Internet, une base de données sans limite ?

Léo Guirauton
Malo Le Mestre
Tanguy Mossion
Louison Vincent
TD2

SOMMAIRE

INTRODUCTION 2

I - L'économie des données 4

1 – Qui sont ces géants du web qui contrôlent nos données ? 5

2 - Que savent les géants du web sur nous ? 7

3 - L'économie des données, modèle des géants du Web 8

4 - En quoi ces entreprises et leur modèle économiques empêche l'anonymat sur Internet 10

II - La collecte et le stockage des données 12

1 - La collecte des données 12

2 - Le stockage des données 14

III - L'aspect juridique des données d'Internet 19

1 - Les lois nouvelles qui protègent nos données 21

2 - Les conditions générales d'utilisation et politique de confidentialité 23

IV - L'anonymat sur Internet 26

Méthode pour se protéger 27

1 - Navigateur 27

2 - L'adresse IP 28

3 - La cryptographie 30

4 - Les précautions à prendre pour garder son anonymat sur le web 32

CONCLUSION 34

Sources 37

INTRODUCTION

Lorsque nous naviguons sur Internet, nous ne pensons pas forcément aux empreintes que nous pourrions laisser : des 0 et des 1, qui semblent à priori inoffensifs. Nous n'accordons peut-être pas assez d'importance à cela. Ainsi, certains géants du web profitent de notre ignorance ou de notre déni pour générer leur chiffre d'affaire. Nos données leur ont permis d'obtenir une puissance équivalente à celle de certains Etats. Suite à cette montée en puissance, des pays ont souhaité pouvoir communiquer. C'est pourquoi, en 2017, le Danemark a nommé un ambassadeur auprès de ces leaders du web.

Chaque action sur Internet, estompe notre anonymat. En effet, au fur et à mesure elles précisent un peu plus notre personnalité, nos goûts, nos projets, permettant de nous connaître mieux que nos proches. De ce fait, les grandes entreprises d'Internet sont les narrateurs omniscients du roman du XXIe siècle. Toutes ces données sont stockées dans des serveurs aux quatres coins du monde, sous de multiples formes, et engendrent une masse inimaginable de données.

Au préalable, il nous semble important de donner une définition de la notion de donnée. Dans le contexte de l'informatique et du web, celle-ci représente un élément d'information exploitable et pouvant être traité afin d'obtenir un résultat. Ici, nos données personnelles sont traitées, avec de puissant algorithmes toujours plus complexe, pour nous analyser dans divers objectifs commerciaux, statistiques ou bien politique.

Ainsi, la situation actuelle d'Internet nous amène à réfléchir sur la problématique suivante : pouvons-nous encore garantir notre anonymat sur Internet ? Si c'est le cas pourquoi est-ce important de le faire et par quelles méthodes ?

I - L'économie des données

Ou comment les géants du web génèrent-ils du profit en exploitant nos données utilisateurs et en quoi cela peut facilement porter atteinte à l'anonymat sur Internet

Dans l'optique de correctement répondre à notre problématique, il est important de poser un contexte particulier qu'est celui des géants du web.

Dans cette partie nous nous intéresserons aux raisons qui poussent les géants du web à empiéter sur notre vie privée, dans l'optique de générer du profit.

1 – Qui sont ces géants du web qui contrôlent nos données ?

Peu sont ceux qui ne connaissent pas encore ces géantes entreprises. 4 ressortent souvent du lot, ont les nomment les GAFA, il s'agit de Alphabet Inc (société mère de Google et coté à 723 Md$ en bourse en Novembre 2018), Apple Inc (838 Md$ en Novembre 2018), Facebook (387 Md$ en Novembre 2018), Amazon (741 Md$ en Novembre 2018). Cependant les géants du web ne se limitent pas à cette poignée d'entreprises.

Comment définit-on un géant du web ?

Ces derniers peuvent être catégorisés en fonction du volume de demande des services, du nombre d'utilisateurs ou bien encore en fonction du volume de données que l'entreprise possède.

Ainsi en 2017, le réseau social Facebook atteint 2 milliards d'inscription sur sa plateforme soit plus d'un quart de la population mondiale. En 2018, Facebook stock plus de 70 milliards d'images. Twitter comptabilise près de 400 millions de messages quotidiens. Chaque jour, les utilisateurs de YouTube comptabilisent 115 000 heures de visionnage de vidéos cumulées.

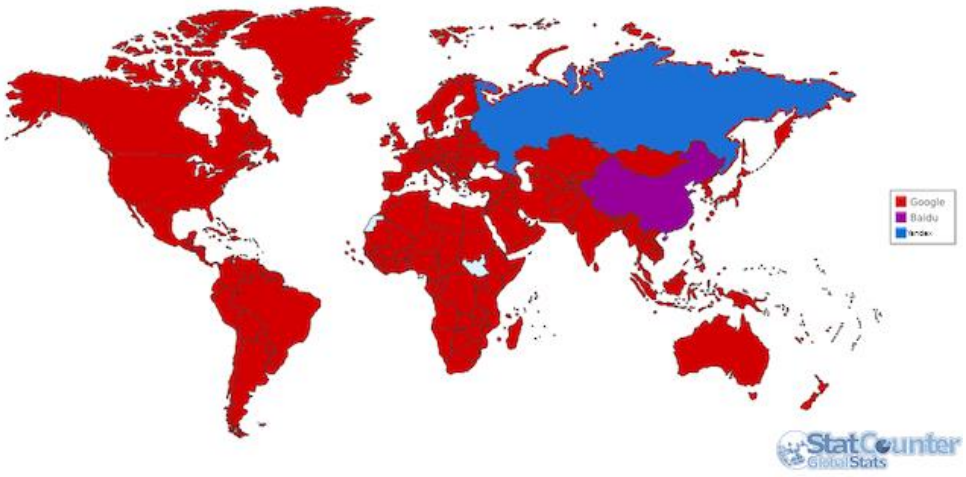
On voit ainsi que ces entreprises captent une très grande portion de l'activité superficielle d'Internet. Ce sont souvent les leaders dans leur domaine comme Oracle pour qui son logiciel Java est installé sur plus de 2 milliards d'appareil connecté à Internet.

Cependant, les entreprises sur lesquelles nous allons nous intéresser seront celles qui concentrent le plus nos données à caractère personnel. Ainsi nous nous pencherons sur le cas de la compagnie *Alphabet* Inc ainsi que du réseau social Facebook.

1.1 - Pourquoi avoir choisi de nous pencher sur c'est deux entreprises en particulier ?

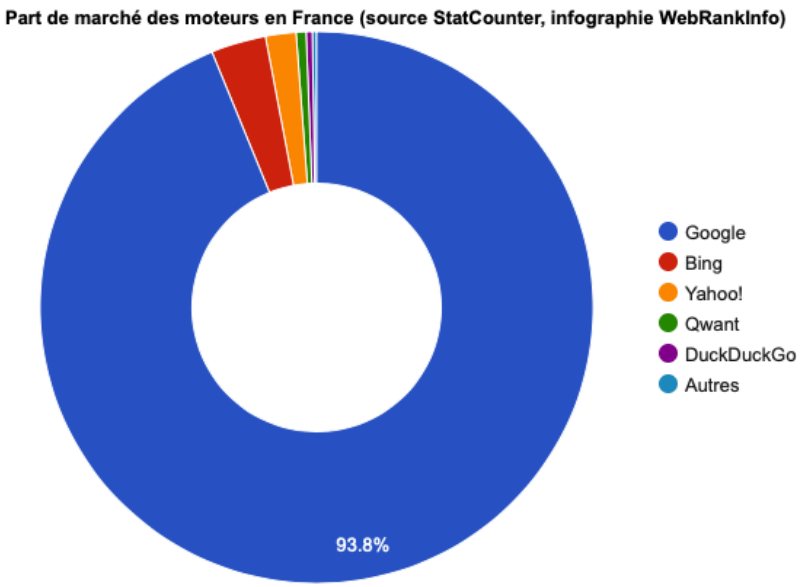
Alphabet Inc. est une entreprise californienne, créée en 2015 comme un conglomérat de sociétés précédemment détenues par la société Google.

Google search, son moteur de recherche, est incontestablement le leader mondial dans son domaine, comme le montre la carte qui indique la part de marché de Google dans chaque pays créée par StatCounter ou encore le diagramme ci dessous à l'échelle nationale. Les seuls pays où Google n'est pas leader sont la Russie (Yandex) et la Chine (Baidu).



