Protéger ses données

Estelle Debouy

Plan du cours

Introduction

Quand Orwell écrivait *1984*, il aurait tout aussi bien pu écrire *2014* : des scandales comme celui révélé par Edward Snowden en 2013 (j'y reviendrai) nous donnent en effet à penser que rien n'a changé. La preuve que le symbole de *1984* est encore très présent aujourd'hui, c'est que lorsque le scandale Snowden est révélé, le livre passe à la 4e place des ventes sur Amazon. Sans parler de la surveillance de masse, nous sommes les premiers à livrer de notre plein gré – et parfois sans en avoir conscience – toutes sortes d'informations sur nous-mêmes. À l'heure des réseaux sociaux et du Cloud, peut-on encore parler de vie privée ?

Dans ce cours j'aborderai cette question en montrant d'abord combien est paradoxale la notion même de vie privée sur internet. Puis je m'intéresserai à ce qu'on appelle l'identité numérique : qu'entend-on par cette expression ? comment protéger cette identité ? Je présenterai rapidement ensuite un réseau qui, parce qu'il s'incarne dans la quête de l'anonymat et de la confidentialité, est l'objet de toutes sortes de représentations fantasmées : le darknet. Et je ne terminerai pas ce tour d'horizon sans donner quelques conseils simples pour protéger ses données personnelles.

On nous espionne

`Nom de code` : opération PRISM : opération menée par la NSA (agence de renseignement des USA) pour accéder aux données de millions de personnes.

`Qui?` Comme la loi interdit l'espionnage des Américains, la NSA espionne des étrangers en contact avec des Américains, puis... les Américains eux-mêmes.

`Quoi?` L'objectif est de constituer un graphe social des personnes et organisations ciblées grâce notamment aux méta-données. Qu'est-ce qu'on appelle méta-données ? Autour des informations contenues dans un fichier, il existe des informations sur ce contenu. Ce sont ces « données sur les données » qu'on appelle « méta-données ». Exemple : le nom du fichier, la date et l'heure de sa création et de sa modification. De nombreux formats de fichiers conservent également des méta-données à l'intérieur du fichier. Elles pourront donc être connues de quiconque aura accès au fichier. La palme revient probablement aux formats d'images comme TIFF ou JPEG : ces fichiers de photo créés par un appareil numérique ou un téléphone portable contiennent un standard de méta-données qui peut contenir la date, l'heure et parfois les coordonnées géographiques de la prise de vue.

C'est ainsi qu'en novembre 2012, John McAfee, fondateur de la compagnie d'antivirus

FIGURE 1 – Exemple de métadonnées dans un fichier image

qui porte son nom et sulfureux personnage impliqué dans une trouble affaire de meurtre au Belize, publiait sur le Web des photos de lui dans sa rocambolesque fuite vers le Guatamala puis la Floride ; or, il n'avait pas pris la précaution de supprimer les méta-données contenues dans la photo et la police a pu le localiser et procéder à son arrestation.

Revenons maintenant à l'opération PRISM. Elle a abouti à la constitution de bases de données énormes... avec beaucoup de déchets [1] mais de gros moyens sont mis en œuvre pour les exploiter [2], sans pour autant y parvenir puisque, selon Snowden, les services secrets américains avaient des informations qui leur auraient permis d'éviter l'attentat du 11 septembre s'ils avaient su exploiter cette « botte de foin » qu'ils avaient collectée. Le problème avec cette surveillance de masse, c'est qu'elle revient surtout à empiler les bottes de foin sans avoir vraiment les moyens de les analyser.

`Pourquoi?` L'information, c'est le pouvoir. Espionnage à triple facette : économique, politique, sécuritaire.

`Comment?`

— 90% des télécommunications circulent par câble sous-marins. Beaucoup de câbles

1. 3 millions de méta-données téléphoniques interceptées par jour.
2. Alors qu'on compte environ 1500 personnes affectées à ce travail à la DGSE, ils seraient plus de 40000 à la NSA.

ont été espionnés par la GB pour le compte des USA [3].

— recherche des failles de sécurité

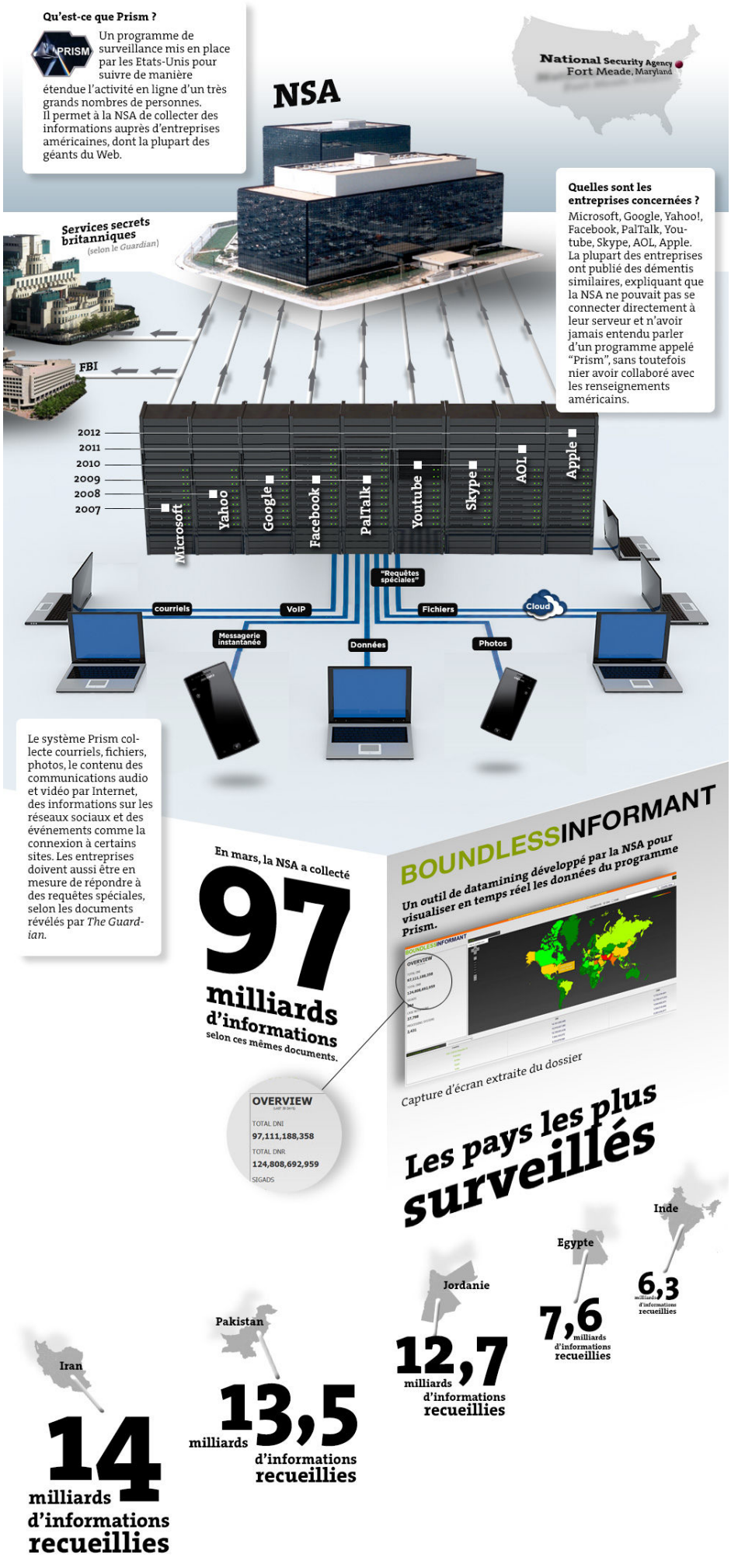
— chevaux de Troie dans les serveurs

— micros espions dans les ambassades

— piratage des VPN du Quai d'Orsay (pour ne citer qu'un exemple qui nous touche)

