ou l'affichage de certaines publicités, et généralement tout le traçage qui en résulte. Cependant, ils remettent en cause le modèle économique d'Internet, ce qui n'est pas toujours acceptable. Il devient urgent de réfléchir à de meilleures solutions pour limiter et mieux contrôler cette surveillance.

Il faudrait aussi développer des systèmes de publicité ciblée qui respectent la vie privée des utilisateurs en adoptant une approche « Privacy-by-Design », car répétons-le encore une fois : ce n'est pas le ciblage qui est en cause ici, mais bien le traçage et le profilage! En effet, la plupart des internautes ne sont pas contre les publicités ciblées, mais souhaitent mieux contrôler les informations qui sont collectées par les publicitaires et les autres entités tierces. ■

Retrouvez toutes les références accompagnant cet article sur http://www.miscmag.com/.

LE FINGERPRINTING : UNE NOUVELLE TECHNIQUE DE TRAÇAGE

Benoit Baudry – benoit.baudry@inria.fr – Chercheur à l'INRIA Rennes
Pierre Laperdrix – pierre.laperdrix@insa-rennes.fr – Doctorant à l'INSA de Rennes

mots-clés : *BROWSER FINGERPRINTING / EMPREINTE DE NAVIGATEUR / TRAÇAGE / VIE PRIVÉE / JAVASRIPT / FLASH / HTML*

Le « browser fingerprinting » désigne l'activité de collecte par un navigateur d'un certain nombre d'informations sur l'appareil d'un internaute pour bâtir une empreinte (fingerprint). De nombreuses études ont montré que cette empreinte est unique dans la très grande majorité des cas et évolue très lentement. Il est ainsi possible de l'utiliser pour tracer les internautes, sans laisser aucune trace sur l'appareil.

1 Présentation du Fingerprinting

En 2010, Peter Eckersley de l'Electronic Frontier Foundation révélait la possibilité d'exploiter le « browser fingerprinting » pour tracer les internautes. Pour illustrer ce phénomène, il a lancé le site https://panopticlick.eff.org/ sur lequel il a installé un script très simple qui récupère 8 attributs du navigateur et de l'environnement du visiteur du site. Il a montré que 94% des 500 000 empreintes récoltées au moment de l'étude étaient uniques, pouvant ainsi être exploitées pour tracer les internautes. Il a aussi démontré qu'il était possible de suivre une même empreinte dans le temps grâce à l'identification de simples évolutions. Depuis lors, de nombreux travaux ont montré qu'il est possible de récolter d'autres attributs pour enrichir l'empreinte et la rendre encore plus identifiable. Des sociétés commerciales ont d'ailleurs commencé à exploiter les empreintes pour tracer les internautes.

1.1 Le browser fingerprinting en pratique

Pour mieux comprendre le phénomène du browser fingerprinting, l'importance des différents attributs ainsi que les différences entre appareils mobiles et ordinateurs (portables ou de bureau), nous avons déployé un script de fingerprinting sur le site https://amiunique.org/ **[1]**. Le tableau 1 présente un exemple d'empreinte récupérée sur AmIUnique. Nous récoltons 13 attributs du navigateur et de son environnement. La première source d'informations provient des en-têtes HTTP que le client envoie systématiquement au serveur, dès qu'il accède à un site. Par exemple, le user agent révèle le type du navigateur, sa version ainsi que le système d'exploitation utilisé par le client, et l'en-tête « Content language » indique la langue demandée pour la page web.

Il est ensuite nécessaire d'exécuter un script pour récupérer certains de ces attributs. Une partie est accessible par les nombreuses APIs fournies par le moteur JavaScript du navigateur (par exemple, les objets JS « window.screen » et « window.navigator » fournissent une grande quantité d'informations à propos du système). Une autre partie des attributs provient des plugins utilisés par le navigateur, qui fournissent des APIs donnant accès à d'autres informations sur le système d'exploitation. Par exemple, le plugin Flash peut fournir la liste des polices installées sur l'appareil grâce à un simple appel de fonction.

Le code ci-dessous est un extrait du script de fingerprinting. On remarque que certains attributs de l'empreinte sont systématiquement fournis pas le client (en-têtes HTTP) et que d'autres sont récupérés directement par une simple requête sur les APIs Flash et JS (la plateforme ou la résolution de l'écran). Enfin, certains attributs ne peuvent être récupérés qu'avec une connaissance plus fine des APIs (WebGL Renderer).