Durant cette étude de cas, nous nous focaliserons aux domaines d'activités de Google et non d'Alphabet puisque c'est cette branche de la société qui exploitent les données utilisateurs pour générer un chiffre d'affaire. Les domaines d'activités de Google sont les suivant :

- Google search (moteur de recherche)
- Youtube
- Android
- Maps
- Ads
- Drive
- Apps



Ce diagramme circulaire met en évidence le monopole de Google sur le marché français.

En ce qui concerne Facebook, le réseau social est, malgré une chute de son nombre d'utilisateurs, de loin le plus grand réseaux social au monde en terme de nombre d'inscrits. De plus, le groupe Facebook se tient en excellente santé et possède 50 entreprises dont Instagram (1 milliard d'utilisateurs actifs), Messenger (1,3 milliards d'utilisateurs actifs), Whatsapp (1,5 milliards d'utilisateurs). On peut ainsi dire que Facebook est le réseaux social par excellence. Toutes les données potentielles que possède Facebook en font l'une des entreprises les plus valorisées au monde (6ème mondiale).

Malgré une diversification progressive des domaines d'activités des deux entreprises, Alphabet, la société mère de Google tire 80% de son chiffre d'affaire d'un modèle économique axé sur l'exploitation des données personnelle.

C'est la même situation du côté du réseau social Facebook, fondé en 2004 par Mark Zuckerberg et ses associés. En effet, la société puise 95% de son chiffre d'affaire dans les données qu'elle recueille. Ainsi, le choix de ces deux géants du Web nous aidera à bien mettre en évidence à quelle point nos données sont le pétrole du XXIème siècle et pourquoi leur grande soiffe de nos données peut facilement empêcher notre anonymat sur internet.

Le fait que ces entreprises dominent leur marché respectif ainsi que l'importante partie de leurs revenus proviennent de l'exploitation des données des utilisateurs, sont des arguments convaincant à l'étude de leurs cas.

2 - Que savent les géants du web sur nous ?

A l'image de Big Brother dans le roman 1984 de George Orwell, l'opinion publique voit de plus en plus les géants du web comme des puiseurs de données, qui sauraient plus de choses sur nous, que nous même. Est-ce déjà une réalité ?

Lorsqu'une compagnie du web reçoit les données que vous leur envoyez, comme par exemple quand vous postez un contenu sur un réseau social, ces données sont différenciables en 2 catégories. Premièrement, les données personnelles. Celles qui ont un lien direct avec l'individu, comme une photo ou un message par exemple.

Les données personnelles sont toujours accompagnées par le second type de données, les métadonnées. Celles ci ne sont donc pas considérées comme des données personnelles mais renforcent la valeur de ces dernières. On retrouve dans les métadonnées de nombreux types de données différents et qui dépendent du services, comme par exemple le système d'exploitation de l'appareil, l'heure d'envoie d'une donnée.

Facebook est le réseaux social par excellence, c'est une bibliothèque qui répertorie chacun de nous comme un livre ou l'historique de nos vies seraient inscrits.

Certaines entreprises sont très gourmandes quand ils s'agit de récolter nos données, et la se posent un problème de confidentialité.

Photo de profil Facebook de Mark Zuckerberg, PDG de Facebook

Sur la capture d'écran ci-dessus, on distingue en rouge les données personnelles et en bleu, les métadonnées relativ es. Quand vous postez un selfie sur Facebook, l'algorithme recueille 3 données principales : photo, texte, reconnaissance faciale. Il en

collecte 17 autres (modèle du téléphone, opérateur, niveau de batterie, réseau wifi, temps passé sur la plateforme, etc.

Cependant il faut savoir que les données recueillies une à une n'ont quasiment pas de valeur. Les données se transforment en pétrole dès lors que l'algorithme rassemble des millions de données du même type. Lorsqu'un réseaux social comme Facebook possède plus de 2 milliards de profils inscrits et que l'on analyse l'impressionnante dynamique de la plateforme, on peut enfin se rendre compte de la machine que cela représente.

Quant à Google, l'algorithme saisi toutes les recherches utilisateurs sur tous les services possibles (drive, maps, docs, search,). Lorsque vous possédez un compte Google, toute votre activité est trackée. Alors Google peut, sur le long terme et en liant toutes vos données, créer un modèle de vos habitudes, de vos centres d'intérêts, de votre entourage et plus encore.

3 - L'économie des données, modèle des géants du Web

3.1 - Définitions et contexte

Parmi les géantes entreprises d'Internet, est arrivé un modèle économique particulier qui valorise lui même un concept assez particulier : l'attention.

La production de ce à quoi nous pouvons porter attention à été démultiplié ces dernières années. Sur Youtube, la plateforme de vidéo d'Alphabet, il y a plus de contenu créé par minutes que l'on puisse regarder en une journée et plus de drama-story par secondes que de choses vraiment dramatiques. Il y a beaucoup d'information et peu d'attention, il y a donc une rareté.

Tout d'abord, il serait intéressant de définir le concept d'attention.

Ce dernier pouvant être défini comme un filtre qui nous permet de ne pas nous noyer dans un flux constant d'information, considérées comme inutiles à la survie de l'individu.

Dans le contexte de l'immense flux de données sur Internet et notamment sur les plateformes connues et que le plus grand nombre utilise (à savoir les géants du web), notre attention en tant qu'utilisateur est capté par le marketing puissant des entreprises. En effet, ces dernières l'ont bien compris, notre attention est vulnérable et nous ne la contrôlons que partiellement, elle est en partie instinctive et se base sur des réflexes.

3.2 - De l'attention aux dollars : comment l'économie de l'attention génère-t-elle du profit ?

Cette attention, il est possible d'en tirer profit, et c'est ce qui exploitent les géants du web. Ils mettent en place un marketing puissant s'inspirant de la psychologie comportementale.

Il devient parfois de plus en plus difficile de discerner un contenu d'une publicité (exemple de Snapchat ou de BuzzFeed).

3.3 - Pourquoi nos données intéressent tant ces entreprises ?

Comme nous l'avons vu dans la première sous partie, 95% du chiffre d'affaire de Facebook provient d'un modèle économique relatif au données et il s'agit de 80% du chiffre d'affaire d'Alphabet, même si cette dernière tend à diversifier son activité

Le profit de ces entreprises à haute valorisation boursière se fait donc très majoritairement grâce aux données acquis de leurs utilisateurs.

Cependant comment peuvent-ils tirer un profit financier à partir de ressources non financières ?

Il y a plusieurs façon de générer un profit à partir des données utilisateurs, la plus connue actuellement étant certainement les publicités ciblées.

Nous n'aborderons pas le principe de référencement des sites web par Google, qui est pourtant une très grande source de revenu, puisque ce dernier se voit être un modèle économique exclusif au moteur de recherche américain.

Cependant, ce qui peut être plus angoissant, c'est que parfois ces données collectées ne sont pas utilisées par l'entreprise elle-même. Vos données personnelles peuvent être vendues sur des places de marché à d'autres entreprises avec lesquelles vous n'êtes pas en contact. C'est notamment le cas des sites qui proposent des services gratuits, comme Facebook ou Google. Le réseau social n'est pas payant et ne vend pas de produits, il doit donc trouver un moyen de se rémunérer : en vendant des informations personnelles. Les ventes se font notamment à des annonceurs publicitaires qui peuvent ainsi fournir de la publicité ciblée qui a plus de chance d'atteindre l'utilisateur.

3.4 – Comment se servent-ils de nos données pour générer un chiffre d'affaire ?

Quand vous ne voyez pas le service, c'est que c'est vous le produit ! C'est en effet comment on pourrait illustrer le modèle économique de certaines entreprise, comme nos deux cas d'études, Google d'Alphabet et Facebook.

Des études montrent que la plupart des utilisateurs de Facebook et de Google ou de tout autres services, ne savent pas qu'ils abandonnent tout droit sur les informations quand ils acceptent les fameuses Conditions Générales d'Utilisation de l'entreprise. C'est ce que nous verrons dans la partie III.

4 - En quoi ces entreprises et leur modèle économiques empêche l'anonymat sur Internet

Au-delà du piratage reste la question de l'utilisation légitime des données personnelles. L'utilisation la plus commune est l'envoi de publicité ciblée à l'utilisateur. C'est l'exemple typique où après réservation d'un billet de train pour Paris, toutes les publicités affichées concernent des logements à Paris.

En attendant, reste la possibilité de demander. N'importe quel internaute peut obliger l'entreprise à lui communiquer les informations dont elle dispose sur lui dans les deux mois. C'est un droit que tout citoyen a, mais que peu utilisent.

Attention, connaître son profil ne donne pas forcément le droit de le supprimer. Autrement cela aboutirait à des situations impossibles. Vous pourriez par exemple écrire à la direction des impôts pour supprimer toutes les informations vous concernant et échapper à l'impôt.