Mais il n'est pas besoin d'être une cible des agents de la NSA pour voir sa vie privée compromise. La preuve.

3. E. Snowden donne le nom des câbles espionnés.


Qu'est-ce que Prism ?
PRISM
Un programme de surveillance mis en place par les Etats-Unis pour suivre de manière étendue l'activité en ligne d'un très grands nombres de personnes. Il permet à la NSA de collecter des informations auprès d'entreprises américaines, dont la plupart des géants du Web.
NSA
National Security Agency
Fort Meade, Maryland
Services secrets britanniques
(selon le Guardian)
FBI
Quelles sont les entreprises concernées ?
Microsoft, Google, Yahoo!, Facebook, PalTalk, Youtube, Skype, AOL, Apple. La plupart des entreprises ont publié des démentis similaires, expliquant que la NSA ne pouvait pas se connecter directement à leur serveur et n'avoir jamais entendu parler d'un programme appelé "Prism", sans toutefois nier avoir collaboré avec les renseignements américains.
2012
2011
2010
2009
2008
2007
Microsoft
Yahoo
Google
Facebook
PalTalk
Youtube
Skype
AOL
Apple
"Requêtes spéciales"
courriels
VoIP
Fichiers
Cloud
Messagerie instantanée
Données
Photos
Le système Prism collecte courriels, fichiers, photos, le contenu des communications audio et vidéo par Internet, des informations sur les réseaux sociaux et des événements comme la connexion à certains sites. Les entreprises doivent aussi être en mesure de répondre à des requêtes spéciales, selon les documents révélés par The Guardian.
En mars, la NSA a collecté
97
milliards
d'informations
selon ces mêmes documents.
OVERVIEW
TOTAL DNI
97,111,188,358
TOTAL DNR
124,808,692,959
SIGADS
BOUNDLESSINFORMANT
Un outil de datamining développé par la NSA pour visualiser en temps réel les données du programme Prism.
Capture d'écran extraite du dossier
Les pays les plus surveillés
Iran
14
milliards
d'informations
recueillies
Pakistan
13,5
milliards
d'informations
recueillies
Jordanie
12,7
milliards
d'informations
recueillies
Egypte
7,6
milliards
d'informations
recueillies
Inde
6,3
milliards
d'informations
recueillies

1 Internet et le paradoxe de la vie privée

1.1 Internet voit tout, sait tout

Des robots qui nous traquent

Prenons l'exemple d'un blog. Sa publicité est assurée par une diffusion propre (l'auteur du blog contacte directement un certain nombre de personnes) et une diffusion automatique faite par des robots. Ce sont des robots dits d'indexation dont la fonction est de rechercher de nouveaux contenus et de les associer à des mots clés (d'où les résultats que vous obtenez quand vous lancez une requête dans un moteur de recherche). Ces robots ne sont pas les seuls à peupler le web : il existe aussi des des robots d'archivage qui explorent un certain nombre de sites pour en faire des sauvegardes. Si vous publiez un contenu erroné, les robots d'archivage risquent très souvent d'en garder trace.

Des sites qui nous traquent ou le cookie : l'espion qui permet de permet de nous profiler

Tout a commencé par une invention géniale : le cookie. Une simple ligne de code déposée sur votre navigateur par les sites Web que vous visitez, et des publicités ciblées s'affichent sur votre écran.

Appelé ainsi en référence aux biscuits que les restaurants offrent au moment de l'addition, le cookie apparaît dès 1994, l'année où le Web s'ouvre au public. Vingt ans plus tard, il reste le socle de la publicité en ligne, une industrie qui a réalisé en 2013 un chiffre d'affaires mondial de 102 milliards de dollars.

Les cookies sont gérés par des sociétés spécialisées qui les déposent, les récoltent, les classent, les analysent, les agrègent et les revendent. Ils servent à vous identifier, à vous pister de site en site, à retenir vos mots de passe, à gérer vos paniers d'achat, à déterminer si votre navigation est lente ou rapide, hésitante ou déterminée, systématique ou superficielle. L'objectif est de vous « profiler », c'est-à-dire de créer des fichiers personnalisés, stockés dans des bases de données. En d'autres termes, de mieux vous connaître afin de vous présenter le bon message publicitaire au bon moment et dans le bon format.

Afin d'affiner le ciblage, les publicitaires croisent les cookies avec d'autres données récoltées sur Internet : votre adresse IP (Internet Protocol : c'est l'équivalent de l'adresse postale. Pour qu'un serveur sache sur quel ordinateur afficher une information, il doit connaître son IP), votre langue usuelle, vos requêtes sur les moteurs de recherche, le modèle de votre ordinateur et de votre navigateur, le type de votre carte de crédit.

FIGURE 2 – Vie privée de Firefox

Le ciblage va jusqu'à modifier le prix d'un produit en fonction du profil. Quand un site de voyage voit que vous venez de consulter un comparateur de prix, il baisse ses prix pour s'aligner sur ceux de ses concurrents, quitte à se rattraper sur les « frais de dossier ». Si vous vous connectez avec un ordinateur à 3 900 euros, le site affichera des chambres d'hôtel plus chères que si vous utilisez un portable à 300 euros. Le libre choix du consommateur, apparemment décuplé par la puissance de l'informatique, semble en fait amoindri.

Vous pouvez certes effacer les cookies, mais de nouveaux arriveront dès que vous reprendrez la navigation. Et si vous les bloquez, la plupart des sites ne fonctionneront plus. Certains cookies ont la vie dure : ceux que dépose Amazon aujourd'hui sont conçus pour durer jusqu'en 2037.

Exemple : dès la page d'accueil du site de e-commerce Priceminister, votre navigateur reçoit d'un coup 44 cookies provenant de 14 agences spécialisées – telles que RichRelevance, Doubleclick, Exelator. Rendez-vous à la rubrique « Téléphonie mobile », vous récoltez 22 nouveaux cookies. Et en cliquant sur la photo d'un smartphone Samsung, vous déclenchez une nouvelle rafale de 42 cookies provenant de 28 sources : en trois clics, vous voilà fiché 108 fois par une quarantaine de bases de données.