Extrait du script de fingerprinting utilisé sur AmIUnique.org.

Attribut	Source	Valeur
User agent	En-tête HTTP	Mozilla/5.0 (X11; Linux x86 64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/ 41.0.2272.118 Safari/537.36
Accept	En-tête HTTP	text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Content encoding	En-tête HTTP	gzip, deflate, sdch
Content language	En-tête HTTP	en-us,en;q=0.5
Liste des plugins	JavaScript	Plugin 1: Chrome PDF Viewer. Plugin 2: Chrome Remote Desktop Viewer...
Fuseau horaire	JavaScript	-60 (UTC+1)
Do Not Track	En-tête HTTP/JavaScript	Oui
Résolution d'écran	JavaScript	1920x1200x24
Liste des polices	Flash	Abyssinica SIL,Aharoni CLM,AR PL UM-ing CN,AR PL UMing HK,AR PL UMing TW...
Plateforme	Flash	Linux 3.19.1-201.fc21.x86 64
WebGL Vendor	JavaScript	NVIDIA Corporation
WebGL Renderer	JavaScript	GeForce GTX 650 Ti/PCIe/SSE2

Tableau 1 : Exemple d'un fingerprint.

```
var platform = window.navigator.platform;
var cookieEnabled = window.navigator.cookieEnabled? "yes" : "no";
var timezone = new Date().getTimezoneOffset();
var resolution = window.screen.width+"x"+window.screen.
height+"x"+window.screen.colorDepth;

try {
    localStorage.fp = "test";
    domLocalStorage = "";
    if (localStorage.fp == "test") {
       domLocalStorage = "yes";
    } else {
       domLocalStorage = "no";
    }
} catch (ex) {
    domLocalStorage = "no";
}

try {
    canvas = document.createElement("canvas");
    canvas.height = 60;
    canvas.width = 400;
    canvasContext = canvas.getContext("2d");
    canvas.style.display = "inline";
    canvasContext.textBaseline = "alphabetic";
    canvasContext.fillStyle = "#f60";
    canvasContext.fillRect(125, 1, 62, 20);
    canvasContext.fillStyle = "#069";
    canvasContext.font = "11pt no-real-font-123";
    canvasContext.fillText("Cwm fjordbank glyphs vext quiz,
\ud83d\ude03", 2, 15);
      canvasContext.fillStyle = "rgba(102, 204, 0, 0.7)";
    canvasContext.font = "18pt Arial";
   canvasContext.fillText("Cwm fjordbank glyphs vext quiz,
\ud83d\ude03", 4, 45);
    canvasData = canvas.toDataURL();
} catch(e){
    canvasData = "Not supported";
}

var canvas = document.createElement('canvas');
var ctx = canvas.getContext("webgl") || canvas.
getContext("experimental-webgl");
if(ctx.getSupportedExtensions().indexOf("WEBGL_debug_renderer_
info") >= 0) {
        webGLVendor = ctx.getParameter(ctx.getExtension('WEBGL_
debug_renderer_info').UNMASKED_VENDOR_WEBGL);
        webGLRenderer = ctx.getParameter(ctx.getExtension('WEBGL_
debug_renderer_info').UNMASKED_RENDERER_WEBGL);
} else {
        webGLVendor = "Not supported";
        webGLRenderer = "Not supported";
}
```

1.2 Browser fingerprinting : un effet de bord de la richesse des navigateurs

Nous voyons que le principe du browser fingerprinting est très simple : récupérer, grâce à un simple script, quelques informations concernant l'appareil sur lequel est installé le navigateur afin d'en constituer une empreinte. Néanmoins, chacune de ces informations semble très banale et très commune, et on peut se demander comment cette empreinte peut être exploitée à des fins de traçage.

Il est important de comprendre que si un cookie est un objet précis, installé sur les machines clientes dans une démarche explicite de garder, d'analyser et de tracer des historiques de navigation, le phénomène du browser fingerprinting est beaucoup plus diffus. La technologie a des contours beaucoup plus flous que la technique des cookies.

