ANNEXE 5 :
DOCUMENT DE SYNTHÈSE SUR L'INTERNET ÉLABORÉ PAR L'INRIA

DIRECTION DE LA RECHERCHE

Document de synthèse sur Internet

Editeur : Frédéric Desprez (Frederic.Desprez@inria.fr)

Contributeurs : Claude Castellucia (claude.castelluccia@inria.fr), Isabelle Chrisment (isabelle.chrisment@inria.fr), Walid Dabbous (walid.dabbous@inria.fr), Arnaud Legout (arnaud.legout@inria.fr), Jonathan Rouzaud-Cornabas (jonathan.rouzaud-cornabas@inria.fr), Luc Saccavini (luc.saccavini@inria.fr), Damien Saucez (Damien.Saucez@inria.fr)

Architecture et socle technologique d'Internet

L'Internet est bâti sur un modèle de réseau de réseaux où coopèrent des acteurs différents (fournisseurs d'infrastructures, de services, de contenus, utilisateurs, etc.) souvent concurrents. Le bon fonctionnement de cette tour de Babel planétaire est garanti par la mise en œuvre d'une base technologique commune dont l'objectif est de répondre à une fonction simple : la transmission de messages. Ces messages doivent pouvoir être transmis entre deux points quelconques du réseau Internet, leur taille pouvant varier de quelques octets à quelques milliards d'octets. Ces messages doivent être transmis de façon fiable, efficace, tout en partageant équitablement la capacité des liens (de quelques dizaines de Mbit/s à des centaines de Gbits/s). Ces messages doivent aussi pouvoir transiter par des supports de transmission variés comme les ondes hertziennes (wifi, téléphonie 2G, 3, 4G...) ou encore des supports physiques (fibre optique, paires torsadées cuivre) dont les caractéristiques techniques (fiabilité, débit, coûts) varient sur plusieurs ordres de grandeur. Une des bases d'Internet est d'avoir l'intelligence à la périphérie du réseau et un cœur de réseau simple.

Pour satisfaire ces contraintes en partie contradictoires, le modèle technologique repose sur deux principes : la communication de « bout en bout » pour la gestion des messages et le « meilleur effort » (*Best Effort*) pour leur acheminement. Le principe du « bout en bout » implique que les deux entités partenaires d'une communication (ex. smartphone / serveur) dialoguent directement pour garantir le bon échange des messages. Chaque message est découpé en paquets que les éléments intermédiaires des réseaux (les routeurs) sont chargés d'acheminer, sans intervenir dans le dialogue entre émetteur et récepteur des messages. Le protocole TCP (*Transmission Control Protocol*) est une implémentation de ce principe, tout objet qui veut se connecter à Internet doit être capable de « faire du TCP » (à de rares exceptions près). Une session TCP est principalement définie par 2 numéros ou ports : le port destination qui définit aussi la nature du service (web, messagerie, vidéo) que l'initiateur de la session veut utiliser et le port source qui permet d'identifier chacune des sessions TCP initiées par un objet donné. Les rôles de serveur et de client sont définis simplement : l'initiateur d'une session TCP est le client et le serveur est son correspondant qui accepte l'ouverture de la session.

Le principe du « meilleur effort » s'applique à la transmission des paquets constituant le message par l'infrastructure réseau. Ce principe implique que les équipements réseaux font ce qu'ils peuvent pour acheminer au mieux les paquets (qui peuvent ne pas arriver à destination, ou arriver dans le désordre). Cette non fiabilité, qui est compensée par TCP, permet de définir un protocole robuste et simple : le protocole IP. Son fonctionnement est basé sur l'algorithme du postier : chaque élément du réseau (ou routeur), choisit le routeur à qui il va transmettre le paquet IP en fonction de son adresse de destination. Cette simplicité du protocole IP permet aux opérateurs d'infrastructure réseau de coopérer sur la base de contrats de services relativement simples à instancier commercialement : l'échange de flux de paquets IP. Elle permet aussi à l'infrastructure d'évoluer dynamiquement (pertes de liens, évolutions topologiques) sans impact sur les communications (sessions TCP) en cours.

Le protocole IP est le véritable cœur technologique de l'Internet. On peut retenir que sa principale caractéristique est que chaque paquet possède une adresse source (qui caractérise l'émetteur) et une adresse destination (qui caractérise le récepteur à qui le paquet doit être remis). Ces adresses IP sont codées sur 32 bits ce qui définit un espace d'adressage de 2 milliards d'objets. Ces 32 bits sont présentés en 4 parties qui conduisent à un format du type « a.b.c.d » où a, b, c et d peuvent prendre des valeurs comprises entre 0 et 255

(exemple : 128.12.3.25). Compte tenu de la croissance d'Internet et en particulier de l'internet des objets (réseaux de capteurs, systèmes embarqués…), cet espace devient insuffisant et une nouvelle version de ce protocole (IPv6) est en cours de déploiement. En IPv6 les adresses sont codées sur 128 bits ce qui étend quasiment « à l'infini » cet espace d'adressage, puisqu'il autorise jusqu'à 7 x 10^{23} adresses IP par m^2 de la surface de la terre !

Du point de vue technologique, il n'y a pas d'autorité centrale, mais des acteurs qui interagissent économiquement, sur la base de protocoles éprouvés, définis dans des standards ouverts publiquement et collectivement élaborés et dont les spécifications sont librement disponibles.