Les cookies tiers ne sont habituellement pas nécessaires pour profiter des ressources disponibles sur Internet. Pour limiter ses traces, il est possible les refuser par défaut. Dans le navigateur Firefox : Menu > Options > Vie privée.

Vous voulez savoir comment et par qui sont utilisées vos données personnelles ? Ins-

tallez le module Collusion de Firefox (https://addons.cdn.mozilla.net/storage/public-staging/363974/collusion-0.27-fx.xpi). Dès lors que vous l'exécutez, ce dernier représente sous forme de graphique les interconnexions entre toutes les parties qui vous suivent ainsi que le cheminement des informations. Il est donc alors plus facile de filtrer ou bloquer les cookies des sites qui vous pistent.

Vérifier l'information Internet est devenue pour beaucoup la principale source d'accès à des contenus culturels et informatifs. Comment être sûr de l'authenticité d'une information ? Quelques réflexes à avoir : vérifier la date de publication, les sources de l'information et ne pas hésiter à faire sa propre recherche. Voir aussi :

— http://www.hoaxbuster.com/
— https://www.lemonde.fr/les-decodeurs/

Mais notre pire ennemi en matière de vie privée est... nous-mêmes !

1.2 Les réseaux sociaux et le cloud computing

Les réseaux sociaux

Facebook, initialement conçu pour permettre de communiquer entre personnes issues de la même école, du même sérail, est un `réseau social` qui, depuis, pousse ses utilisateurs à y mener une vie publique, tout en y révélant un maximum de données personnelles, afin de pouvoir profiler ses utilisateurs, et commercialiser ces profils clients auprès d'annonceurs pour y afficher de la publicité comportementale et personnalisée, en vertu de l'adage qui veut que « Si c'est gratuit, c'est que vous êtes le produit » [4].

Et cela fonctionne ! À tel point que Bill Thompson, un célèbre éditorialiste à la BBC, spécialisé dans les technologies, expliquait, de manière certes provocatrice, lors de la conférence Lift en 2009 à Berlin :

> Les utilisateurs de Twitter, Tumblr et autres outils de réseaux sociaux, partagent plus de données, avec plus de gens que le FBI de Hoover ou la Stasi n'auraient jamais pu en rêver. Et nous le faisons de notre propre chef, espérant pouvoir en bénéficier de toutes sortes de manières.

Il en déduit qu'il faut en finir avec la notion de vie privée. Et en effet il apparaît comme une évidence qu'il n'y a pas de vie privée sur Facebook : sur un réseau social, on

4. Gmail scanne vos courriels privés, Google archive les mots-clefs que vous recherchez, Facebook surveille les articles, pages et billets que vous consultez – quand bien même vous ne les auriez pas partagés. Pour autant, Facebook et Google n'ont que faire de votre vie privée. Ce qui intéresse ces marchands de données, c'est de vendre et donc d'afficher des publicités personnalisées, en fonction de vos profils et ce, quels qu'ils soient : ils ne s'intéressent pas aux individus), ils ciblent des consommateurs.

mène une vie sociale, voire une vie publique.

Cette prise de conscience a été favorisée par un certain nombre de polémiques qui ont défrayé la chronique (messages sensés être privés mais visibles dans la partie publique, licenciements à cause de Facebook), montrant ainsi à quel point les internautes attendent de Facebook qu'il protège leur vie privée... alors même que, et souvent, ils ne la protègent pas eux-mêmes correctement.

Activité : maîtriser son compte Facebook https://www.slideshare.net/JulieGarnier2/matriser-son-compte-facebook

Le Cloud

C'est une technologie qui permet de stocker des données sur des serveurs distants et d'y avoir accès n'importe quand et depuis n'importe quel appareil connecté à Internet. L'inconvénient majeur du Cloud est que la sécurité de nos données est entre les mains de tiers qui peuvent avoir accès et même utiliser nos données sans même que nous nous en rendions compte. Exemple : comment faire confiance à des prestataires américains (ou plus précisément à des prestataires dont les serveurs sont sur le sol américain) quand on sait que le Patriot Act est une loi aux États-Unis qui donne au gouvernement le droit de regard sur les données collectées et stockées sur son sol ? D'où quelques bonnes habitudes à prendre :

- avoir un mot de passe suffisamment fort et unique pour chaque service (on y reviendra)
- bien lire les règles de confidentialité et les conditions d'utilisation
- ne pas stocker des documents très sensibles dans le Cloud
- chiffrer ses données (on verra ce que cela signifie)
- opter pour une solution de partage de fichiers fiable : https://upload.disroot.org/)

Alors, LA solution – radicale, certes – serait-elle de renoncer à internet ? Même pas, car, comme nous allons le voir, même hors connexion, nous laissons des traces à tous les étages.

2 Traces à tous les étages

2.1 Ma vie disséquée à travers mes données personnelles

Du lever au coucher, on sait depuis quelques années que nos vies se copient en temps presque réel dans des bases de données, parfois sans notre véritable consentement. La vie numérisée dessine-t-elle un portrait fidèle de ce que je suis ? Un journaliste du *Monde* s'est amusé à retracer une journée type : au réveil, son premier réflexe consiste à allumer son IPhone qui, instantanément, se géolocalise afin d'« améliorer ses performances et proposer des informations utiles en fonction des lieux où vous êtes » (sic !). Cela dit, Apple assure ne pas stocker ces données dans un datacenter et pourtant la NSA a ajouté l'entreprise à son programme Prism, qui permet d'accéder de manière privilégiée aux données de plusieurs géants du Web.

La pluie conduit ensuite notre journaliste à prendre le métro dont le portique s'ouvre après le passage du badge, le Pass Navigo qui est associé à toute son identité, sauf si on fait le choix, moyennant finance, d'un badge anonyme appelé Pass découverte. Puis : arrivée sur son lieu de travail, là encore avec badge à l'accueil. À peine arrivé au bureau, il projette d'aller au cinéma le soir même et se fait la réflexion que sa carte UGC doit enregistrer l'ensemble des informations et des films qu'il est allé voir. Quand il essaie de se renseigner, il se heurte ou bien à de la réticence ou bien à de l'incompréhension, alors même que la loi informatique et libertés de 1978 prévoit explicitement un droit quasiment inconditionnel d'accès aux données personnelles.

Il poursuit sa journée et énumère toutes les traces numériques qu'il laisse : carte de cantine (qui stocke l'historique de ses consommations pendant 13 mois), carte vitale quand il passe à la pharmacie en sortant de son travail, les recherches qu'il a effectuées dans Google quand il rentre chez lui : il constate qu'en parcourant la liste des requêtes qui sont enregistrées, c'est tout simplement ses intérêts professionnels, lubies, passe-temps qui sont soigneusement classés par ordre chronologique. Il en vient à la conclusion que, mises bout à bout, ces bases de données réunissent ses goûts, ses habitudes, ses obsessions, ses loisirs, ses centres d'intérêt et en vient à s'interroger : dispersées sur des ordinateurs aux quatre coins du monde, ces données, souvent analysées, résistent encore aux croisements et recoupements divers. Mais pour combien de temps ?