La possibilité d'exploiter les empreintes à des fins de traçage est un « accident » rendu possible par l'apparition de nouvelles technologies, indépendantes de la volonté des sites marchands. Au cours des années, les navigateurs se sont enrichis pour permettre plus

d'interactions avec les usagers, pour permettre aux internautes de personnaliser leur navigateur, et pour permettre l'affichage de contenus multimédias très riches. Cet enrichissement des fonctionnalités des navigateurs s'est fait grâce à deux technologies : des architectures à plugins qui permettent à chaque internaute de spécialiser son navigateur ; des APIs très riches qui permettent aux développeurs de proposer du contenu web dynamique, attractif et adapté aux environnements des internautes (par exemple, le site est automatiquement affiché dans la langue choisie par l'internaute).

L'apparition du browser fingerprinting comme technique de traçage est un effet de bord direct de cet enrichissement des navigateurs. D'une part, chaque internaute personnalise son navigateur et son environnement (différents systèmes d'exploitation, différents appareils, etc.) et, par conséquent, l'empreinte d'un navigateur est très probablement différente de toutes les autres. D'autre part, l'enrichissement constant des APIs donne accès à de plus en plus d'informations, permettant de bâtir des empreintes de plus en plus riches, qui ont une probabilité d'autant plus grande d'être uniques.

La conséquence majeure du détournement des fonctionnalités du navigateur à des fins de constitution d'une empreinte est qu'il extrêmement difficile de détecter une activité de fingerprinting de la part d'un site, et donc de contrôler ou d'empêcher une telle activité. Par exemple, si un internaute détecte qu'un script récupère la résolution de son écran, il est impossible de savoir si cette information est utilisée à des fins légitimes pour ajuster la page web à la taille de l'appareil ou si c'est un élément d'une signature plus grande, récoltée à des fins de traçage.

2 Analyse de la diversité des fingerprints

Nous discutons ici de la différence entre les attributs récoltés sur amiunique.org en termes d'identification, et nous distinguons l'importance de ces attributs entre les ordinateurs fixes ou portables et les appareils mobiles. Les mesures et observations présentées s'appuient sur l'analyse de 75 000 empreintes récoltées entre novembre 2014 et mai 2015.

2.1 Attributs discriminants

La liste des polices de caractères et celle des plugins constitue souvent les attributs les plus discriminants dans une empreinte. La liste des polices dépend directement du système d'exploitation et des logiciels installés. Il suffit d'avoir un logiciel qui est très peu utilisé ou d'avoir téléchargé et installé une police de caractères très spéciale depuis Internet pour avoir une empreinte très facilement reconnaissable. L'utilisation d'un système d'exploitation atypique ou peu répandu comme FreeBSD, Arch Linux ou openSUSE contribue aussi très fortement à avoir une empreinte unique. Pour ce qui est des plugins, de très nombreuses empreintes récoltées possèdent des plugins qui sont très peu répandus et qui rendent leur utilisateur directement identifiable (plus de 96% des plugins observés sur AmIUnique sont présents dans moins de 1% des empreintes récoltées). Les plugins recensés concernent tous les usages : des modules pour fournir une sécurité et une identification accrue sur certains sites comme le plugin Identity Protection Technology d'Intel à des plugins pour faciliter le lancement de certains jeux comme le plugin Uplay d'Ubisoft.

Ces derniers temps, nous pouvons observer une évolution parmi les empreintes récupérées sur des navigateurs Chrome. Il y a 2 ans, Google a décidé d'arrêter le support des plugins qui suivent l'architecture NPAPI car, selon eux, ils sont source « de bugs, de crashs, d'incidents de sécurité et de complexité de code » **[2]**. Depuis la version 43, sortie en avril 2015, les plugins NPAPI sont désactivés par défaut, et les plugins qui étaient utilisés par une très petite minorité d'internautes ne sont donc plus présents. L'impact que cette décision a sur le traçage par empreintes reste à évaluer, mais un changement est certainement à prévoir.

2.2 Analyse d'un attribut : le canvas fingerprinting

Fig. 1 : Test de l'API Canvas du navigateur en procédant au rendu d'une image en suivant les instructions présentes dans le deuxième bloc « try » de l'extrait de code de la partie 1.1.