En synthèse technique de cette description succincte des protocoles de base d'Internet, on peut retenir que chaque objet connecté à Internet doit posséder 1 adresse IP (au moins). Une session TCP/IP est ainsi identifiée de façon unique par 4 numéros : 2 adresses IP et 2 ports TCP (dits chacun source et destination) correspondants aux deux entités qui échangent des messages. L'identification de machines par des numéros IP n'est pas commode pour des utilisateurs, il faut donc pouvoir faire cette identification de façon plus ergonomique, c'est le rôle d'un autre composant clé d'Internet : le DNS.

Le DNS

Alors qu'au début de l'Internet le nombre de machines interconnectées était petit, il était facile pour un humain de retenir l'adresse IP de chaque machine et de s'y connecter directement en utilisant cette adresse. Cependant, la taille du réseau a rapidement augmenté et ce sont des milliers d'adresses que les utilisateurs devaient retenir. Pour pallier à cette complexité, un système d'indirection a été proposé. Le principe étant que chaque machine dispose d'un nom non ambigu et unique que les utilisateurs peuvent utiliser pour se connecter à la machine. Comme le protocole IP ne comprend pas les noms, mais uniquement les adresses, ces noms doivent être traduits en adresses. La première approche a été de construire un fichier commun, le fichier "hosts" qui est installé sur toutes les machines du réseau. A chaque changement, le fichier doit être corrigé et synchronisé entre toutes les machines du réseau. Le réseau prenant de l'ampleur, il devint rapidement ardu de garder le fichier "hosts" synchronisé sur toutes les machines et à cela s'ajoutait la complexité de nommer les machines de manière unique. Pour cette raison les concepts de noms de domaines et de DNS (*Domain Name System*) ont été proposés et déployés dès les années quatre-vingt.

Le DNS est donc un annuaire réparti qui permet de passer d'un espace de nommage humainement compréhensible pour désigner les machines connectées à Internet, aux informations techniques comme leurs adresses IP. Par exemple le site www.assemblee-nationale.fr a comme adresse IP 89.185.59.149. Quand un internaute écrit le texte suivant http://www.assemblee-nationale.fr/ dans son navigateur cette information provoque l'ouverture d'une session TCP sur le port destination 80 (service=web) vers la machine d'adresse IP destination 89.185.59.149.

L'espace de nommage du DNS est hiérarchique. Un arbre de nommage est ainsi créé par délégation de sous espaces en cascade. Chaque nœud représente une entité à nommer (ex., un domaine administratif, un serveur). Afin d'éviter toute ambiguïté, tous les fils d'un nœud reçoivent un nom différent et le nom global pour un nœud correspond à la concaténation du nom de chaque nœud suivant la hiérarchie entre lui et le sommet de l'arbre. Le premier niveau est constitué de deux ensembles : les domaines nationaux ou ccTLD (*country code Top Level Domain*) et les domaines génériques ou gTLD (*general Top Level Domain*). Les ccTLD réfèrent des pays (.fr pour la France, .es pour l'Espagne, etc) les gTLD des domaines généraux (.com pour commercial, .edu pour les sites académiques, .org pour les organisations). Le DNS comprend actuellement 320 TLD, 265 millions de domaines dont 105 dans le gTLD .com. L'ICANN (*Internet Corporation for Assigned Names and Numbers*), est une autorité de régulation qui a été mise en place pour gérer le sommet de la hiérarchie du système de nommage dans l'Internet. Le rôle de l'ICANN est d'une part de déterminer les règles de nommage et de s'assurer du bon fonctionnement des serveurs qui composent la racine du DNS. Le rôle de l'ICANN est aussi déterminant dans la création de nouveaux TLD ou même des alphabets pouvant être utilisés pour construire des noms.

La gestion des (13 systèmes) racines est l'une des 3 missions principales de l'ICANN. La gouvernance de l'ICANN est régie par un long document, initié en 1999 et dont la dernière mouture date de 2002[1]. Ce texte fait de nombreuses références à la loi californienne « CNPBCL » (*California Nonprofit Public Benefit Corporation Law*). Là où il est évident que pour maintenir de l'ordre dans un système de nommage de l'ampleur du DNS il est nécessaire de se référer à une autorité centrale, il y a une certaine controverse autour de l'ICANN et de ses liens plus ou moins proches avec le gouvernement américain du fait que l'ICANN est intrinsèquement lié au département du commerce des Etats-Unis. Par exemple, d'aucun pourrait voir une sur-représentativité d'organisations basées aux Etats-Unis dans l'attribution des serveurs racines du DNS[2]. Ceci dit, cela vient principalement de raisons historiques. Toutefois, comme l'attribution se fait de manière assez statique suivant des schémas administratifs complexes, il est assez ardu de pouvoir déployer une racine DNS officielle, ce qui pourrait mettre en péril la neutralité du système de nommage qui est le pilier de l'Internet tel que nous le connaissons aujourd'hui.

La possibilité de créer de nouveaux gTLD a été décidée en 2008 et a été lancée en 2012. Le nombre de gTLD va très probablement augmenter dans un avenir proche, avec l'arrivée de gTLD comme .paris, .banque, .sncf par exemple. Il y a eu 1800 demandes de nouveaux gTLD déposées auprès de l'ICANN, mais compte tenu des critères de sélection et du coût élevé (180k$ pour l'examen complet d'un dossier, puis 25k$/an), le nombre final de nouveaux gTLD est pour l'instant estimé à 1200.