Enfin, beaucoup d'internautes sont contents d'obtenir de meilleurs services grâce à leurs données personnelles, ils préféreraient juste qu'elles ne soient pas vendues sans leur "accord". La plupart disent ne rien avoir à cacher et pensent être invulnérables, cependant on ne sait pas comment nos données peuvent être ré-utilisées dans l'avenir. Mais comme les géants du web peuvent-ils se permettre de stocker et conserver toutes les données qu'elles captent, il est concevable que ces données soient, après avoir été vendues, retournées contre certaines personnes lors d'un procès ou dans d'autres situations compromettantes. On se rapproche de plus en plus de scénarios de science fiction ou une entités possède un pouvoir d'omniscience sur le reste du monde. La croissance des GAFA devrait se prolonger sur une long terme, et ces derniers sont pourtant déjà plus riches et plus puissants que certains pays,

La série télévisée Black Mirror met très bien en garde contre ce genre de scénario et bien d'autres encore à propos des abus d'usage des technologies du numérique à l'avenir.

Dans les prochaines parties, nous verrons comment les données utilisateurs sont stockées par les entreprises du web

II - La collecte et le stockage des données

1 - La collecte des données

Une des premières conditions à accepter lorsqu'on navigue sur Internet est le fait qu'utiliser un service c'est en accepter toutes les clauses.

En effet, il faut retenir un point : les messages en bas de page parlant d'accepter les cookies ou la collecte d'information n'ont aucune valeur juridique.

Par exemple, si vous utilisez Facebook, vous acceptez que l'entreprise collecte les informations qu'elle souhaite et ce même si c'est un photo mise en "privé" ou encore des informations issues de messages dit "messages privés". Il faut retenir qu'une donnée personnelle n'appartient pas à l'individu, elle le concerne, il faut bien faire la différence et la garder en tête.

De manière plus générale, lorsqu'un individu s'inscrit sur un site à l'aide de son adresse électronique, son prénom, son nom ou encore parfois d'autres informations telles que le numéro de téléphone par exemple, il faut savoir que toutes ces données sont conservées par le site.

C'est grâce à ces données et aux algorithmes de traitement de données que les sites sont capables de nous comprendre et de nous analyser. Par exemple, ces algorithmes sont capables de comprendre nos intentions d'achats ou nos goûts, sans que nous n'ayons eu le choix de communiquer des informations.

Cependant, comme nous l'avons vu précédemment, il ne faut pas avoir peur que ces informations servent à espionner l'utilisateur, en effet, celles-ci sont surtout vendues à des sites dans l'objectif de réaliser de la publicité ciblée.

Un des majeurs problèmes concernant la collecte d'information est le manque de connaissances de l'utilisateur sur le sujet. Pourtant, depuis la loi informatique et libertés de 1978 il est obligatoire d'informer les utilisateurs de la collecte de leurs données.

1.1 - Les cookies, un moyen de collecte de données

Un des outils utilisé afin de collecter les données et ensuite de les utiliser est nommé les cookies. Un cookie est un fichier contenu dans le disque dur de l'ordinateur de l'utilisateur et exploitable par le navigateur Internet. Ce fichier permet au serveur web de le reconnaître d'une page à l'autre.

De cette manière, lorsqu'un utilisateur se connecte sur un site et va naviguer dessus celui-ci va retenir un maximum d'informations. Par exemple, sur un site de de shopping, le site va retenir sur quelles sections l'utilisateur navigue et les stocker dans des cookies. Ce sont ces cookies qui vont expliquer le fait que l'utilisateur ait de nombreuses publicités montrant des baskets si l'utilisateur a navigué dans des sections dédiées à cela sur le site de shopping.

Cependant, il ne faut pas s'inquiéter à propos de l'espionnage ou du vol de données personnelles car les cookies ne doivent pas contenir directement les informations et leur durée de vie doit être limitée. Enfin, le cookie ne contient en aucun cas des informations que l'utilisateur n'a pas données ou des des informations sur le contenu de son ordinateur. Autrement dit les cookies ne peuvent pas collecter d'informations concernant le système de l'utilisateur. En conclusion, les cookies n'ont rien de dangereux pour l'utilisateur s'ils sont bien conçus.

1.2 - Snapchat : le cas d'une application mobile

Nous avons vu divers moyens de collecter des informations et nous allons maintenant parler d'un exemple d'application pratiquant la collecte d'informations personnelles : Snapchat.

Snapchat est une application disponible gratuitement sur les diverses plateformes de téléchargement d'applications telles que AppStore ou encore Play Store. Celle-ci permet de partager des photos et des vidéos à une liste de contact ou à quiconque nous à ajouter à sa liste d'amis. Snapchat a été développée par des étudiants de l'université Stanford (Californie). En 2018, Snapchat représente 188 millions d'utilisateurs actifs dans le monde (dont 13 millions en France).

Les utilisateurs de l'application sont relativement jeunes. Ceci peut s'expliquer par le fait, notamment, que l'âge requis pour télécharger et utiliser cette application est fixé à 13 ans. La plus importante spécificité de l'application est le fait que le temps pour regarder la photo ou la vidéo peut être limitée. En effet, l'utilisateur envoyant le média peut fixer le temps limite pour le visionner. Cette période de temps peut par exemple être d'une seconde, dix secondes, mais aussi, sans limite de temps (cela n'était pas disponible au début de l'application).

Cependant, il est important de ne pas confondre le temps d'accessibilité au média pour les utilisateurs de l'application et le fait qu'elle ne soit pas conservée par *Snapchat Inc.*. C'est notamment grâce aux informations qu'elle conserve que l'entreprise est estimée à 24 milliards de dollars en 2017.

2 - Le stockage des données

Une fois collectées, les données personnelles sont envoyées vers le serveur de stockage de données d'une entreprise. Ces serveurs peuvent prendre différentes formes que nous verrons ultérieurement dans cette partie. Contrairement à ce que peut penser l'idée générale, les données sont toujours stockées dans un lieu physique. Il peut se trouver n'importe où dans le monde. Cependant, l'utilisateur ne peut pas savoir où se trouve ce lieu physique et de ce fait ne sait pas quelles lois sont appliquées à cet endroit.

Si le lieu de stockage est situé en France ou dans un autre pays de l'Union Européenne alors c'est la législation européenne qui est appliquée. Les données stockées dans l'Union Européenne ne peuvent en sortir qu'avec l'autorisation de la CNIL (Commission Nationale Informatique et Libertés).

Nous allons maintenant voir des formes que peuvent prendre les serveur de stockage de données.

2.2 - Un serveur de stockage connecté au réseau local de l'entreprise

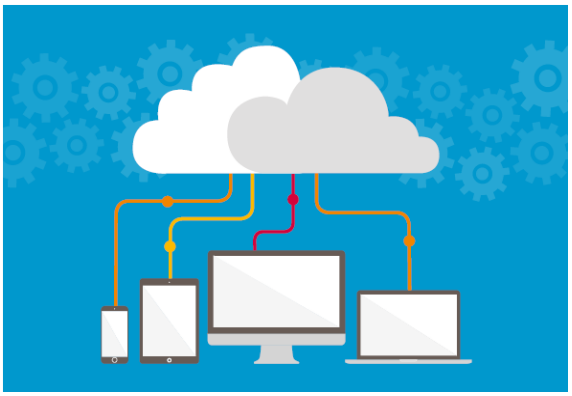
Ce type de serveur doit fonctionner continuellement et doit être composé de matériels fiables en plusieurs exemplaires afin de maintenir un service toute la journée quelque soit l'heure. Ce serveur permet le partage de fichier, de données ou de bases de données, il peut gérer les e-mails de l'entreprise et le fait que les utilisateurs du réseau puissent se servir de périphériques communs tels que les imprimantes. Ce type de serveur peut aussi servir d'hébergeur de site Internet.

Dans ce type on retrouve un sous-type permettant seulement le partage de fichiers. Ces serveurs sont appelés les Network Attached Storage (NAS) ou en français "serveur de stockage en réseau". Ce type de serveur permet, en plus du partage de fichiers entre utilisateurs du réseau, d'accéder à ces fichiers depuis n'importe quel réseau via Internet.