Certains attributs peuvent donner des informations sur plusieurs couches du système. L'image générée en utilisant l'API Canvas en fait partie (exemple dans la figure 1). Tout d'abord, le rendu des 2 chaînes de caractères dans l'image peut varier d'un appareil à un autre. Dans le script de fingerprinting, nous demandons au navigateur d'utiliser une police qui n'existe pas. Cette demande spéciale se traduit dans le navigateur par l'utilisation d'une police de caractères dite de « fallback », c'est-à-dire une police qui est utilisée quand celle demandée n'existe pas. Selon le système d'exploitation et les polices installées, cette police de fallback n'est pas la même, et cette différence peut être utilisée pour distinguer deux appareils. Dans un second temps, le dernier caractère peut révéler de précieuses informations sur le système utilisé. C'est un « emoji », un pictogramme d'une émotion ou d'une action (à ne pas confondre avec les émoticônes qui ont le même but, mais qui sont représentées par une suite de caractères comme ":(" ou "<3"). Chaque emoji a son propre caractère Unicode **[3]**, et c'est à la charge des développeurs des polices de caractères de donner leur propre représentation de chaque emoji. Cela se traduit par une grande diversité d'emojis entre systèmes d'exploitation. On peut même observer des

différences entre constructeurs qui utilisent Android sur leurs smartphones. Les emojis dans la figure 2 illustrent cette diversité qui est récupérable par un script de fingerprinting et qui donne des informations sur l'appareil utilisé.

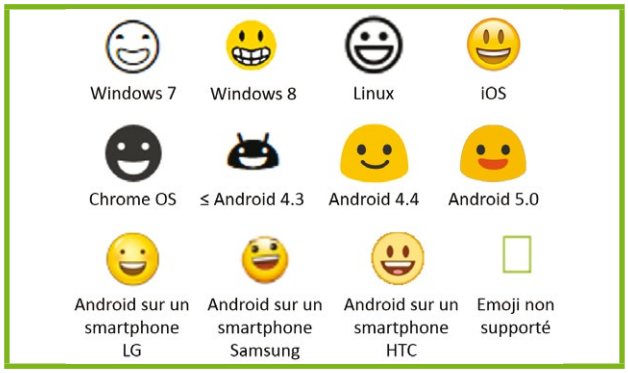

Fig. 2 : Emojis sur différents systèmes d'exploitation et sur différents appareils.

Enfin, comme démontré par Mowery et al. **[4]**, l'agencement des pixels dans l'image peut varier entre des systèmes ayant la même couche logicielle, mais du matériel différent. Selon la carte graphique, le processeur et les pilotes utilisés, une même image générée à partir d'un même ensemble de polices peut présenter des variations de quelques pixels qui sont détectables et donc utilisables pour du fingerprinting.

2.3 Comparaison PC/smartphone

Les smartphones et tablettes sont devenus les supports les plus répandus pour naviguer sur Internet. Une analyse des empreintes de ces appareils révèle qu'ils présentent moins de diversité sur beaucoup d'attributs, comparés à un ordinateur fixe ou portable. Ceci est une conséquence directe de l'absence de plugins. Cependant, le user agent est souvent plus riche et peut révéler des informations très précieuses sur le système d'un appareil. Par exemple, sur les systèmes Android, les mises à jour des firmwares sont directement fournies par le constructeur. Le navigateur indique alors le modèle exact du téléphone ainsi que la version exacte du firmware comme on peut le voir ci-dessous avec un Sony Xperia Z3 D5803 sous Android 4.4.4 :

Mozilla /5.0 (Linux; **Android 4.4.4; D5803** *Build /23.0.1. A .5.77) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/39.0.2171.93 Mobile Safari /537.36*

En revanche, sur le même système, si l'utilisateur décide de naviguer avec Firefox et non plus avec Chrome, beaucoup moins d'informations sont visibles :

*Mozilla /5.0 (***Android; Mobile** *rv :34.0) Gecko /34.0 Firefox /34.0*

Comme Firefox est installé à partir du Google Play Store, le user agent renvoyé est générique et n'a pas été compilé pour l'appareil concerné. Selon le navigateur utilisé, plus ou moins d'informations peuvent donc être visibles et exploitées dans une empreinte.