Chaque gérant de TLD va déléguer la gestion d'une sous-zone à une entité qui en fait la demande légitime. Ainsi l'AFNIC qui gère le .fr va déléguer au prestataire qui gère le site web de l'Assemblée Nationale le domaine 'assemblee-nationale.fr', ce dernier pourra ensuite organiser l'espace de nommage 'assemblee-nationale.fr' à sa guise. Il peut nommer directement des machines, créer d'autres sous-domaines, etc.

La structure hiérarchique du nommage DNS crée une pseudo sémantique (.fr = France, .com = sté commerciale) qui peut être trompeuse. En effet, une machine sous .fr peut être située n'importe où sur la planète, et sur le territoire français il peut y avoir des machines de pratiquement n'importe quel TLD.

Une partie de la sécurité d'Internet repose sur le DNS, en particulier l'authentification des serveurs applicatifs. Comment être sûr que l'adresse IP 89.185.59.149 est bien celle du serveur www.assemblee-nationale.fr ? Un protocole complémentaire (DNSSEC) permet de garantir par signature numérique que les informations DNS sont correctes. Le déploiement de DNSSEC commencé en 2009 par la signature de la racine du DNS se poursuit doucement. Par exemple sur .fr 20 000 domaines étaient signés fin 2012 (sur 2,5 millions).

Quand un internaute connecte son terminal (*smartphone*, PC, tablette) à Internet, via un point d'accès privé (ADSL domestique), d'entreprise ou public, cette connexion est fonctionnelle quand le point d'accès a affecté à son terminal une adresse IP, l'adresse IP du routeur le plus proche et l'adresse IP du serveur DNS à utiliser.

Racines ouvertes

En alternative au système de racines officiel, la mouvance des racines ouvertes (*open roots* en anglais) est de plus en plus marquée. L'idée des racines ouvertes est de ne plus lier le devenir des racines DNS à une seule autorité, mais de déployer les racines de manière coopérative à la place. La majorité des racines ouvertes reprennent le schéma de l'ICANN avec la délégation des racines à des entités bien déterminées. Cependant d'autres approches se veulent complètement coopératives et reprennent le schéma bien connu des réseaux pair-à-pair où chaque utilisateur du système devient aussi contributeur. L'utilisation des racines ouvertes permet de casser l'aspect monopolistique qui règne actuellement dans la gestion du sommet de la hiérarchie DNS et offre de ce fait plus de flexibilité et de liberté dans la gestion des noms dans l'Internet, par exemple en autorisant l'utilisation des alphabets et langues locales. Les approches fortement décentralisées permettent aussi de lutter efficacement contre la censure et la répression et donc d'assurer un accès libre à l'Internet et

[1] www.icann.org/en/about/governance/bylaws

[2] Voir http://fr.wikipedia.org/wiki/Serveur_racine_du_DNS (tableau repris en **Annexe 1**)

d'empêcher les formes de répressions numériques qui sont de plus en plus nombreuses[3,4]. Cette liberté peut cependant compromettre la cohérence de l'ensemble car elle supprime l'unicité globale des noms dans l'Internet. Les racines ouvertes posent aussi des problèmes juridiques avec un contrôle difficile de l'attribution des noms aux différentes entités et donc le risque pour les marques de perdre le contrôle sur leur nom dans l'Internet.

Sur ce sujet de racines ouvertes, Julien Naillet, en charge de la communication de l'Afnic (Association française pour le nommage Internet en coopération) déclare : « *Nous accueillons avec intérêt toute initiative en faveur de l'innovation et de la concurrence. Néanmoins, il nous semble essentiel de garantir aux utilisateurs l'unicité des noms de domaine en usage. La multiplication des racines, bien que pouvant offrir de nouvelles fonctionnalités au cas par cas, est donc une voie sur laquelle nous ne souhaitons pas nous engager.* » Il est important de garder une autorité sur le sujet mais qu'elle soit sous contrôle indépendant et ouvert. Plus précisément, la participation dans l'autorité devrait être libre, gratuite, non-gouvernementale et non-commerciale afin de laisser la tribune à tout acteur du monde numérique, tant individuel qu'organisationnel. La raison pour laquelle il est important de maintenir une autorité est lié au besoin de conserver un système assurant l'unicité des noms à l'échelle globale, et sans ambiguïté, tout comme recommandé par l'Afnic. Un autre point qui nous semble important est de rendre complètement libre et décentralisé le déploiement de serveurs racines alors qu'aujourd'hui seules les entités approuvées par l'ICANN sont autorisées à déployer des serveurs racines officiels. L'aspect décentralisé étant à mettre en avant afin de réduire le risque de censure au maximum.