Mais savez-vous bien ce qu'on entend par « données personnelles » ? Quels sont vos droits sur vos données ?

2.2 Vos droits sur vos données

Législation

Définition de ce qu'on appelle données personnelles (article 2 de la Loi informatique et libertés de 1978) :

> Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne.

Notons d'abord que la nature des informations n'a aucune importance. La question clé est de savoir s'il y a un lien avec une personne physique. Dans ce cas, alors il s'agit de données personnelles.

Mais le loi va plus loin et définit dans l'article 8 ce qu'on appelle données sensibles :

> Il est interdit de collecter ou de traiter des données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales, ethniques, les opinions politiques, philosophiques, religieuses ou l'appartenance syndicale, ou qui sont relatives à la santé ou à la vie sexuelle des personnes.

Une des difficultés qu'il faut souligner est liée aux inférences. Exemple : si un responsable de traitements, par le biais d'une application smartphone, collecte ma géolocalisation de façon systématique et précise, alors, il est capable très facilement d'inférer, c'est-à-dire de déduire un certain nombre de choses, par exemple, si je fréquente régulièrement un lieu de culte, et on tombe alors dans la catégorie des données sensibles.

Certaines obligations s'imposent, par ailleurs, au responsable de traitements d'après l'article 6 : il doit effectuer une collecte et un traitement de façon loyale et licite, mais aussi, pour des finalités bien précises et légitimes. Ainsi, il est totalement hors de question de collecter ma géolocalisation en permanence si l'objectif annoncé pour cette collecte est de personnaliser un service utilisé de façon ponctuelle.

Vos droits sur vos données

La loi informatique et libertés nous donne 4 droits fondamentaux : l'accès, la rectification, l'opposition et l'oubli.

1. Le droit d'accès nous permet d'obtenir une copie de toutes les informations qu'un service possède sur nous. Pour obtenir toutes ces informations, la CNIL, Commission nationale informatique et libertés, dispose de formulaires afin de vous faciliter les démarches.
2. Nous disposons aussi d'un droit de rectification. Si une personne publie des données et informations qui sont inexactes et qui vous concernent, vous pouvez demander que ces informations soient rectifiées. La demande se fait auprès du service du responsable du traitement, et de la même façon, en cas de refus ou sans réponse, il suffit de porter plainte à la CNIL afin d'obtenir rectification. Le droit de rectification a cependant des limites. Il ne couvre pas les contenus artistiques, littéraires ou journalistiques. Ce n'est pas parce qu'une peinture vous représentant vous déplaît que vous serez capable de la faire rectifier auprès de l'artiste.
3. Le troisième droit est le droit d'opposition qui vous permet, lors de la collecte ou de la diffusion de données, de refuser que vos données soient prises en compte.
4. Le droit à l'oubli est un droit un peu plus spécifique. Il concerne uniquement les moteurs de recherche. On pourrait l'appeler exactement le droit au déréférencement. En substance, il consiste à faire éliminer des résultats de requêtes de moteurs de recherche. Par exemple, vous souhaitez qu'un lien que vous considérez diffamant soit éliminé d'un moteur de recherche quand une personne va taper votre nom et prénom.

Quelle confiance dans les services qui gèrent nos données ?

La question ne date pas d'aujourd'hui malheureusement. Peut-être certains d'entre vous se souviennent-ils d'une affaire qui défraya la chronique en... 1974 ! Le ministère de l'Intérieur autorisa le croisement des fichiers informatiques administratifs en utilisant le numéro de sécurité sociale, créant une base de données centralisée de toute la population : c'est le projet SAFARI, qui fut révélé par *Le Monde*.

Comme le projet suscita une vive opposition, il fut retiré et fut créée une commission Informatique et Libertés afin de réfléchir « au devenir des libertés individuelles et publiques dans la quête permanente d'information ».

Voir https://www.cnil.fr/fr/maitriser-mes-donnees.

Nous enregistrons beaucoup de données (identifiants, authentifiants, images, conversations, idées, données de santé, données bancaires) dans des services, des réseaux sociaux, des clouds, des messageries électroniques. Que doit-on attendre de tels services ? Le res-

• • • LE MONDE — 21 mars 1974 — Page 9

JUSTICE

Tandis que le ministère de l'intérieur développe la centralisation de ses renseignements

Une division de l'informatique est créée à la chancellerie

En ordre dispersé, les départements ministériels tentent de développer à leur profit, à leur seul usage, l'informatique et son outil, l'ordinateur. Ce n'est pas tout à fait un hasard si, à l'époque où le Journal officiel va publier un arrêté créant une « division de l'informatique » au ministère de la justice, celui de l'intérieur met la dernière main à la mise en route d'un ordinateur puissant destiné à rassembler la masse énorme des renseignements grapillés sur tout le territoire; pas un hasard non plus si le projet SAFARI (Système automatisé pour les fichiers administratifs et le répertoire des individus) destiné à définir chaque Français par un « identifiant », qui ne définisse que lui, maintenant terminé, est l'objet de convoitises ardentes; le ministère de l'intérieur y souhaite jouer le premier rôle. En effet, une telle banque de données, soubassement opérationnel de toute autre collecte de renseignements, donnera à qui la possédera, une puissance sans égale.

Ainsi se trouve d'évidence posé un problème fondamental, même s'il est rebattu : celui des rapports des libertés publiques et de l'informatique. Son importance exigerait qu'il en fût, au Parlement, publiquement débattu. Tel ne paraît pas être, pourtant, la solution envisagée par le premier ministre dans les directives qu'il vient d'adresser au ministère de la justice, intéressé au premier chef si l'on s'en rapporte à la Constitution qui dans son article 66 fait de l'autorité judiciaire le gardien des libertés individuelles.

« Safari » ou la chasse aux Français

Rue Jules-Breton, à Paris-13e, dans des locaux du ministre de l'intérieur, un ordinateur Iris-80 avec bi-processeur est en cours de mise en marche. A travers la France, les d'tférents services de police détiennent, selon la confidence faite par un très haut magistrat, 100 millions de fiches, réparties dans 400 f.chiers. Ainsi se trouve posées — et, à terme, théoriquement résolues — les données d'un problème comprenant, d'une part, l'énormité des renseignements collectés ; de l'autre, la méthode à définir pour faire de cet ensemble une source unique, à tous égards, de renseignements.

L'histoire du très puissant appareil qu'est l'Ir.s-80 est exemplaire du secret qui entoure l'épanouissement de l'informatique dans les administrations, quelles que puissent être les informations qui filtrent ici et là.

l'origine, budgétairement, n'était pas du tout prévu pour la tâche qu'il a finalement assurée, mais pour « traiter » les données administratives du Fichier national des constructeurs (F.N.C.). Il s'agit donc apparemment d'un détournement man.teste de crédits d'études, ce qui n'était sans doute pas le vœu du Parlement qui les vota.