3 Le fingerprinting en pratique

3.1 Quelques exemples d'adoption

Comme toute technique de traçage, le fingerprinting peut être utilisé de manière constructive en fournissant des outils d'authentification avancée permettant de vérifier l'authenticité d'un appareil et de détecter des comportements anormaux. Nous n'avons malheureusement aujourd'hui pas d'exemples concrets d'une telle pratique même si certains sites web sensibles doivent très probablement s'appuyer sur de tels outils pour contrôler le trafic au sein de leurs services.

Le fingerprinting peut aussi rentrer dans le cadre d'une utilisation abusive, voire destructive en construisant des profils complets d'utilisateurs et en analysant leurs habitudes de navigation. De telles pratiques peuvent être considérées comme de sérieuses atteintes au respect de la vie privée des internautes.

En 2013, des chercheurs de Leuven en Belgique ont analysé les 10 000 sites les plus populaires pour déterminer la propagation du fingerprinting sur Internet **[5]**. Ils en ont conclu que cette technique était beaucoup plus répandue que les estimations des précédentes recherches sur le sujet. Plusieurs sociétés de publicité comme BlueCava **[6]**, AddThis **[7]** ou Revenviews **[8]** se sont déjà tournées vers le fingerprinting pour renforcer leur traçage et compléter l'usage des cookies.

Les géants du Web comme Google, Yahoo, Twitter ou Amazon n'indiquent pas dans leur politique de vie privée l'usage du fingerprinting, mais ils précisent qu'ils collectent des « données relatives à l'appareil utilisé » pour « améliorer les services proposés aux utilisateurs ». Google a même mis à jour sa politique de vie privée en juin 2015 pour indiquer l'usage de « technologies similaires » aux cookies et non plus « d'identifiants anonymes » **[9]**.

Enfin, la récolte d'informations sur la configuration d'un appareil peut mener à des attaques ciblées. Par exemple, si vous identifiez qu'une machine possède une version vulnérable du plugin Flash ou une version non patchée d'OpenSSL, une attaque peut être développée sur mesure pour cette machine.

Le fingerprinting est donc déjà largement utilisé sur Internet et c'est une technique en pleine expansion qui présente des atouts, mais aussi des inconvénients en matière de sécurité.

3.2 Des contre-mesures partielles

Il n'existe aujourd'hui aucune solution permettant d'empêcher toute tentative de collecte d'empreinte ou de rendre une empreinte anonyme. Néanmoins, il existe un ensemble de techniques qui permettent de cacher ou de modifier certains des attributs qui forment une empreinte.

3.2.1 Do Not Track : une opportunité manquée ?

En 2011, un en-tête HTTP appelé « Do Not Track » a été intégré dans les principaux navigateurs. Son objectif est simple : informer un site web de son souhait de ne pas être tracé. Cette solution est censée couvrir toutes formes de traçage existantes et futures (cookies, fingerprinting). Le problème est qu'un en-tête HTTP est un simple signal qui ne bloque ni la mise en place de cookies ni l'exécution de scripts de traçage. Ainsi, malgré des recommandations très claires de la part des autorités (par exemple, la CNIL « considère que lorsqu'un internaute décide d'activer une option de type « do not track », aucun profil ne devrait être réalisé sur cet internaute et sur son terminal »), la demande est ignorée par de très nombreux sites. Même les deux géants que sont Facebook et Google l'ignorent sous prétexte que l'appellation « Do Not Track » n'est pas claire pour les internautes **[10]**.

Yahoo a arrêté le support de cet en-tête en novembre 2014 à cause d'un manque d'efficacité et d'adoption par l'industrie du Web pour ensuite revenir sur sa décision un an plus tard à la suite d'un nouveau partenariat avec Mozilla **[11]**. De son côté, le moteur de recherche DuckDuckGo se distingue de ses concurrents par le fait qu'il ne collecte pas et qu'il ne partage pas d'informations personnelles **[12]**.

L'adoption d'un standard efficace contre le traçage sur Internet est souhaitée et demandée par les internautes. Cependant, les grands acteurs du Web freinent le développement d'un tel standard à cause des impacts économiques qui sont en jeu.