BGP: le protocole qui assure le fonctionnement global d'Internet

Comme cela a été expliqué précédemment, Internet est un réseau de réseaux. Chacun de ces réseaux est appelé *Autonomous System* (AS[5]), il en existe 45 000 (début 2014). Le protocole BGP (*Border Gateway Protocol*) permet à ces différents AS de communiquer, c'est-à-dire d'échanger du trafic de paquets IP. Pour cela, chaque AS va dire à ses voisins quels préfixes (ensembles d'adresses IP) il sait router. Soit parce que les machines correspondantes à ces préfixes sont dans son AS, soit parce qu'il a un voisin qui sait router ces préfixes. Ces informations de routage vont permettre aux AS de cœur de connaître toutes les routes de l'Internet (500000 actuellement) et d'acheminer n'importe quel paquet vers sa destination. Par exemple c'est l'AS CLARANET-AS N° 8426 de la sté ClaraNET Ltd qui possède le préfixe incluant l'adresse IP 89.185.59.149 et qui annonce à ses voisins qu'il sait router les paquets IP correspondants. BGP est, avec TCP/IP et DNS, une des briques essentielles d'Internet car c'est ce protocole qui permet aux différents AS qui forment Internet de communiquer. Tout comme ces deux derniers, il est basé sur des protocoles ouverts. De plus, BGP est construit sur une confiance absolue entre les différents AS. Comme un AS peut annoncer qu'il est capable de router des préfixes correspondant à des terminaux qui ne sont pas dans son réseau il a la possibilité de détourner par son système des données à destination d'un autre. Ces attaques ont été vues plusieurs fois ces dernières années[6]. Elles vont permettre de facilement espionner le trafic à destination ou en provenance d'une partie du réseau sans avoir à installer du matériel ou aller modifier les systèmes qui sont attaqués.

Pour transférer des données d'un AS à un autre, des opérateurs dits de transit IP (ou de cœur) vont fournir les liaisons qui permettent ce transfert. Certains de ces transits sont payants en fonction de la capacité du lien et du volume données qui vont y transiter. Un acteur majeur est Open Transit qui appartient à Orange mais aussi Cogent ou Level 3 qui sont des sociétés étrangères. Ces acteurs ne sont pas visibles par les utilisateurs mais disposent d'un grand pouvoir puisqu'ils sont en charge d'interconnecter les différents systèmes qui fondent Internet mais ils pourraient aussi surveiller le trafic de leurs clients passant dans leurs liaisons. Pour autant, afin de diminuer les coûts, il existe des accords dit de peering entre les différents acteurs d'Internet. Un accord de peering permet d'échanger gratuitement du trafic entre 2 systèmes (AS). De tels accords existent entre la plupart des fournisseurs de contenus et d'accès sur Internet. Cela se base sur un trafic symétrique (entrant et

[3] http://www.lefigaro.fr/secteur/high-tech/2011/01/26/01007-20110126ARTFIG00639-face-a-la-revolte-l-egypte-muscle-sa-censure-du-web.php
[4] http://www.lefigaro.fr/flash-actu/2011/10/29/97001-20111029FILWWW00423-la-syrie-a-censure-internet.php
[5] Internet est modélisé et organisé par un ensemble d'AS communicants entre eux.

[6] http://www.bortzmeyer.org/bgp-shunt.html

sortant). Mais de part les nouveaux modèles de consommation sur Internet, il y a eu une perte de symétrie. Par exemple, les clients de Free vont regarder plus de vidéos sur Youtube qu'ils vont en envoyer. L'excédent doit passer par des transits IP payants. Il est nécessaire de conserver ces accords de peering afin de conserver une bonne qualité du réseau. Des combats commerciaux existent pour savoir qui doit payer pour augmenter la capacité des liaisons de transit IP ainsi que ceux de peering. Un exemple concret est le ralentissement de Youtube pour les clients de Free de fait d'une bataille commerciale entre les 2 sociétés. Une des caractéristiques originales du protocole BGP est donc de prendre en compte (à travers son paramétrage) aussi bien des caractéristiques technologiques (optimisation du routage des paquets IP) que des éléments commerciaux (implémentation de l'accord de peering entre deux opérateurs d'AS).

Elaboration des standards

L'élaboration des standards de l'Internet est principalement portée par l'IETF (*Internet Engineering Task Force*). Il s'agit d'une organisation ouverte où les participations sont individuelles. Toute personne peut participer (il suffit que son employeur le paye pour cela), et son poids dans l'élaboration des standards sera directement lié à sa capacité à agir et à la qualité de ses propositions de protocoles. L'objectif premier de l'IETF est de produire des spécifications de protocoles appelées RFC les plus efficaces, les plus simples et les plus lisibles possibles afin de faciliter leur implémentation dans les systèmes. Pour cela, quelques règles simples sont mises en œuvre : les documents sont tous publics, et discutés publiquement. De plus un projet de spécification ne peut passer la première étape de validation que si deux implémentations en ont été faites de façon indépendante et coopèrent comme prévu. Ces points méthodologiques, garantissent d'être à l'état de l'art, l'efficacité des protocoles et la lisibilité des documents de référence.

A titre indicatif, l'IETF produit environ 250 RFC par an dont la moitié sont des propositions de standards. Au bout d'un processus qui peut durer plusieurs années, 2 à 3 de ces propositions deviennent chaque année des standards aboutis ayant un caractère obligatoire. L'implication de la France est notable dans ce processus et représente 4% de cette production à travers des auteurs issus du monde académique et industriels (France-Telecom, Orange, INRIA, AFNIC, Alcatel Lucent, Renater, SFR, Bouygues Telecom, Institut Telecom, ...). D'autres organismes participent à l'élaboration des standards de l'Internet, comme le W3C (web), l'IEEE et l'ETSI (liens physiques, hertziens, filaires), l'OGF (services répartis) et d'autres encore sur des domaines très spécifiques.