De vastes ambitions

Il n'y a pas que cela. Le ministère de l'inténeur a d'encore plus vastes ambitions. Détenteurs, déjà, du fichier national du remembrement, les services de M. Jacques Chirac font de grands efforts pour, affirme-t-on, s'en adjoindre d'autres : le cadastre, le fichier de la direction nationale des impôts et, plus grave peut-être, celui du ministère du

Ce n'est pas, pourtant, que les avertissements aient manqué. Le Conseil d'Etat en 1970, puis le ministère de la justice en 1972 (qui avait rappelé le rôle dévolu à l'autorité judiciaire de « *gardien des libertés individuelles* » et donc réclamé voix au chapitre) ont insisté sur la nécessité d'une intervention législative qui préciserait les quelques éléments essentiels de l'emploi de l'informatique appliquée aux particuliers : réglementation de l'acces des tiers aux fichiers, de l'intercommunication de ceux-ci, droit de rectification des personnes fichées si les renseignements retenus sont inexacts, etc.

De plus, tous les exemples étrangers incitent à ce débat sur une utilisation de l'informatique à laquelle, par définition, il ne s'agit pas de renoncer, mais à qui doivent

outil, il n'apparaît pas — sauf erreurs négligeables, relativement — que l'accès des tiers ou le droit à contrôle des personnes visées — par demande d'un extrait — ait jamais provoqué des bavures préjudiciables à la légalité.

De même, le fichier national des conducteurs, dans sa partie judiciaire, est prévu par une loi, et il faut regretter que les textes d'application ait permis des illégalités injustifiables — mais connues (*le Monde* du 8 mars).

« A la hussarde »

Fort, pourtant, de ces avantages, le ministère de la justice paraît curieusement se laisser dépasser par des querelles internes peu compréhensibles. L'arrêté signé le 18 mars par M. Jean Taittinger le montre.

dite commission. D'autant qu'il est d'ores et déjà connu que M. Touffait a été rayé de la liste des « sachants ». Il semble d'ailleurs que les réactions vives qui sont enregistrées portent moins sur le renouvellement des structures, jugées inévitables, que sur la méthode « *à la hussarde* » employée par tel membre de l'entourage de M. Taittinger pour mener à bien ses projets de rénovation de la gestion dans le domaine judiciaire.

Est-ce à dire de plus que les choix que l'on entend promouvoir soient nécessairement les plus opportuns ? Tout indique, pour l'instant, que, si le ministère de l'intérieur a définitivement choisi le « matériel lourd » pour s'équiper, la chancellerie, au contraire, s'oriente vers un réseau de mini-ordinateurs placés auprès de chaque tribunal de grande instance important.

d'un *aggiornamento* technique, a illustré son discours par un large reportage sur les équipements du tribunal de Bobigny — plus réduits, donc plus rapides à réaliser, ainsi plus vite source d'orgueil pour leurs créateurs.

C'est donc un doute global qui pèse sur les intentions du gouvernement, en général, et du ministère de la justice, en particulier : ce dernier département, qui rappelle à tous sa mission de protection des libertés individuelles, a apparemment accepté sans broncher la suppression d'un éventuel débat public, ce qui jette sur les déclarations « libérales » de M. Taittinger en d'autres domaines une suspicion qui n'est pas de bon aloi.

Mais, dans cette entreprise, le ministère de la justice, même s'il fait

FIGURE 3 – Article du *Monde* du 21 mars 1974

pect de la vie privée, c'est-à-dire que ne soient pas divulguées nos données personnelles. Le risque zéro n'existe pas. Un exemple est, à ce titre, tout à fait significatif : en 2014, des pirates ont réussi à récupérer les données d'un demi-milliard d'utilisateurs de Yahoo. On s'est rendu compte de ces pertes de données seulement en juillet 2016 quand on a commencé à voir, sur certains réseaux, la mise en vente de l'ordre de 200 millions de mots de passe de comptes Yahoo. Assez rapidement, Yahoo a reconnu ces pertes de données et il a fallu attendre décembre 2016 pour que Yahoo force tous les utilisateurs à changer leur mot de passe et renouveler la sécurité de leur compte. Voir : « Des millions de comptes Yahoo ! piratés en 2014 », FR3, 23 sep 2016 : http://sites.ina.fr/cnil-40-ans/focus/chapitre/5/medias/5823342_001_014 et https://www.lemonde.fr/pixels/article/2018/06/13/piratage-de-yahoo-en-2014-le-groupe-condamne-a-280-000-euros-d-am 5314242_4408996.html.

Comment vérifier si un site est de confiance, comment vérifier sa web-reputation ? Dans Firefox, on peut utiliser l'extension WOT (Web Of Trust) : https://addons.mozilla.org/en-US/firefox/addon/wot-safe-browsing-tool/. C'est une extension de Firefox conçue pour sécuriser la navigation des internautes sur la toile. Elle sert à vérifier si un site est sûr ou non. L'application intègre aussi une liste noire de sites peu recommandables. Une fois WOT installée, une icône représentée par un feu de couleur apparaît dans la barre d'outils du navigateur. Les sites sont alors classifiés suivant des codes couleurs, à savoir le vert pour bon, le jaune pour douteux et le rouge pour

malveillant.

2.3 L'identité numérique

En France, toute personne physique a une identité juridique, qui est constituée d'un nom, d'une adresse, d'un état civil. Il existe d'autres formes d'identités, une identité administrative qui va être caractérisée par un code INSEE ou un numéro fiscal. Ces identités permettent d'identifier de façon unique un individu afin de lui octroyer des droits (le code INSEE permet l'accès aux soins par exemple), ou des devoirs (ex. : vérifier que la personne a bien acquitté ses impôts avec le numéro fiscal). Quand on va sur Internet ou quand on utilise des technologies, on a exactement les mêmes besoins d'identification, la technologie a besoin d'identifier les utilisateurs de façon unique.

Ce qui va constituer l'identité numérique, ce sont tous les identifiants qui vont nous permettre de faire le lien à la technologie. Donc concrètement, qu'est-ce qu'une identité numérique ? Les identifiants que vous avez choisis, vos adresses mail, vos identifiants Facebook, votre adresse IP, etc. Une autre famille d'adresses et d'identifiants utilisés pour acheminer l'information sur Internet est les adresses MAC (rien à voir avec Apple) : ce sont des adresses qui sont associées à l'équipement qui permet de se connecter au réseau.

Compte tenu du tableau que je viens de dresser de notre vie privée sur internet, on peut comprendre que certains choisissent le darknet. Mais de quoi parle-t-on au juste ?