3.2.2 Modifier son fingerprint : une fausse bonne idée

Une idée pour se protéger du fingerprinting est de modifier la valeur des attributs renvoyés au serveur ou de masquer les informations discriminantes. Par exemple, il existe de nombreuses extensions pour modifier le user agent telles que « User agent switcher » **[13]** sous Chrome ou « Masking agent » **[14]** sous Firefox. L'usage de ces extensions pose deux problèmes.

Le premier vient du fait qu'en modifiant une partie des données renvoyées au serveur, l'utilisateur a une empreinte dite incohérente, c'est-à-dire une empreinte qui ne peut correspondre à aucun navigateur. Par exemple, un appareil qui dit être un iPhone avec une résolution d'écran de 1920x1080 est tout simplement impossible. Cette incohérence rend ainsi l'appareil détectable, car il sera un des seuls appareils dans le monde à mentir avec ces valeurs spécifiques.

Le deuxième problème vient du fait que, même si un individu cherche à cacher des informations comme le système d'exploitation de son appareil ou le navigateur utilisé, les moteurs JavaScript ont des APIs tellement riches qu'il est possible de vérifier la valeur d'un même attribut par de très nombreux moyens. Par exemple, il est possible de trouver la valeur du système d'exploitation dans le user agent renvoyé, dans l'attribut « platform » en JavaScript et en Flash, dans la liste et les extensions des fichiers des plugins utilisés, dans le type des emojis, etc.

Au final, mentir sur son empreinte est inefficace et rend l'utilisateur encore plus visible qu'avec les valeurs réelles.

3.2.3 Bloquer les scripts : utilisation d'extensions

La meilleure défense aujourd'hui contre le browser fingerprinting est de bloquer en amont les scripts qui sont utilisés pour récupérer des informations concernant votre appareil.

L'usage d'extensions comme AdBlock **[15]**, Ghostery **[16]**, Disconnect **[17]** ou Privacy Badger **[18]** est recommandé (plus d'extensions sur https://amiunique.org/tools). Fonctionnant sur le même principe qu'un antivirus, elles bloquent tous les scripts de fingerprinting qui sont présents dans leurs bases de données. Par exemple, sur le site LeMonde.fr, AdBlock et Ghostery bloquent plus d'une soixantaine de traqueurs (publicités, outils statistiques, boutons pour partager sur les réseaux sociaux, widgets...).

Pour encore plus de protection, l'extension NoScript **[19]** est très utile pour bloquer les scripts JavaScript dans une page web. Enfin, pour les personnes les plus soucieuses de leur vie privée, nous recommandons l'usage du navigateur Tor **[20]**. Le but de Tor est d'améliorer l'anonymat des internautes sur Internet en routant tout le trafic réseau de ses usagers à travers un réseau dédié. Les techniques de routage en oignon procurent une très grande confidentialité sur la provenance des paquets réseaux. Le navigateur Tor, basé sur de nombreuses modifications de Firefox, limite au maximum les capacités de traçage et fournit une empreinte unique (même si la réalité est un petit peu différente). Le système d'exploitation Tails **[21]**, qui embarque le navigateur Tor, va plus loin en modifiant le système d'exploitation pour ne laisser aucune trace sur l'ordinateur utilisé (aucune écriture disque, vidage de la mémoire...).

Conclusion

En conclusion de cet article, nous pouvons nous interroger sur les manières de limiter ou d'empêcher ce traçage par empreintes. Est-ce possible ou est-ce que la diversité toujours croissante des appareils pouvant se connecter à Internet annonce déjà un combat perdu d'avance ? Il est très clair qu'il est nécessaire que les navigateurs informent les serveurs d'une partie de leur configuration pour offrir un plus grand confort de navigation aux utilisateurs. Mais ne pourrait-on pas limiter la quantité d'informations discriminantes fournies pour qu'une empreinte ne soit tout simplement plus unique ? Dire qu'un téléphone est sous Android paraît tout à fait légitime, mais indiquer le modèle exact de son smartphone et la version précise de son firmware paraît extrêmement superflu et ouvre des portes à du traçage non voulu et à des attaques ciblées.