Indépendance numérique de la France

A côté de la gouvernance ISOC, la véritable gouvernance revient de plus en plus aux acteurs économiques. On observe un glissement historique depuis les équipementiers, telco jusqu'aux producteurs de contenus actuellement.

L'économie numérique est reconnue aujourd'hui comme une des premières économies mondiales, mais l'importance de l'indépendance numérique des états est largement sous-estimée ce qui peut conduire à moyen terme à une profonde modification des équilibres géopolitiques.

La France sur les 40 dernières années a fait le choix d'un protectionnisme d'anciens modèles économiques au détriment de nouveaux modèles innovants. Ce choix a conduit à une forte dépendance numérique envers notamment les États-Unis. Deux cas emblématiques illustrent ce choix : l'adoption d'Internet et le téléchargement pair-à-pair. L'Internet a longtemps été opposé au minitel en France, les responsables politiques et dirigeants de grands groupes considérant Internet comme étant un produit américain devant être freiné face au minitel. En effet, le minitel avait un modèle économique stable et bien maîtrisé : l'accès payant a tous les services avec une facturation sur la ligne téléphonique rattachée. Cette lutte contre Internet fut paradoxale puisque des chercheurs français jouèrent un rôle de premier plan dans sa conception, notamment Louis Pouzin et Christian Huitema, chercheurs visionnaires qui comprirent très tôt le potentiel énorme d'Internet, mais qui ne furent pas suivis par les dirigeants de l'époque.

Le deuxième cas emblématique est celui du téléchargement en pair-à-pair (utilisant majoritairement BitTorrent) qui fut stigmatisé. Cependant, ce soudain succès du téléchargement pair-à-pair montrait que les nouveaux modes de communication et les attentes du grand public ne correspondaient plus aux modes historiques de diffusion des contenus audiovisuels. Dans ce cas également, la France a été précurseur avec

Azureus/Vuze (le client BitTorrent le plus populaire créé par un français Olivier Chalouhi qui est parti en Californie pour développer sa société), Wizzgo (le premier magnétoscope numérique déporté condamné par décision de justice à arrêter ce service pourtant précurseur des actuels services de vidéos à la demande), et les succès relatifs que sont Dailymotion et Deezer (qui sont pourtant parfaitement compétitifs avec les leaders mondiaux Youtube et Spotify, mais qui manquent de contenus).

Bien que le contexte soit très différent dans ces deux cas, la même erreur fut reproduite : vouloir conserver un ancien modèle économique bien maîtrisé et refuser un nouveau modèle économique en rupture. Cette stratégie protectionniste n'est plus adaptée dans un contexte mondialisé tel que le permet Internet. En effet, n'importe qui peut accéder de manière totalement transparente à n'importe quel service opéré depuis n'importe quel pays. Lorsqu'un modèle économique alternatif apparait, la seule option valable est de le soutenir, puisqu'en cas d'absence de soutien, ce modèle sera développé à l'étranger sans aucun retour financier pour la France ni aucune possibilité de régulation.

Les conséquences d'une dépendance numérique sont nombreuses :

- perte d'emplois en France au profit de pays ayant développé les nouveaux modèles économiques ;
- difficulté à légiférer et réguler ces nouveaux modèles qui reposent sur de nouveaux services accessibles par Internet ;
- risques d'atteinte à la vie privée puisque les données ne sont pas hébergées en France est qu'il est très difficile d'influer sur un choix stratégique d'une société basée à l'étranger ;
- risque de pénurie numérique lorsqu'une seule société basée à l'étranger a le monopole de la fourniture d'un service à forte valeur ajoutée (Google, Facebook, Twitter, iTunes, Dropbox, Amazon, etc.)

L'identification des enjeux liés à l'indépendance numérique de la France et les choix stratégiques afférents relève de compétences scientifiques, technologiques et politiques de haut niveau. Aucun acteur isolé ne peut prendre de décisions éclairées, il est par conséquent fondamental de renforcer les discussions et collaboration entre les chercheurs, les entrepreneurs et les décideurs politiques.

Sécurité et chiffrement

Au niveau d'Internet, la plupart des services utilisent des mots de passe et le chiffrage des communications. Pour les mots de passe, il arrive trop souvent que les bases de données des sites web ne soit pas chiffrées (ou faiblement) et que des grands ensembles de mots de passe se retrouvent dans le domaine public. Comme il n'y a pas de loi forçant les entreprises et les sites web à divulguer ces attaques et mises à disposition de données, il n'est pas forcément simple de savoir si un mot de passe a été compromis. De plus, une grande partie du chiffrage des communications repose sur le protocole SSL (https par ex). Hors, tout comme les noms de domaines, les certificats SSL utilisés pour l'authentification du serveur reposent sur une structure stricte et très contrôlée par les USA. La corruption du système de certificat aurait des impacts encore plus grands en termes de sécurité que celles des serveurs DNS. En effet, il serait possible de se faire passer pour un site commercial ou gouvernemental et de lire les informations envoyées entre les utilisateurs et le site.