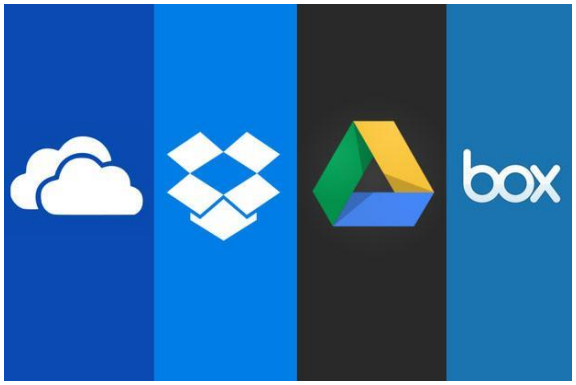
2.3 - L'utilisation d'un Cloud pour le stockage de données

L'utilisation d'un cloud désigne le fait de se servir d'Internet pour stocker des données. Ces données "stockées sur Internet" sont en réalité stockées sur des disques durs de serveurs dans des datas centers (en français centre de données ou encore centre de traitement de données) qui peuvent se trouver n'importe où dans le monde. Un data center est un espace centralisant des données informatique par le regroupement d'un grand nombre de machines. Les possesseurs de ces serveurs sont des grandes entreprises telles que Google, Amazon, Apple ou encore Microsoft. Un data center permet la sauvegarde de données de manière durable avec des équipements de haute qualité.

On comprend avec cette explication du fonctionnement des clouds que les grandes entreprises possédant les serveurs de stockage de données stockent leur propre données aussi dans leurs serveurs.



Schématisation du fonctionnement d'un cloud



Des exemples de Cloud : OneDrive, DropBox, Google Drive, Box FR

2.4 - Le cas de Google

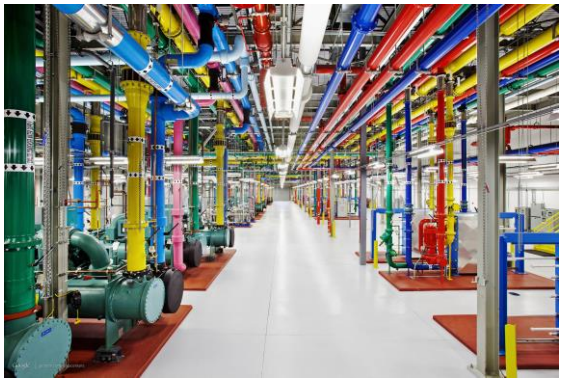
Nous allons donc parler des lieux de stockage d'une de ces grandes entreprises : les data centers de Google.

En effet, Google comme de nombreuses autres grandes entreprises dispose de data centers permettant de stocker leurs propres données mais mettent aussi leur service à la vente pour d'autre entreprises. Cela fait plus de 10 ans que Google s'est lancé dans la création de centres de données et ils font partis des plus fiables et économiques.

Google est la première entreprise de services Internet à avoir acquis des certifications externes élevées pour le respect de normes sur l'environnement et la sécurité des locaux et des employés (dans tous ses centres de données aux États-Unis). Google a notamment obtenu les certifications ISO 14001 et OHSAS 18001. Pour Google, l'ISO 14001 concerne un ensemble de normes destinées à aider les sociétés à minimiser l'impact environnemental de leurs opérations. Quant à elle, l'OHSAS 18001, définit des normes de sécurité et de santé des employés.

Centre de données de Google à Singapour

De plus, depuis quelques années, Google a axé sa communication vers un axe environnemental. C'est pour cette raison que "Google possède l'une des infrastructures de centres de données les plus écologiques de toutes les entreprises du monde" (- Datacenter Dynamics). Pour mener des actions écologiques Google a notamment amélioré l'économie de ces centres de données, en effet ils utilisent 50% d'énergie en moins que les centres de données habituels. Google utilise aussi des stratégies de recyclage de l'eau visant à réduire l'impact de ses installations sur les communautés locales. Pour finir l'utilisation d'énergies renouvelables (ex : énergie éolienne, énergie solaire) est aussi très importante car elles alimentent 30% des opérations de Google.



2.5 - Le vol de données

L'intérieur des centres de données de Google

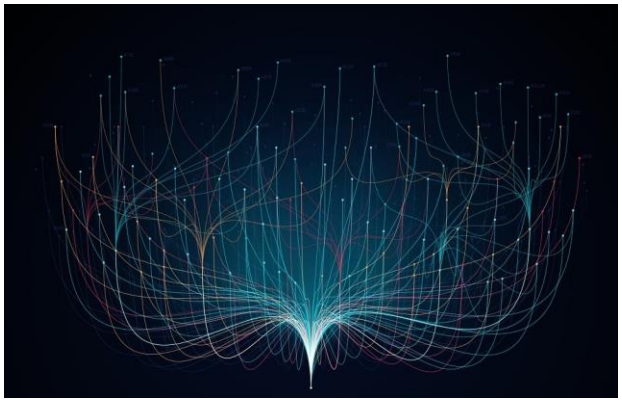
Maintenant que nous savons que nos données sont stockés dans un lieu physique nous pouvons nous demander s'il existe un fort risque de vol de données.

En effet, on peut se demander si n'importe quelle personne ayant plus ou moins des connaissances informatiques est capable de voler des données personnelles d'utilisateur. Il faut savoir un point important : dans leur lieu de stockage presque aucune donnée n'est cryptée, exceptés les mots de passe. Le cryptage est un "système de protection informatique destiné à garantir l'intégrité et l'inviolabilité de données pendant leur transmission ou leur stockage" (- *Larousse*). Cependant, lors du trajet, les données sont en majeure partie cryptées afin d'éviter qu'une personne ayant des intentions malveillantes puisse les intercepter.

Il y a donc bien plus de données qui sont cryptées durant le trajet que dans leur lieu de stockage. A contrario du stockage de l'argent il est donc plus simple de voler les données une fois qu'elles sont stockés que lors du trajet.

2.6 - Le concept de Big Data

Nous venons de voir que lorsqu'on navigue sur Internet la collecte de données est omniprésente, et, avec elle s'accompagne le stockage de ces données. Si la collecte est ubiquiste on peut alors supposer que la quantité de données collectées soit immense. C'est en effet le cas car de nos jours environ 2,5 trillions d'octets de données sont créés tous les jours. Cette masse de données est appelée "Big Data" ou en français mégadonnées. Ce terme de Big Data "désigne des ensembles de données devenus si volumineux qu'ils dépassent l'intuition et les capacités humaines d'analyse et même celles des outils informatiques classiques de gestion de base de données ou de l'information" (- *Wikipédia*). Cette notion est donc apparue avec l'explosion quantitative des données numériques. Le Big Data permet à tous les utilisateurs d'Internet d'accéder à des bases de données géantes.



Représentation schématique du Big Data

Nous avons donc mis en évidence que la collecte des données est omniprésente lors de la navigation sur Internet et nous avons aussi vu où celles-ci sont stockées et dans quelles conditions. La collecte étant présente partout, sur n'importe quelle page Internet, on peut alors se demander si cela empêche pour autant l'utilisateur d'être anonyme. Nous pouvons nous demander s'il existe divers moyens, pour se protéger, que nous pouvons réaliser nous même ou encore s'il existe des lois protégeant nos données et la manière dont elles sont utilisées.

III - L'aspect juridique des données d'Internet

1 - Les lois nouvelles qui protègent nos données

Depuis plus de 20 ans les lois sur la protection des données se multiplient en France, mais aussi à l'internationale. En effet, plus de dix lois ont été votées en France pour accroître la sécurité et la privacité des données des citoyens français. Nous allons voir quelles sont les dernières lois qui ont renforcées la protection des données des utilisateurs européens, et la dernière loi qui a permis la collecte massive des données de millions d'utilisateurs. Ces lois ne sont, toutefois pas acceptées à l'unanimité, et critiquées sur certains aspects.

1.1 - Bouclier de protection des données aux Etats-Unis et en Union Européenne

Cette loi sur la protection des données, négociée en 2015-2016, est un accord entre les Etats-Unis et l'Union Européenne. Il remplace le *Safe Harbour*, qui a été invalidé en 2015. Cette accord renforce la protection des données des citoyens européens.

Les entreprises tierces qui traitent les données des utilisateurs sont désormais plus contrôlées par les instances américaines.

L'accès aux données des citoyens européens est désormais que pour des raisons de sécurité nationale. Ce droit permet donc aux Etats-Unis de vérifier toutes les données des citoyens européen. En effet, ils peuvent ainsi inspecter toutes les données qui circulent en espérant en trouver quelques une menaçant la sécurité de leur pays.

Les citoyens européens peuvent désormais intenter un procès pour une violation du droit sur la protection des données personnelles contre les entreprises américaines.

Cependant, cet accord n'empêche pas les collectes massives et systématiques des données personnelles par les autorités américaines, bien au delà du strict nécessaire. De plus plusieurs associations françaises dénoncent le fait qu'un citoyen européen ne peut accéder à un recours juridictionnel effectif, donc il ne pourra porter plainte contre les instances institutionnelles américaines.