Il faut donc réussir à concevoir des navigateurs qui permettent de supporter une très grande variété de configurations tout en respectant la vie privée et en divulguant le moins d'informations possible sur le système utilisé. Limiter ou supprimer la liste des plugins et des polices de caractères, enlever les informations discriminantes et inutiles dans les en-têtes HTTP, fournir des polices de caractères avec le navigateur : ce sont ici quelques pistes à explorer qui permettraient de répondre aux inquiétudes soulevées par le fingerprinting et seuls les développeurs de navigateur ont le pouvoir d'y apporter une solution, dans une approche dite de « privacy by design » **[22]**.

L'obligation pour les développeurs de logiciels de veiller au respect de la vie privée dès la conception de leur produit n'est toutefois pas encore inscrite dans le droit français. À défaut de pouvoir, à court terme et par la contrainte légale, tarir la source alimentant les techniques de traçage, la limitation du fingerprinting peut être envisagée du côté de ceux qui l'exploitent pour suivre les internautes.

C'est sous cet angle que les législateurs français comme européens encadrent les pratiques de traçage, que ce soit par usage des cookies ou du fingerprinting. La question est en effet : « À qui profite le traçage ?». Les informations dont résulte l'empreinte sont certes rendues disponibles par le navigateur selon une configuration définie par leur concepteur ; pour autant ces informations étaient-elles destinées à être exploitées à des fins de suivi ? À l'évidence non. Ce nouveau mode d'exploitation a été défini par les organismes tirant profit du traçage (par exemple, pour monétiser des espaces publicitaires, pour améliorer leurs techniques de ciblage et contourner l'éventuel blocage des cookies, etc.). Or, le suivi par fingerprinting, en tant qu' « action tendant à accéder, par voie de transmission électronique, à des informations déjà stockées dans [un] équipement terminal de communications électroniques » nécessite de recueillir l'accord préalable et informé de l'internaute et de lui donner la possibilité de refuser à tout moment ce suivi **[23]**.

Le fingerprinting n'en est qu'à ses débuts et seul le futur nous renseignera sur l'évolution et la prédominance de cette technique sur le traçage sur Internet dans les années à venir. ■

Références

[1] Le code du site AmIUnique.org, y compris le script de collecte d'empreinte, est disponible en open source : https://github.com/DIVERSIFY-project/amiunique

[2] https://blog.chromium.org/2013/09/saying-goodbye-to-our-old-friend-npapi.html

[3] http://unicode.org/faq/emoji_dingbats.html

[4] K. Mowery and H. Shacham. *Pixel perfect : Fingerprinting canvas in HTML5*. In M. Fredrikson, editor, Proceedings of W2SP 2012. IEEE Computer Society, May 2012, http://cseweb.ucsd.edu/~hovav/papers/ms12.html

[5] G. Acar, M. Juarez, N. Nikiforakis, C. Diaz, S. Gürses, F. Piessens, and B. Preneel. *Fpdetective : dusting the web for fingerprinters*. In Proc. of the Conf. On Computer & Communications Security (CCS), pages 1129–1140. ACM, 2013, https://www.cosic.esat.kuleuven.be/fpdetective/

[6] http://bluecava.com/

[7] https://www.addthis.com/

[8] http://revenviews.com/

[9] https://www.google.com/policies/privacy/#infocollect https://www.google.com/policies/privacy/archive/20150501-20150605/

[10] http://www.forbes.com/sites/eliseackerman/2013/02/27/big-internet-companies-struggle-over-proper-response-to-consumers-do-not-track-requests/

[11] http://yahoopolicy.tumblr.com/post/84363620568/yahoos-default-a-personalized-experience

[12] https://duckduckgo.com/privacy, http://donttrack.us/

[13] https://chrome.google.com/webstore/detail/user-agent-switcher-for-c/djflhoibgkdhkhhcedjiklpkjnoahfmg

[14] https://addons.mozilla.org/en-us/firefox/addon/masking-agent/

[15] https://adblockplus.org/fr/

[16] https://www.ghostery.com/fr/home

[17] https://disconnect.me/

[18] https://www.eff.org/privacybadger

[19] https://noscript.net/

[20] https://www.torproject.org/

[21] https://tails.boum.org/

[22] Ou principe de « protection des données dès la conception » introduit dans la proposition de règlement relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (COM(2012)0011)

[23] Article 32 II de la loi du 6 janvier 1978 modifiée (dite « Informatique et Libertés »)