Il est possible de chiffrer les données avant de les envoyer sur un canal qui est sécurisé ou pas. On peut prendre l'exemple du site mega.co.nz (ou de cryptocat) qui place le chiffrement du côté de l'utilisateur. Par conséquent, même si les données sont lues lorsqu'elles circulent ou lorsqu'elles sont stockées, elles sont chiffrées. Cela permet de garantir une meilleure sécurité à un hébergeur car il ne peut pas savoir ce qu'il stocke et donc en être tenu responsable. Par contre, dans le cas du Cloud, on se retrouve avec des données chiffrées dans le Cloud et il faut les déchiffrer pour pouvoir les traiter. Ce problème au niveau du Cloud peut être réglé par la cryptographie homomorphique. En effet, en utilisant cette technologie, il est possible de faire des traitements directement sur les données chiffrées et donc faire ces traitements sur des plateformes dont on a moins confiance. Pour le moment, les systèmes de cryptographie homomorphique ont souvent un coût en termes de performance trop important pour être utilisables et/ou ne permettent que l'une des 2 opérations élémentaires (addition, multiplication).

Enfin, il est maintenant reconnu qu'aussi bien des cœurs de réseaux que des câbles sous-marins, bien qu'utilisant des fibres optiques comme support de transmission ont été espionnés[7]. Cela pose le problème de la sécurité physique de ces infrastructures qui sont pour le moment considérées comme étant de confiance. Pour éviter ces problèmes de communication et détecter les tentatives d'espionnage, certaines banques suisses ont installé un réseau basé sur le cryptage quantique qui rend les méthodes actuelles d'espionnage beaucoup plus complexes. Pour parer à l'espionnage des fibres optiques par écoute directe de la fibre optique, certains équipementiers de matériels optiques (dont Alcatel) proposent des solutions de chiffrement au niveau optique jusqu'à des débits de 10Gb/s par lambdas actuellement.

DataVeillance (Surveillance par les données)

Les révélations récentes sur les programmes d'espionnage de la NSA ont suscité l'émoi et l'inquiétude en France et un peu partout dans le monde. Ces révélations ont eu le mérite de susciter des discussions/débats et des prises de conscience des hommes politiques et des citoyens. Bien que la surveillance soit nécessaire dans certains cas, si elle est bien encadrée, la surveillance de masse est dangereuse pour la démocratie et peu efficace[8]. Un autre mode de surveillance qui fait moins parler de lui et qui a été un peu éclipsé par l'impact médiatique de l'affaire NSA, mais qui est au moins autant, voire plus, inquiétant (par son ampleur, par ses acteurs, par ses motivations, par sa furtivité, par son manque de transparence) est la *surveillance de masse sur Internet par des entités privées* (publicitaires, agrégateurs/courtiers de données, réseaux sociaux, fournisseurs de services etc.) qui collectent de plus en plus de données ou métadonnées sur les Internautes[9]. On peut distinguer deux types de collectes : les collectes involontaires et les collectes volontaires (entre autre par les réseaux sociaux).

Collectes passives et involontaires (online tracking). Il existe des centaines d'entreprises qui « tracent » les utilisateurs, souvent à leur insu, lorsqu'ils utilisent l'Internet. L'objectif de ces entreprises est d'identifier les sites visités par les utilisateurs, pour identifier leurs centres d'intérêts et construire des profils. Ces profils sont souvent complétés avec des informations recueillies sur diverses bases de données publiques ou privées. Ils incluent typiquement l'âge, la race, le sexe, le nombre d'enfants, le niveau d'éducation, les achats récents, etc. Ces courtiers en données (*data brokers*), tel qu'Acxiom.com, revendent souvent ces données à des banques, publicitaires, agences de crédits, assurances, partis politiques etc. Le marché est tellement juteux que certaines entreprises proposent même aux utilisateurs de racheter leurs données pour quelques dollars[10] ! Une étude récente effectuée dans le cadre du projet Cnil-Inria Mobilitics a montré que les téléphones mobiles sont eux-aussi largement ciblés et qu'une grande majorité des applications mobiles exfiltre des données privées vers des entités tierces. La plupart des applications mobiles étant « gratuites », les données personnelles deviennent une monnaie virtuelle. Ce phénomène est d'autant plus inquiétant que les mobiles possèdent et génèrent beaucoup d'informations personnelles (localisations, listes de contacts, ...). De plus il est très difficile, voire impossible, d'échapper à cette traque[11].

Collectes « volontaires ». Les données que les utilisateurs publient sur les divers réseaux sociaux (Facebook, Twitter,...) sont, bien entendu, aussi utilisées pour enrichir les profils construits par les divers « *Data Brokers* ». Le développement des gadgets connectés (Fitbit, Withings,...) et du mouvement « *Quantified Self* » qui permet à chacun de « se mesurer » (fréquence cardiaque, calories, sommeil, etc.) pour mieux se connaître est un progrès incontestable et aura probablement un impact positif sur la santé. Cependant les données collectées,

---

[7] **Les câbles sous-marins, clé de voûte de la cybersurveillance**
http://www.lemonde.fr/technologies/article/2013/08/23/les-cables-sous-marins-cle-de-voute-de-la-cybersurveillance_3465101_651865.html

[8] **Surveillance is necessary for security, but not mass-surveillance**
http://www.theguardian.com/commentisfree/2014/feb/11/surveillance-myths-nsa-reform-freedom-act?CMP=twt_gu

[9] **How your Data are being deeply mined** http://www.nybooks.com/articles/archives/2014/jan/09/how-your-data-are-being-deeply-mined/?pagination=false

[10] http://www.technologyreview.com/news/524621/sell-your-personal-data-for-8-a-month/ et http://stream.wsj.com/story/markets/SS-2-5/SS-2-453476/

[11] http://juliaangwin.com/privacy-tools-opting-out-from-data-brokers/

éminemment sensibles car liées à la santé, peuvent aussi être aussi très convoitées par les « *data brokers* ». En effet, imaginer la source d'information que ces données peuvent être pour vos assureurs ou banques (par exemple lors d'une demande de prêt)!