3 Le darknet

Voici quelques explications qui visent à démythifier ce côté obscur du net. Le darknet est avant tout un fait social : il s'incarne dans la quête de l'anonymat et de la confidentialité. Avant d'aller plus loin, juste un mot pour clarifier ces concepts :

— `anonymat` : consiste à dissimuler son identité (les actions peuvent être connues, mais pas leurs auteurs) ;

— `confidentialité` : consiste à interdire l'accès à l'information aux tiers (repose essentiellement sur le chiffrage) :

— `vie privée` : préserver sa vie privée signifie qu'on ne souhaite pas que certaines de nos activités soient observées à notre insu. Pour E. Snowden, « répondre je n'ai rien à cacher en matière de vie privée revient à affirmer qu'on se fiche de la liberté d'expression parce qu'on n'a rien à dire ».

Définition du darknet : un sous-réseau d'internet utilisant des protocoles spécifiques et intégrant nativement des fonctions d'anonymisation.

Ne pas confondre darknet et deep web : ce qu'on appelle le deep web, c'est toute la partie du web que ne peuvent indexer les moteurs de recherche et donc à laquelle on ne peut accéder directement (pages générées par l'interrogation d'une base de données via un formulaire ou encore sites auxquels on accède après identification). Si vous consultez certaines statistiques sur le site de l'INSEE (donc à partir d'un formulaire qui cherche dans une base de données), vous êtes dans le deep web mais certainement pas dans le darknet ! Le darknet repose avant tout sur la recherche de la préservation de l'anonymat. Et ce n'est possible que grâce au chiffrage.

3.1 La question du chiffrage

Les outils modernes du chiffrage reposent sur des principes très anciens. La cryptographie remonte à l'Antiquité : Suétone écrit dans la *Vie de César*, 56 [5] :

> On possède enfin de César des lettres à Cicéron, et sa correspondance avec ses amis sur ses affaires domestiques. Il y employait, pour les choses tout à fait secrètes, une espèce de chiffre qui en rendait le sens inintelligible (les lettres étant disposées de manière à ne pouvoir jamais former un mot), et

5. De façon insolite, le code de César a été réemployé notamment au début d'internet et des forums, à travers le ROT-13. Le ROT-13 désigne simplement le code de César, où on a choisi une ROTation de 13 lettres (A–>N...). L'idée n'est pas de diffuser des messages cryptés, mais de faire en sorte que le message ne soit pas lu involontairement, par exemple s'il dévoile l'intrigue d'un film ou donne la réponse à une devinette.

> qui consistait, je le dis pour ceux qui voudront les déchiffrer, à changer le rang des lettres dans l'alphabet, en écrivant la quatrième pour la première, c'est-à-dire le d pour l'a, et ainsi de suite.

À vous de jouer : que veut dire : DOHD MDFWD HVW

Plus récemment, on eut recours à des systèmes reposant sur le même principe mais utilisant plusieurs alphabets de substitution (appelés systèmes polyalphabétiques) : le plus célèbre est peut-être l'Enigma utilisée par l'armée allemande pendant la seconde guerre mondiale. C'est Alan Turing qui découvrit le système de code d'Enigma en 1944 (et qui posa à l'occasion les bases du premier ordinateur !).

FIGURE 4 – Enigma 1940 (Wikimedia Commons, licence CC BY-SA 3.0)

C'est en 1976 qu'apparaît le concept qui va révolutionner la cryptographie : le chiffrage à clé publique. Le principe est le suivant : Alice et Bertrand souhaitent communiquer.

/home/estelle/Documents/informatique/fifti/fichiers/miscellanees/images/chiffrement.png

FIGURE 5 – Chiffrement asymétrique

Alice crée deux clés, une clé qui servira au chiffrage (clé publique) et la clé correspondante qui servira au déchiffrage (clé privée). C'est la première clé qu'Alice distribuera à tous les Bernard potentiels. Bernard, pour répondre, chiffre le message avec la clé publique d'Alice qui le déchiffre avec sa clé privée.

Pendant longtemps, la cryptographie a exclusivement été une affaire d'État, cantonnée aux opérations militaires et diplomatiques (et jusqu'en 1990 les systèmes de chiffrage étaient considérés en France comme des armes de guerre). Mais avec l'explosion des échanges sur internet, la problématique a changé : il est devenu indispensable pour tout un chacun de garantir la sécurité de ses transactions. C'est cette pression économique qui a obligé les États à assouplir leur contrôle et en 2004 la loi pour la confiance dans l'économie numérique déclare dans son article 30 : « l'utilisation des moyens de cryptologie est libre ». Et pourtant, pour les États, la cryptologie libre généralisée est d'abord considérée comme une atteinte à leurs capacités d'investigation. La NSA (National Security Agency), créée en 1952, est au cœur de ces problématiques puisqu'elle a centralisé l'effort américain dans le domaine de la cryptographie. Les révélations d'E. Snowden ont notamment dévoilé l'existence d'un programme (*Bullrun*), doté de 250 millions de dollars en 2013, visant à affaiblir les systèmes de chiffrage. La NSA n'hésite pas à intervenir auprès des opérateurs internet et des entreprises logicielles pour qu'ils mettent en place des portes dérobées. On comprend bien pourquoi les solutions propriétaires ne peuvent être dignes de confiance, par opposition aux logiciels open source qui sont vus et audités par une vaste communauté, raison pour laquelle il est difficile d'y introduire des fonctions cachées.

Quels sont les outils faciles à utiliser dont nous disposons ? AxCrypt : https://www.axcrypt.net/fr/.

3.2 Les outils du darknet

Les outils du darknet existent pour les grands systèmes d'exploitation : pas d'obligation de passer par Linux, même si son usage est un atout certain si on veut préserver son anonymat, pour au moins trois raisons :

1. les outils de sécurité et d'anonymisation sont d'abord développés pour Linux ;

2. le pire ennemi de l'anonymat est tout simplement la compromission de son propre système : si un programme enregistre tout ce que vous tapez, la notion d'anonymat n'a plus de sens. Or la probabilité de voir son système infecté est beaucoup plus faible sous Linux car les virus et autres malwares sont conçus pour s'attaquer au plus grand nombre ;
3. Linux est un système open source (voir ce qu'on a dit plus haut à ce sujet).

Dans le cadre de ce cours, je me contenterai de présenter un seul outil : Tor.

Tor Qu'est-ce que Tor ? Il s'agit d'abord d'un navigateur qui a recours à des serveurs intermédiaires multiples re-routant l'information sans avoir accès à celle-ci ni connaître son origine ou sa destination finale. La technique consiste donc à cacher les utilisateurs parmi les utilisateurs.