Cette accord reste très flou même au sein du groupe de travail chargé de la création de cette loi. En effet, Isabelle Falque-Pierrotin, présidente de ce groupe de travail, précise qu'il est difficile de comprendre tous les documents et les annexes qui composent cette loi, car certains se contredisent.

1.2 - RGPD, qu'est-ce que ça change?

La RGPD, qui est un règlement sur la protection des données de l'Union Européenne. Ce règlement a pour but de renforcer et d'unifier la protection des données pour les individus au sein de l'Union européenne. Ce texte a été adopté le 14 avril 2016, après quatre années de négociation. Cette loi remplace la directive qui datait de 1995, dépassée depuis longtemps.

Affiche de l'Union Européenne sur la RGPD

Ces principaux objectifs sont d'accroître à la fois, la protection des particuliers concernés par un traitement de leurs données et la responsabilisation des acteurs de ce traitement. Ces principes peuvent être appliqués grâce à l'augmentation du pouvoir des autorités européennes et nationales de régulation.

Cette loi modifie de nombreux aspects de la précédente directive. En effet, celle-ci harmonise les différentes politiques appliquées dans les 28 états membres de l'Union Européenne, et met ainsi fin à la fragmentation juridique entre les Etats.

De plus, elle s'applique aux entreprises basées hors de l'Union Européenne. Ces sociétés sont également soumises au règlement dès qu'elles ciblent les résidents de l'Union Européenne par le profilage ou proposent des biens et services à des résidents européens. Cette loi cible toutes les grandes entreprises du net qui profitaient de l'absence de loi claires sur leur statut pour revendre les données de leurs utilisateurs. Cette loi bloque actuellement plusieurs site qui n'ont pas mis à jour leurs politiques de confidentialité pour protéger les utilisateurs.

La RGPD donne des sanctions plus importantes, à hauteur de 20% du chiffre d'affaires annuel mondial d'une entreprise, ce qui est suffisant pour la mettre grandement en danger.

Cette loi permet un changement majeur, le droit à l'oubli et à l'effacement. En effet, tout utilisateur peut contraindre une entreprise à effacer les données personnelles des utilisateurs. Concrètement, cela permet aux individus concernés de demander aux référenceurs de retirer certaines informations anciennes si celles-ci sont nuisibles à l'utilisateur.

La RGPD exige que chaque utilisateur puisse avoir accès à tout instant à un fichier comprenant toutes ses données stockées en ligne. Malheureusement ce dossier est très difficile d'accès, en effet il est très difficile pour une entreprise de regrouper toutes nos données, et ainsi cette entreprise fait en sorte que cette procédure soit longue et compliqué pour décourager les utilisateurs qui voudront accéder à leur dossier.

Toutefois cette loi est très controversée, surtout au niveau l'article 13. En effet, cette article demande aux entreprises d'Internet de réaliser un filtrage automatique de tous les contenus que l'utilisateur téléverse sur leurs plateformes. Ainsi, beaucoup s'inquiètent du fait qu'Internet passe d'une plateforme ouverte de partage et d'innovation à un outil de surveillance et de contrôle automatisé de ses utilisateurs.

L'article 22 est aussi critiqué car il interdit le profilage sauf si l'utilisateur donne son consentement explicite. De ce fait, si l'on veut se créer un compte sur Facebook, Google, ou un autre site, on doit accepter la politique de confidentialité qui contient un traitement de nos données, et donc accepter le profilage.

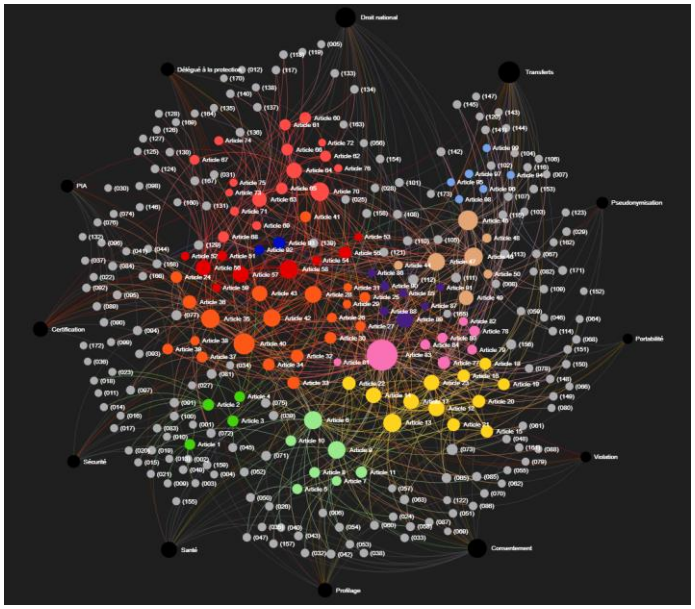


Schéma montrant l'interaction entre les différentes annexes et articles de la RGPD

1.3 - La loi américaine qui facilite la collecte de nos données

La majorité des entreprises qui gèrent les données d'une grande partie des utilisateurs sont basées aux Etats-Unis. Par conséquent, le droit américain s'applique à ces entreprises. Malheureusement, ce droit est peu contraignant sur la consultation des données des utilisateurs. Notamment le *USA Patriot Act* voté en 2001, en réaction aux attentats du 11 septembre 2001.

Ce texte permet aux autorités de la sécurité d'accéder aux données informatiques détenues par les particuliers et les entreprises, sans autorisation préalable et sans en informer les utilisateurs. Ce texte a été prolongé deux fois, jusqu'en 2015.

Cependant cette loi est très critiquée. En effet, elle porte atteinte aux libertés individuelles, elle diminue la liberté d'expression, et permet aux autorités des Etats-Unis de demander aux entreprises américaines de leurs fournir des données, dites sensibles, même si celle-ci sont stockées en Europe, et donc concernent les citoyens européens.

De plus, les lobbys des GAFA sont extrêmement puissants aux Etats-Unis. Ils ont ainsi pu faire annuler ou limiter, de nombreuses lois américaines à l'encontre de leurs business model, basé sur les données des utilisateurs.

2 - Les conditions générales d'utilisation et politique de confidentialité

Les CGU, ou conditions générales d'utilisation, sont les règles qui régissent l'utilisation d'un service en ligne entre le service et ses utilisateurs. Elles sont souvent liées avec une autre règle, la politique de confidentialité. Nous acceptons celles-ci lors de l'inscription sur un service en ligne, sans le lire, dans la plupart des cas. Ainsi en les acceptants, on accepte la collecte, l'utilisation de nos données. Ces conditions utilisent souvent des termes compliqués, des paragraphes assommant, et bien d'autres stratagèmes pour nous éviter de nous renseigner sur le véritable sens cachés, et leurs conséquences.

En effet, selon un sondage OpinionWay, 7 français sur 10 ne lisent peu ou pas les conditions générales d'utilisation. De plus, la moitié des sondés ne comprennent pas le sens des textes concernant les conditions générales d'utilisation et la politique de confidentialité.

Nous allons donc voir l'évolution de ces textes, notamment ceux des GAFAs, des années 2000 à nos jours.

2.1 - Les années 2000

Le début des années 2000 marque un tournant dans la législation des données sur Internet. En effet, on assiste à une prise de conscience générale sur les traces que nous laissons sur Internet et leurs possible utilisation.

Apple

Apple est une entreprise se souciant de la privacité des données laissées par ces utilisateurs. En effet, la seule utilisation des données générées sera soit pour collecté des données sur le fonctionnement de leurs système ou dans un cadre légal. C'est à dire pour prévenir ou prendre des mesures concernant les activités illégales.

Google

La politique de confidentialité de Google décrit deux choses totalement différentes entre les conditions générales d'utilisation de l'an 2000 et celles de l'an 2001.

En effet, celles de 2000, dit utiliser les cookies pour stocker les préférences de l'utilisateur, et permet de savoir quel est l'ordinateur associé à ces cookies mais ne peut pas dire qui est cet utilisateur. Alors que celles de 2001, stockent toujours les cookies et suit les tendances des utilisateurs en fonction de leurs recherches. Ces données ne seront pas divulguées sauf par un processus légal et valide. Ces deux politiques de confidentialité montrent ainsi deux choses différentes : la première que vous êtes totalement anonyme alors que la seconde montre que vous ne l'êtes absolument pas.

Les programmes de surveillance américains

Suite aux attentats du 11 septembre 2001 qui ont profondément marqué les Etats-Unis, les années 2000 sont marquées par des programmes massifs de surveillance américains.