Les dangers. Il existe au moins trois conséquences de ces pratiques qui méritent d'être discutées :

Discrimination par les données : Ces profils peuvent être utilisés pour catégoriser les utilisateurs selon différents critères (par exemple « acheteurs impulsifs », « acheteurs influençables », etc.) afin de fournir des traitements différents. Cette catégorisation peut aboutir à des discriminations inacceptables. Par ailleurs, ces profils étant générés automatiquement, par des algorithmes, ils peuvent parfois être erronés et produire des aberrations.

Manipulation par les données : Les profils peuvent aussi être utilisés pour manipuler les gens en leur présentant les informations de façon à influencer leurs décisions.

Surveillance par les données : Finalement, comme les récentes révélations sur la NSA l'ont montré, ces données et profils peuvent être revendus à divers gouvernements à des fins de surveillance de masse.

Le développement de la collecte des données personnelles est préoccupant car il est aujourd'hui impossible d'y échapper. Cette collecte est omniprésente, à la fois sur l'Internet mais aussi dans le monde physique (vous êtes filmés en permanence dans les magasins, les rues. Vous laissez des traces lorsque vous payez avec votre carte bancaire, prenez les transports en commun, empruntez les autoroutes, utilisez votre téléphone mobile, etc.). Il y a ainsi une inversion entre ce qui est perçu par l'Internaute qui se sent consommateur de services (gratuits) alors qu'il est en fait producteur de données à forte valeur (son comportement, ses connaissances, ses opinions, ...). La technologie se développe beaucoup plus rapidement que la législation sur la protection des données personnelles et des citoyens. Bien que la technologie puisse apporter des solutions, elle ne peut malheureusement pas résoudre tous les problèmes. Il est important d'avoir plus de transparence et de contrôle sur la façon dont nos données sont collectées et utilisées. Il est aussi important de lancer un débat sur la question du « *Big Data* » comme vient de le faire la Maison Blanche[12].

Cloud, Internet of Things

Le Cloud Computing pose des problèmes encore plus complexes d'un point de vue de la sécurité. En effet, le matériel est maintenant virtualisé et ce matériel est disposé dans un lieu distant auquel le client n'a pas accès. Cela amène le concept de confiance sans contact. Il faut faire confiance envers la société qui fournit le service de ne pas espionner les données qui transitent et d'avoir une sécurité suffisante pour garantir sa sécurité et la nôtre. Les méthodes d'évaluation et de certification telles que ebios sont caduques dans ce cas. En effet, il faudrait évaluer la sécurité du système d'information de l'entreprise mais aussi celui de l'ensemble des systèmes d'information de ses fournisseurs (Cloud, etc.). En effet, en sécurité, le niveau de sécurité global est toujours celui du système le plus faible.

Pour autant, est-ce une bonne idée d'avoir un Internet/Cloud purement européen et/ou français comme par exemple l'Internet iranien ou Chinois qui ne sont pas « parfaits » mais permettent via un contrôle très fort de la totalité des infrastructures, services et points d'accès à Internet de limiter la propagation des données. Le Cloud souverain (Cloud spécifique à un pays) ne permet pas de s'affranchir des problèmes de sécurité inhérents aux piratages (par exemple, voir le piratage de Belgium Telecom par la NSA pour récupérer les méta-données sur tous les appels en Belgique). Il permet par contre de s'assurer que la loi nationale/régionale s'applique sur les données et les processus de ce Cloud. Pour autant, cette approche n'a pas tellement de sens car ce n'est pas uniquement le lieu géographique qui compte mais la totalité de la pile matérielle et logicielle qui doivent être souverains dans ce cas. Il faut donc pouvoir redévelopper et produire l'intégralité de la plate-forme Cloud depuis les puces électroniques jusqu'au service SaaS. Cela n'est clairement pas réaliste car équivalent au principe d'interdire tel ou tel fournisseur. Huawei (constructeur chinois d'équipements réseaux) a été interdit car les américains n'ont pas confiance dans le logiciel et les composants qui se trouvent dans ses routeurs. En

[12] Bring on the Big Data Debate, https://cdt.org/blogs/0202bring-big-data-debate

effet, ils pourraient contenir des logiciels permettant la surveillance par la Chine à l'insu des gens qui installent ce matériel. Pour autant, il faut se rappeler qu'en 2008, une agence fédérale américaine s'était rendu compte que certains des routeurs Cisco qu'ils utilisaient étaient des contrefaçons chinoises[13]. Sachant qu'il est possible d'introduire des mécanismes de surveillance cachés (chevaux de troie par exemple) aussi bien dans le logiciel et le matériel, cela aurait pu avoir l'impact que souhaitent éviter les américains en bannissant Huawei. Cela justifie un contrôle de bout en bout du développement à la production. Une entreprise américaine produisant en Chine n'est pas plus sûre qu'une entreprise Chinoise produisant aux USA.