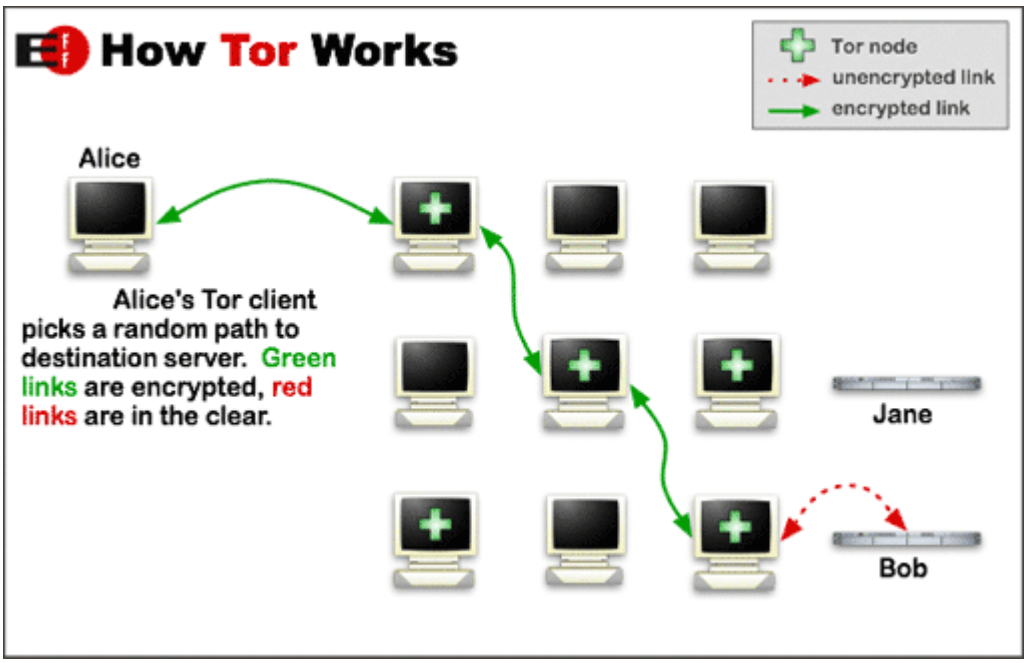


FIGURE 6 – Comment fonctionne Tor

C'est au départ un projet de la U.S. Navy. Au lieu d'une connexion directe entre l'utilisateur et le serveur, on utilise un ensemble de relais via des connexions chiffrées. On a ainsi un ensemble de couches de chiffrage, d'où la référence à l'oignon ! Pour accéder au réseau Tor, il suffit d'installer le *Tor Browser* qui s'installe comme n'importe quelle application. À partir de là, la navigation est anonyme. On peut aller sur n'importe quel site web sans qu'il puisse nous identifier. Tor n'est pas limité à la navigation anonyme :

il héberge également un web caché, c'est-à-dire un ensemble de sites auxquels on accède uniquement à travers Tor. Ils ont pour extension .onion. C'est le cas, par exemple, de l'adresse de dépôt des documents de WikiLeaks (j'y reviendrai). Le succès de Tor est impressionnant et il a dépassé les 5 millions d'utilisateurs au lendemain des révélations de Snowden. Tout un chacun peut utiliser Tor. On estime à environ 2 millions le nombre d'utilisateurs de Tor chaque jour.

3.3 Darknet et libertés

L'affaire Snowden et la surveillance de masse

Le darknet est au cœur de l'affaire Snowden dont j'ai dit un mot au début de ce cours. Fin décembre 2012, le journaliste Glen Greenwald, connu pour ses critiques des systèmes de surveillance, a reçu un email disant que son correspondant anonyme disposait d'informations importantes qu'il ne pourrait communiquer qu'après avoir obtenu sa clé PGP. Vous voyez que, dès le début, les révélations qui ont ébranlé le monde n'ont été possibles que grâce aux technologies qye nous venons d'évoquer. Puis, c'est dans un hôtel d'Hong-Kong que les deux hommes se sont rencontrés : E. Snowden a remis au journaliste des documents montrant que les services américains se livraient à de l'espionnage de masse de leurs concitoyens, ce que la constitution interdit formellement. Ce fut l'objet d'un premier article, suivi bientôt d'un deuxième qui révélait, quant à lui, le programme PRISM. Voici quelle était la conclusion à laquelle arrivait G. Greenwald (*Nulle part où se cacher*, Lattès, 2014) :

> Prises dans leur intégralité, les archives Snowden conduisent à une conclusion fort simple : le gouvernement américain a bâti un système qui s'est fixé comme objectif l'élimination complète, à l'échelle planétaire, de toute vie privée électronique.

Et ce n'est pas l'apanage des États-Unis : l'ancien directeur technique de la DGSE expliquait que son objectif était « d'écouter tout l'internet mondial » (propos rapportés par *Libération*, 29/09/2015).

Cette question du contrôle permanent, de cette surveillance constante, est l'objet de réflexion pas seulement chez les informaticiens mais aussi chez les philosophes. C'est le cas de Gilles Deleuze qui écrivait déjà en 1990 [6] :

> Nous entrons dans des sociétés de contrôle [7] qui fonctionnent, non plus par

6. Dans un article intitulé « Post-scriptum sur les sociétés de contrôle » (*Pourparlers*, 1990).
7. Après la fin des institutions disciplinaires théorisées par Michel Foucault.

> enfermement, mais par contrôle continu et communication instantanée. [...] Face aux formes prochaines de contrôle incessant en milieu ouvert, il se peut que les plus durs enfermements nous apparaissent appartenir à un passé délicieux et bienveillant. [...] L'important ce sera peut-être de créer des vacuoles de non-communication, des interrupteurs, pour échapper au contrôle.

Le darknet est précisément l'un de ces interrupteurs.

Wikileaks ou le droit d'alerte

Durant l'été 1971, le New York Times révélait les *Pentagon Papers*, montrant l'ampleur des mensonges et manipulations du gouvernement américain sur le Vietnam, faisant ainsi basculer l'opinion publique. Daniel Ellsberg, à l'origine des révélations, est l'un des premiers lanceurs d'alerte moderne. L'année suivante, William Mark Felt était à l'origine du scandale du Watergate, qui conduisait à la démission de Nixon. Wikileaks incarne aujourd'hui le lien entre alertes et liberté d'information. La plate-forme inaugurée par Julien Assange publie ses premiers documents à la fin de l'année 2006 mais c'est notamment la révélation des procédures en vigueur à Guantanamo l'année suivante qui ont commencé à attirer l'attention des médias. Sur son site, Wikileaks déclare [8] :

> Les principes généraux sur lesquels notre travail s'appuie sont la protection de la liberté d'expression et de sa diffusion par les médias, l'amélioration de notre histoire commune et le droit de chaque personne de créer l'histoire. Nous tirons ces principes de la Déclaration universelle des droits de l'homme. En particulier l'article 19 inspire le travail de nos journalistes et autres volontaires.

Comment fonctionne Wikileaks ? Ceux qui souhaitent communiquer des documents à Wikileaks doivent se connecter via Tor à une adresse en .onion (web caché) qu'on trouve sur leur site (voir capture d'écran).

Wikileaks et ses homologues sont une démonstration de la puissance du darknet...