Cela se traduit en 2002 par le *Total Information Awareness*, un programme ayant pour but de collecter et de traiter toutes les données générées sur Internet pour prévoir les crimes et les actes de terrorisme, et donc les éviter. Ce programme fût un grand échec, dû à une communication ratée et un titre ne présentant rien de bon aux yeux de la population.

Cependant, l'échec du programme précédent ne va pas décourager les instances américaines dans leur but de contrôler la circulation des données d'Internet. En effet, en 2005, la *National Security Agency* devient celle que nous connaissons maintenant : une agence de surveillance massive des données générées sur Internet mais aussi des conversations téléphoniques du monde entier. Cette agence de surveillance va accomplir exactement la même chose que le *Total Information Awareness*, avec un budget de plus de

10 milliards de dollars américains. Cette agence est toutefois très critiquée sur cette omniscience, ces pratiques douteuses sont dénoncées depuis sa création jusqu'à aujourd'hui, mais cela ne semble pas arrêter la volonté américaine de tout savoir sur tout le monde, en violant l'espace privé de chaque personne connectée à Internet.

2.2 - Les années 2010

Google

En 2012, Google modifie sa politique de confidentialité et rassemble toutes vos données sur un seul profil. Cela confirme donc que nos données stockées par Google sont associés à un nom.

Facebook

Facebook change les paramètres de partage sans prévenir ses utilisateurs : les informations partagées sont prioritairement publiées en public au lieu du privé comme choix de base. Cela entraîne une forte publication d'informations non consenties par les utilisateurs, qui ne se sont pas rendus compte du changement. Ainsi, au fur et à mesure des années, Facebook rend le partage d'informations de plus en plus public comme choix initial. En 2010, tout est partagé par défaut sauf les contacts et notre date de naissance.

De plus, des experts en droit dénoncent le fait que Facebook soit traité comme un objet d'utilité public, alors que l'on devrait traiter celle-ci comme une entreprise. Ainsi, lorsqu'une loi va à l'encontre du business model de cette entreprise, elle lance des armées d'avocats pour que cette loi affecte le moins possible les revenus de Facebook sur l'utilisation des données des utilisateurs.

Comme nous l'avons vu, les lois régissant les données personnelles sont nombreuses, et ont des buts différents. Elles sont votés dans un concept particulier qui ne convient pas à l'évolution constante d'Internet, d'où l'importance de les renouveler régulièrement.

Ainsi les lois restent un bras de fer constant entre les Etats et les entreprises basant leur commerce sur les données, les premiers voulant protéger les utilisateurs et contrôler Internet pour leur sécurité nationale, et les seconds pour une libéralisation totale d'un Internet neutre. Nous pouvons espérer que de nouvelles lois complèteront les actuelles, dans l'optique de protéger nos données face aux subterfuges que les GAFA appliqueront pour contourner ces dernières.

IV - L'anonymat sur Internet

L'anonymat sur Internet est un problème connu mais peu de gens prennent des précautions à ce sujet. L'opinion publique pense que le fait de rester anonyme sur Internet est un état inatteignable réservé seulement à l'élite, et c'est une idée reçue. Certes un anonymat complet est impossible mais le but à atteindre est de décourager celui qui veut prendre vos informations personnelles, pour cela il faut créer un maximum d'obstacles.

Définition de l'anonymat

L'anonymat est un état, c'est lorsque l'on ignore l'identité d'une personne.

Sur Internet est le fait de séparer sa vie sur Internet de sa vraie vie, une personne sur Internet ne peut pas savoir qui vous êtes en vrai et vice versa.

L'anonymat complet est impossible

A chaque fois que l'on se connecte à Internet, on laisse des traces de notre passage. Que ce soit inconsciemment avec notre adresse IP ou volontairement parce que nous avons donné des informations sur nous en s'inscrivant sur un site : partager des photos de nous, donner notre vrai nom ou des informations pouvant être reliées à notre identité.

Ainsi le simple fait de se connecter nous ferait perdre notre anonymat complet, également beaucoup d'entre nous ont déjà baissé leur vigilance et ont partager des informations sur eux un jour.

De ce fait, l'anonymat complet n'existe pas.

Méthode pour se protéger

Le fait de protéger son anonymat sur Internet est quelque chose qui peut se faire simplement, et gratuitement. Ces méthodes peuvent être appliqué séparément en fonction du type de navigation que l'on utilise.

1 - Navigateur

1.1 - Historique & Cookies

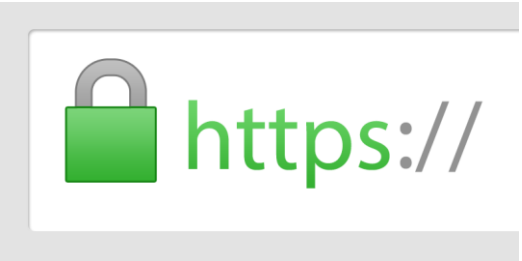
Le navigateur permet deux types de navigation, une dite publique et l'autre dite privée. La navigation privée permet d'empêcher votre ordinateur de stocker vos données de navigation sur votre ordinateur, ainsi une personne utilisant le même ordinateur que vous ne saura pas quel site vous avez visité. Cependant ce type de navigation n'empêche en aucun cas votre Fournisseur d'accès de connaître votre passage sur le net.

De la même façon que l'historique, les Cookies peuvent ne pas être enregistrer sur votre ordinateur. Néanmoins, leur suppression n'empêchera en rien votre employeur ou votre fournisseur d'accès de savoir ce que vous avez fait.

1.2 - Le protocole HTTPS

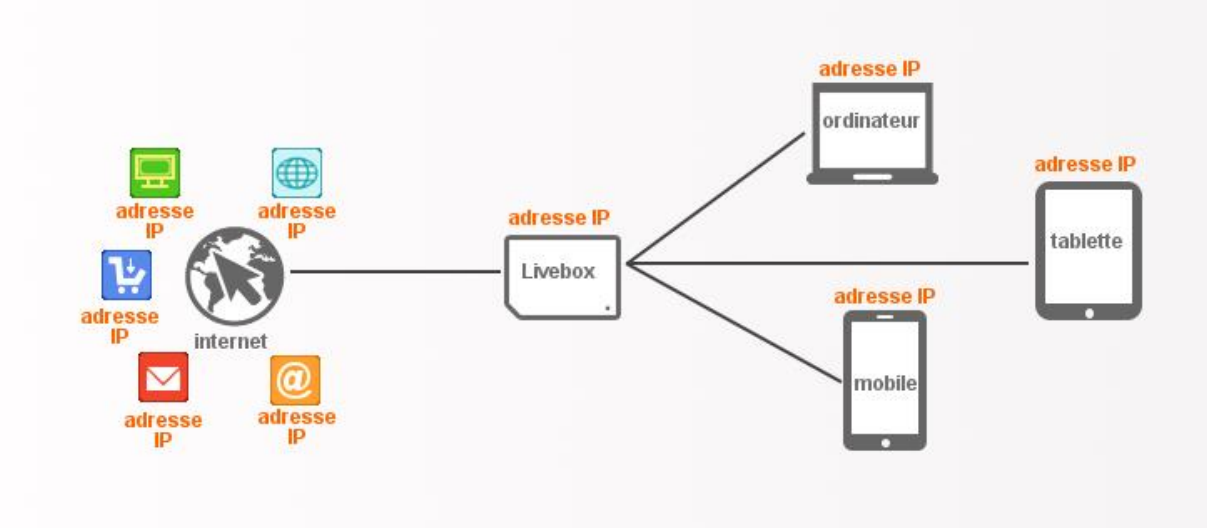
L'HTTPS, ou HyperText Transfer Protocol Secure, est une mention que l'on voit souvent au début de l'URL d'un site, mais on voit également souvent son cousin HTTP. La différence est que le premier est permet de chiffrer les données envoyées entre le site et l'utilisateur (entre le serveur et le client), la communication est ainsi rendu plus difficile pour un hacker de prendre vos données. Le HTTPS empêche donc qu'un serveur intermédiaire utilise une mémoire cache pour mémoriser l'information.

Il existe des extensions de navigateur qui permettent de forcer une transmission d'information via HTTPS, les sites en HTTP sont donc forcés de d'envoyer et de recevoir des données chiffrées.