Bien sur, il est possible de réguler les flux d'informations de l'ensemble d'un pays. Quelques sociétés disposent et vendent des équipements permettant à des gouvernements ou des entreprises de surveiller et interdire des flux réseau ciblés. Au niveau d'un pays, pour repérer un flux interdit pour par exemple essayer de contrer la fuite d'information industrielle suite à un pirate, il faudrait utiliser des méthodes fortement intrusives en marquant l'ensemble des documents avec des *watermark* ou DRM et inspecter l'ensemble des paquets/flux sortant de France pour voir si on repère ces documents. Un pays avec un budget conséquent comme la Chine ne réussit pas à imposer un tel contrôle qui a plus de chance de se retourner contre le citoyen que d'améliorer la sécurité des entreprises.

Après si le but est de contrôler les données au sein d'un Cloud, il est possible de pouvoir exprimer ce genre de problème comme une propriété de sécurité/privacy et ensuite d'avoir un (ou un ensemble de) mécanisme(s) qui va (vont) s'assurer que cette propriété est bien respectée. Ces mécanismes de contrôle peuvent être augmentés avec des mécanismes d'assurance qui permettent de fournir la preuve à l'utilisateur que ses demandes de sécurité ont été respectées (une sorte de log).

Pour autant, il faut se méfier des certifications de sécurité. Mettre en place des audits spécifiques au Cloud via une certification donnée par un tiers peut amener de la confiance. Pour autant, les certifications reconnues et utilisées par le secteur de la sécurité informatique tel qu'eBios, CC, ISO 27001, etc. ont montré la limite de ces approches et la décorrélation entre une sécurité effective et une sécurité sur le papier. De plus, il faut que les résultats d'un audit soient appliqués. Pouvoir fournir un Cloud sécurisé passe par un processus de R&D complexe visant à analyser, développer, évaluer chacun des composants (matériels et logiciels) du Cloud. Le but étant de construire progressivement une base de confiance suffisamment large pour fournir ce Cloud sécurisé.

Pour finir, les objets connectés sont de plus en plus nombreux. Mais comme les téléphones portables il y a quelques années, ils ne sont jamais mis à jour ou ont une sécurité approximative. On peut ainsi trouver des dizaines (centaines) de milliers de dispositifs connectés à Internet avec le login/mot de passe par défaut qui est de plus disponible dans la documentation. Enfin, les mécanismes de mise à jour automatique n'existent pas toujours. Pour le moment, les problèmes liés à ces objets connectés sont minimes de part leur faible déploiement et la faible sensibilité des données qu'ils traitent. Pour autant, il devient de plus en plus critique de s'intéresser à leur sécurité de part leur omniprésence de plus en plus forte dans des dispositifs critiques de la vie de tous les jours (domotique, *wearable computing*, etc.).

[13] http://it.slashdot.org/story/08/05/09/164201/fbi-says-military-had-counterfeit-cisco-routers

DIRECTION DE LA RECHERCHE

Annexe 1 : Les treize serveurs* racines du DNS

Lettre	adresse IPv4	adresse IPv6	Autonomous System	Ancien nom	Société	Localisation	Sites (global/local)	Logiciel
A	198.41.0.4	2001:503:ba3e::2:30	AS19836	ns.internic.net	VeriSign, USA	trafic distribué par anycast	6 (6/0)	BIND
B	192.228.79.201	2001:478:65::53 (pas encore dans la zone)	AS4	ns1.isi.edu	USC-ISI (en), USA	Marina Del Rey, Californie, États-Unis	1 (1/0)	BIND
C	192.33.4.12	2001:500:2::c (pas encore dans la zone)	AS2149	c.psi.net	Cogent Communications, USA	trafic distribué par anycast	6 (6/0)	BIND
D	199.7.91.13	2001:500:2d::d	AS27	terp.umd.edu	Université du Maryland, USA	College Park, Maryland, États-Unis	1 (1/0)	BIND
E	192.203.230.10		AS297	ns.nasa.gov	NASA, USA	Mountain View, Californie, États-Unis	1 (1/0)	BIND
F	192.5.5.241	2001:500:2f::f	AS3557	ns.isc.org	Internet Systems Consortium, USA	trafic distribué par anycast	49 (2/47)	BIND, ?
« G »	192.112.36.4		AS5927	ns.nic.ddn.mil	Defense Information Systems Agency (en), USA	trafic distribué par anycast	6 (6/0)	BIND
H	128.63.2.53	2001:500:1::803f:235	AS13	aos.arl.army.mil	United States Army Research Laboratory (en), USA	Aberdeen, Maryland, États-Unis	1 (1/0)	NSD
I	192.36.148.17	2001:7fe::53	AS29216	nic.nordu.net	Autonomica, Suède	trafic distribué par anycast	36	BIND
J	192.58.128.30	2001:503:c27::2:30	AS26415		VeriSign, USA	trafic distribué par anycast	70 (63/7)	BIND
K	193.0.14.129	2001:7fd::1	AS25152		RIPE NCC, Hollande	trafic distribué par anycast	18 (5/13)	NSD
L	199.7.83.42	2001:500:3::42	AS20144		ICANN, USA	trafic distribué par anycast	38 (37/1)	NSD
M	202.12.27.33	2001:dc3::35	AS7500		WIDE Project, Japon	trafic distribué par anycast	6 (5/1)	BIND

(*) Le terme serveur est à comprendre comme système, chacun de ces « systèmes racines » pouvant comporter un grand nombre de serveurs répartis sur tous les continents pour des questions de robustesse.