8. Voici le texte original : « The broader principles on which our work is based are the defence of freedom of speech and media publishing, the improvement of our common historical record and the support of the rights of all people to create new history. We derive these principles from the Universal Declaration of Human Rights. In particular, Article 19 inspires the work of our journalists and other volunteers. » (source : https://wikileaks.org/About.html)

FIGURE 7 – Soumettre des documents à Wikileaks

4 Comment se protéger : les quatre conseils faciles à suivre

4.1 Privilégier les logiciels libres

Qu'entend-on par logiciel libre ? Il s'agit d'un logiciel dont le code source (c'est-à-dire la recette de fabrication) est ouvert, lisible par tout un chacun. Exemple : LibreOffice (comme son nom l'indique est un logiciel libre) par opposition à Microsoft Word qui est un logiciel propriétaire.

L'ouverture du code source d'un programme est d'une importance capitale, car seule la lecture du code source, avant la compilation, permet de savoir à quelles instructions obéit le programme. Beaucoup de programmes fermés sont faits pour envoyer aux sociétés qui ont concédé la licence d'utilisation des informations collectées sur les propriétaires des ordinateurs sur lesquels ils sont exécutés. Les contrats de licence présentent d'ailleurs ces politiques de collecte de données en les atténuant : le plus souvent, il est question de lutter contre le piratage des logiciels... C'est ainsi, par exemple, qu'il est désormais impossible d'initialiser un ordinateur *Apple* sans entrer des informations détaillées sur son identité et son adresse. Un autre exemple nous est fourni par la société *Skype* : depuis son rachat par *Microsoft*, toutes les conversations qui passent par son logiciel de communication sont enregistrées «pour des raisons de sécurité»...

Dernière chose à préciser quand on parle de logiciels : avoir des logiciels (et un OS) libres est fondamental, mais encore faut-il veiller à ce qu'ils soient à jour : navigateur, antivirus, bureautique, pare-feu personnel, etc. La plupart des attaques tentent d'utiliser les failles d'un ordinateur (failles du système d'exploitation ou des logiciels). En général, les agresseurs recherchent les ordinateurs dont les logiciels n'ont pas été mis à jour afin d'utiliser la faille non corrigée et ainsi parviennent à s'y introduire. C'est pourquoi il est fondamental de mettre à jour tous ses logiciels afin de corriger ces failles. Mettre à jour vos logiciels, c'est une chose, mais il faut aussi penser à mettre à jour votre OS. Sous Windows, c'est Windows Update qui vous permettra de mettre à jour votre ordinateur. Sous Mac, Préférences Système > App Store.

4.2 Choisir un bon navigateur

Firefox est un bon choix. Sachez qu'il existe aussi une version de *Firefox* entièrement libre, dans le sens où les auteurs, à la différence de ceux de *Firefox*, ne distribuent que des *plugins* ou des modules complémentaires (*addons*) entièrement libres : il s'agit de *Icecat* (`http://www.gnu.org/software/gnuzilla/`). En outre, *Icecat* est livré avec un module complémentaire de protection de la vie privée : *Gnuzilla privacy extension* qui peut aussi être installé dans *Firefox* à partir de l'adresse suivante : `http://ftp.gnu.org/gnu/gnuzilla/` (choisir ensuite votre version de Firefox, et cliquer sur `privacy_ext.xpi`).

Pour améliorer la sécurité et la confidentialité de la navigation internet, que ce soit à partir de *Icecat* ou de *Firefox*, voici les conseils à suivre :

1. voir les modules complémentaires qu'on trouve sur Prism-break (`http://prism-break.org/`) ou sur `http://www.eff.org` et notamment le module « HTTPS Everywhere » (`https://www.eff.org/https-everywhere`) ;
2. le module *Adblock Plus* : `https://addons.mozilla.org/fr/firefox/addon/adblock-plus/` : indispensable, il bloque la plupart des publicités, ce qui change réellement le confort d'utilisation du Web ;

Sur quelques idées reçues concernant la navigation « privée » :

— quand on utilise l'option « Effacer mes traces » du navigateur Firefox, ce dernier ne fait pas mieux que de supprimer les fichiers. Certes les données sont devenues inaccessibles pour Firefox, mais elles sont toujours accessibles en regardant directement le disque dur.

— quand on utilise l'option « Navigation privée » du navigateur Firefox, ce dernier n'enregistre aucune information au sujet des sites qu'on a visités. C'est donc utile

en cas d'ordinateur partagé, mais cela ne garantit pas l'anonymat sur internet : le FAI (Fournisseur d'Accès à Internet) ou encore les sites visités eux-mêmes peuvent toujours garder trace des pages consultées.

4.3 Choisir un bon moteur de recherche

Je vous en propose deux : Startpage ou encore DuckDuckGo qui, tous les deux s'engagent à respecter votre vie privée. Voici ce que l'on peut lire sur le site de startpage :

> Chaque fois que vous utilisez un moteur de recherches courant, vos données de recherches sont enregistrées. Les principaux moteurs de recherche enregistrent votre adresse IP et utilisent des *cookies* de suivi pour enregistrer vos termes de recherche, la date et l'heure de votre visite, et les liens sur lesquels vous avez cliqué. Ils enregistrent ensuite ces informations dans une énorme base de données.
>
> Ces recherches révèlent une quantité choquante d'informations personnelles à votre sujet, comme vos intérêts, votre situation de famille, vos penchants politiques, votre état de santé, etc. Ces informations constituent aujourd'hui une mine d'or pour les sociétés de marketing, les autorités gouvernementales, les hackers et les criminels : tous ceux qui ne demandent qu'à mettre la main sur vos données de recherche privées.

4.4 Choisir un bon mot de passe

Les principes

Le choix du mot de passe est une recette savante qui va combiner majuscules, minuscules, chiffres et caractères spéciaux. Plus le mot de passe va contenir de caractères différents, donc majuscules, minuscules, chiffres ou tous les caractères possibles, y compris les caractères spéciaux, plus le nombre de mots de passe possibles va être élevé. Et c'est très important quand on sait qu'il existe des logiciels "casseurs de mots de passe" qui sont capables de tester un milliard de mots de passe à la seconde... Un bon moyen mnémotechnique : prenez les premières lettres d'une citation ou d'un vers. Par exemple : Rns2cifpà.

Gérer et stocker ses mots de passe

Comme il est prudent d'utiliser un mot de passe unique par service, il faut donc retenir un mot de passe pour accéder à son mail, un autre pour ouvrir sa session dans l'ordinateur et d'autres encore pour des services en ligne tels que des forums, des sites web, etc. La solution est d'utiliser un logiciel de gestion de mots de passe, qui permet de stocker, générer et gérer vos mots de passe en toute sécurité. Le principe d'un logiciel de gestion de mots de passe est le suivant : retenez un seul mot de passe, et nous nous chargeons de crypter tous les autres.

Prenons l'exemple de KeePass (https://keepass.info/), qui est un logiciel libre, gratuit et compatible avec tous les OS : Windows, Linux, Mac OS X. La première étape consiste à créer une base de données qui stockera tous vos mots de passe. Vous devez maintenant définir le mot de passe principal de votre base de données. C'est l'unique mot de passe que vous allez retenir. Puis vous pouvez saisir votre première entrée.