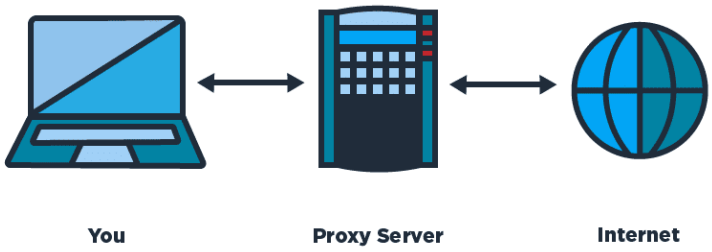
2 - L'adresse IP

L'adresse IP est en quelque sorte la « carte d'identité » de notre connexion sur Internet, chaque site garde en mémoires les adresses IP qui ont visité la page, ce sont les « logs ». Si quelqu'un a accès à ces logs, alors il aura accès aux adresse IP des utilisateurs, donc il saura qui est connecté, où et quand.



2.1 - Les serveurs proxy

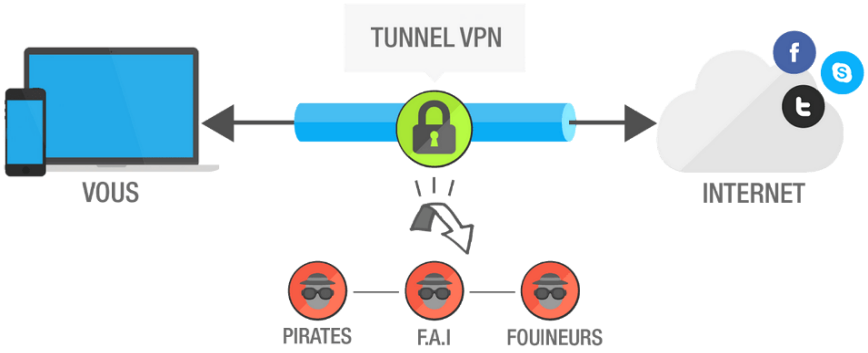
Un serveur proxy est un ordinateur qui permet à un utilisateur de posséder une connexion indirecte à Internet. Ainsi les données envoyées par l'utilisateur passent d'abord par le serveur proxy puis vers le site Internet et inversement le site enverra les données au proxy, et le proxy les enverra à l'utilisateur. Cependant le proxy ne crypte pas le trafic de données, il utilisera donc le protocole HTTP et non HTTPS, cette caractéristique lui permet d'être utilisable par un grand nombre d'utilisateurs simultanément.



Les serveurs proxy sont disponibles gratuitement ou en payant. Cependant, les proxy gratuits sont utilisés par des milliers d'internautes en même temps donc on peut potentiellement utiliser la même connexion qu'un utilisateur effectuant des actions illégales.

2.2 - L'utilisation de VPN

Le VPN (Virtual Private Network) est une technologie qui crée une connexion sécurisée pour l'utilisateur. Cependant, il ne passe pas par un serveur intermédiaire comme pour le proxy. Le VPN fabrique un tunnel crypté entre le serveur et le client, ainsi le lien entre votre IP et celle qui est utilisé via le VPN n'est connu que par le VPN. Le fournisseur d'accès internet n'aura ni accès à l'adresse IP utilisée ni aux sites visités.



Tout comme les serveurs proxy les VPN sont disponibles gratuitement ou en payant, mais les VPN gratuits gardent en mémoire les liens entre les IP et sont susceptibles de les vendre alors que ceux qui sont payant n'enregistrent pas ces données.

2.3 - Tor, un réseau superposé

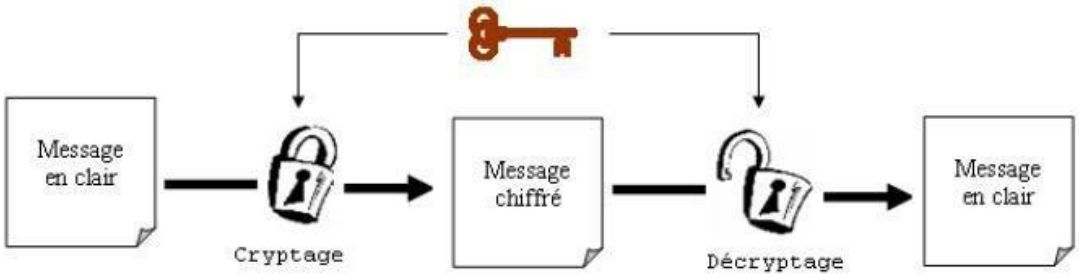
Lorsqu'un ordinateur se connecte à un site via Tor, le flux de données traverse aléatoirement un réseau de serveur. Ainsi on ne voit ni d'où la connexion provient, ni ce qu'elle contient car elle est chiffrée. Tor est un réseau composé de plusieurs couches de serveur appelés "nœuds du réseau".

Cependant ce navigateur rend la navigation plutôt lente dû au nombre conséquent de serveurs de connexion. De plus la sécurité de Tor n'est pas parfaite, en effet elle présente des failles connues. En effet entre le dernier noeud du serveur Tor et le destinataire des

données envoyées, le flux de donnée n'est pas chiffrée et donc les informations transmises sont vulnérables.

3 - La cryptographie

Depuis 1990, les logiciels de cryptographie ne sont plus réservés par l'armée et sont



Cryptographie à clef privée : la même clef est utilisée pour crypter et décrypter.

donc utilisables dans le domaine du public. On peut donc facilement chiffrer ses communications grâce à l'installation de logiciels.

3.1 - Les messageries instantanées et e-mails

Il existe de nombreux plugins qui utilisent le protocole de cryptographie OTR (Off-the-Record Messaging), il fourni de nombreux services de :

- ❖ Chiffrement, les messages ne sont pas lisible par autrui.
- ❖ Authentification, s'assurer que l'on parle bien à la bonne personne.

- ❖ Déni plausible, supprimer tous liens entre le message et expéditeur.
- ❖ Confidentialité persistante, protéger ses messages antérieurs.

Des logiciels comme Adium, Pidgin ou encore Miranda fournissent ce genre de services. Les e-mails sont très souvent surveillés, plus que les messageries instantanées. Il serait donc plus nécessaire de crypter les e-mails, dont le logiciel de cryptage le plus connu est PGP.

4 - Les précautions à prendre pour garder son anonymat sur le web

4.1 - Un système d'exploitation portable

Lors de l'utilisation d'un ordinateur public il y a un risque de laisser des données sur nous dessus. Il existe un moyen d'utiliser ces ordinateurs sans laisser de trace avec un système d'exploitation sur une clé USB. Ce système d'exploitation est sous Linux étant donné qu'il est moins gourmand en ressource et en place, sur ce système il faut installer les logiciels de cryptage précédent. Pour l'utiliser il suffit d'aller dans le BIOS et de changer les options de démarrage, il faut faire démarrer l'ordinateur sur la clé au lieu du disque dur contenant le système d'exploitation de l'ordinateur.

Cette méthode est, certes un peu compliqué pour les non-initiés mais elle est efficace pour ne pas laisser de trace lorsque que nous ne sommes pas sur notre ordinateur personnel.

4.2 - Les Logiciels Open Source

Les logiciels Open Source, ou logiciels libres, sont des logiciels dont le code est disponible et modifiable par tout le monde. Ces logiciels sont sans cesse analysés par une communauté de programmeurs, la communauté du libre. Cette façons de procéder permet à ces logiciels d'être dépouillés de toute fonctionnalités malveillantes qui permettrait la collecte des données personnelles. Les logiciels dits propriétaires ne donnent pas accès aux codes sources et il y donc un risque au niveau de la sécurité.

4.3 - Le Cloud

Le Cloud permet de stocker des données en ligne, le simple fait d'énoncer le principe du cloud montre à quelle point il présente un risque pour préserver son identité sur Internet. Lorsque nous mettons des données sur un cloud, les données ne nous appartiennent plus complètement et au moindre bug il est possible de perdre ou de se faire dérober ses données. C'est déjà arriver avec la fermeture de Megaupload qui avait un service de cloud, tous les utilisateurs avaient perdu leurs données. Ce n'est donc pas une manière sûre de mettre en sécurité ses données.

4.4 - Identités sur Internet

Une action basique mais indispensable est de changer son identité sur Internet, d'utiliser un pseudonyme qui n'a aucun rapport avec notre identité dans la vie réelle. Pour être encore plus anonyme il faut utiliser des adresses e-mail et des pseudonymes différents pour chaque service utilisé sur Internet.

L'identité sur Internet se forme lors de la création de compte, et nous créons souvent des comptes que l'on oublie par la suite. Mais ils sont toujours relié à notre adresse e-mail. le site web deseat.me est une solution à ce type de problème, il permet de mettre en évidence les relations entre votre adresse e-mail et tous vos comptes, et il vous permet également de supprimer ces liens.

4.5 - Les métadonnées

Les métadonnées sont des données de données, elles permettent de décrire une autre donnée, sa date de création, de dernière modification, la machine a l'origine de la création du fichier. Il est donc conseillé de supprimer ces métadonnées avant la publication d'un

fichier. Il existe différents logiciels pour effacer ces données, comme par exemple MAT (Metadata Anonymisation Toolkit) ou QuickFix.

CONCLUSION

En conclusion, après s'être intéressés à l'économie, la collecte, et le stockage des données, nous en sommes venus à la conclusion qu'Internet permettait d'accumuler une masse de données importante. Ces données sont récoltées par des entreprises comme Facebook ou encore Alphabet afin de notamment les revendre pour générer des publicités ciblées. Ensuite, nous avons parlé de l'aspect juridique des données sur Internet et des lois nous protégeant ou encore des lois permettant la collecte des données par les entreprises ou les Etats. Enfin, nous avons montré différentes méthodes permettant de se protéger de la collecte des données ou d'essayer d'être anonyme au maximum.

Au terme de ce travail, nous avons compris l'importance pour les grandes entreprises du web de collecter nos données. Avec cela nous comprenons que rester anonyme sur Internet ne sera pas chose aisée. Même si nos données sont assez bien sécurisées lors de leur transfert ou de leur stockage il est compréhensible et souhaitable de ne pas vouloir que l'on collecte nos données. Malheureusement, les lois concernant les données numériques ne peuvent pas se confronter à l'influence et à la volonté des grandes entreprises du web. Il en revient donc à l'utilisateur, s'il le souhaite, d'oeuvrer pour conserver au maximum son anonymat. Nous avons vu pour cela divers moyens plus ou moins accessible à tous. Cependant, même en réunissant l'ensemble de ces outils il semble impossible de préserver totalement son anonymat car nous ne prenons pas conscience de l'importance de conserver celui-ci dès notre première utilisation d'Internet. De plus, lors de notre première utilisation d'internet, nous n'avons pas non plus toutes les compétences nécessaires à la protection de nos données personnelles.

La liberté était une des caractéristiques fondamentales d'Internet à ses débuts. Cependant, comme nous l'avons vu avec la collecte de données, cette liberté s'est estompée. En effet, l'utilisateur n'a pas facilement accès au fait de choisir ou non de transmettre ses données. Néanmoins, Internet est un moyen de partage et cela devrait sembler logique que chacun des contenus que l'on peut y poster dessus soit collecté. On

peut remarquer un réel manque de prise de conscience de l'utilisateur sur les conséquences que peuvent avoir une publication. D'autre part, le manque d'anonymat total permet la traçabilité de nos actes, celle-ci pouvant limiter la pratique du cyber-harcèlement. Dans la réalité ce dernier est énormément pratiqué, les réseaux sociaux étant, malheureusement, un excellent vecteur d'application. L'anonymat connaît ici une certaine limite. Imaginons maintenant que nous soyons entièrement anonyme, le cyber-harcèlement n'en serait alors que décuplé. L'importance de le contrer est d'autant plus grande qu'il y a de plus en plus de personnes de tout âge qui utilise Internet. La présence d'un anonymat ou non sur Internet est alors discutable. Effectivement celui-ci est en accord avec les débuts d'Internet mais ses utilisateurs et son utilisation évoluent, peut-être que son changement à lui est aussi inévitable. C'est pourquoi la collecte et l'utilisation de nos données se doit d'être réglementée comme nous l'avons vu précédemment. En effet, des lois sont nécessaires car certains acteurs du web dépassent les limites de leur liberté en empiétant sur celle des autres.

Nous pouvons nous appuyer sur l'exemple de l'affaire du Cambridge Analytica. Effectivement, entre 2013 et 2016, l'entreprise collecte les données de 50 millions d'utilisateurs de Facebook sans leur accord, donc de façon totalement illégale. Cette collecte est illégale car seulement 270 000 personnes ont donné leur accord et qu'en réalité les données de beaucoup de leurs amis Facebook ont elles aussi été collectées.. L'affaire prend beaucoup d'ampleur dès lors que l'on apprend qu'une enquête américaine découvre que ces informations auraient été utilisées pour élaborer un logiciel permettant de prédire et d'influencer le vote des électeurs dans la direction du candidat républicain Donald Trump lors de la campagne présidentielle américaine de 2016.

Pour finir, Internet et les grandes entreprises qui lui sont liées sont en constante évolution. Avec eux, les lois s'appliquant sur Internet changent pour garantir un fonctionnement viable de celui-ci. En effet, Internet est un lieu où l'utilisateur se sert de sites mais en échange, le site en fait de même pour lui. L'interaction entre ces deux acteurs est primordiale et doit rester réglementée pour pouvoir qu'Internet puisse perdurer. Cette interaction démultipliée par le nombre d'utilisateurs et de sites permet à Internet d'être une base de données sans réelle fin.

Sources

I - L'économie des données

https://fr.scribd.com/document/386820061/DCN-Google-Data-Collection-Paper#from_embed
http://gs.statcounter.com
https://www.journaldugeek.com/2018/04/16/voila-genre-de-donnees-geants-web-collectent-sachiez/
https://www.zdnet.fr/actualites/android-de-google-aime-beaucoup-trop-nos-donnees-personnelles-39872643.htm
http://www.slate.fr/story/109791/test-google-donnees-confidentialite
http://www.internetactu.net/2012/02/27/quand-vous-ne-voyez-pas-le-service-cest-que-vous-etes-le-produit/
https://www.youtube.com/watch?v=rMV1WaWGb3I
https://www.youtube.com/watch?v=ms2lHt9b9CY
https://fr.wikipedia.org/wiki/Géants_du_Web

II - Collecte et stockage des données

https://www.la-croix.com/Sciences/Sciences-et-ethique/Que-deviennent-donnees-personnelles-lnternet-2016-12-06-1200808326
https://fr.wikipedia.org/wiki/Snapchat
https://www.blogdumoderateur.com/chiffres-reseaux-sociaux/
https://www.commentcamarche.com/contents/1041-cookies-internet
https://www.supergeek.fr/blog/cloud-serveur-ou-nas-le-bon-choix-pour-les-petites-entreprises/
https://www.lebigdata.fr/definition-big-data
https://www.google.fr/about/datacenters/
https://fr.wikipedia.org/wiki/Big_data
https://lokoyote.eu/le-vrai-danger-de-la-collecte-de-donnees/
https://www.insurancespeaker-wavestone.com/2014/02/donnees-personnelles-vie-privee-big-data-quels-sont-les-risques/
https://www.culture-informatique.net/cest-quoi-le-cloud/
https://www.sas.com/fr_fr/insights/big-data/what-is-big-data.htm
https://www.microsoft.com/fr-fr/security/resources/cookie-whatis.aspxl
https://www.culture-informatique.net/cest-quoi-les-cookies/

III - L'aspect juridique des données

https://fr.wikipedia.org/wiki/Politique_de_confidentialit%C3%A9
https://fr.wikipedia.org/wiki/Bouclier_de_protection_des_donn%C3%A9es_UE-%C3%89tats-Unis
https://www.cnil.fr/fr/textes-officiels-europeens-protection-donnees
https://fr.wikipedia.org/wiki/R%C3%A8glement_g%C3%A9n%C3%A9ral_sur_la_protection_des_donn%C3%A9es
https://www.cnil.fr/fr/reglement-europeen-protection-donnees/dataviz#
https://www.contrepoints.org/2018/06/29/319184-apres-rgpd-larticle-13-va-t-il-tuer-les-memes-internet
https://fr.wikipedia.org/wiki/USA_PATRIOT_Act
https://www.youtube.com/watch?v=LIiLoT4Po-c
https://www.isoc.fr/cgu-opinionway/
https://fr.wikipedia.org/wiki/Information_Awareness_Office
https://fr.wikipedia.org/wiki/National_Security_Agency

IV - L'anonymat sur internet

https://www.leblogduhacker.fr/tor-garantit-pas-lanonymat/
https://fr.wikipedia.org/wiki/Anonymat_sur_Internet
https://www.youtube.com/watch?v=LIiLoT4Po-c
https://www.nouvelobs.com/rue89/rue89-internet/20160823.RUE7578/vie-privee-le-guide-pour-rester-anonyme-sur-internet.html
https://www.paradisfiscaux20.com/comment-etre-anonyme-sur-internet.html
https://www.commentcamarche.com/contents/1040-securite-anonymat-sur-internet
https://desgeeksetdeslettres.com/difference-proxy-vpn-anonyme
https://fr.wikipedia.org/wiki/HyperText_Transfer_Protocol_Secure
https://fr.wikipedia.org/wiki/Off-the-Record_Messaging