Chapitre 1

La mécanique d'Internet

Ce premier chapitre est le chapitre technique du livre. Il est divisé en deux parties. La première présente les bases du fonctionnement d'Internet avec ses spécifités techniques. La seconde partie parle de sécurité, des dangers mais introduit aussi quelques notions mathématiques afin de démistifier la cryptographie. Si on considère qu'Internet est un réseau physique avec des protocoles de communication, des logiciels et finalement des utilisateurs qui forment la couche sociale, ce chapitre se concentre sur les premiers niveaux.

1.1 Le réseau

La grande force d'Internet est de permettre aux machines de communiquer entre elles. Historiquement d'autres systèmes ont permis la même chose, mais Internet a gagné la compétition pour devenir l'objet indispensable qu'il est aujourd'hui.

Grâce à Internet et aux logiciels toujours plus conviviaux, des milliards de personnes peuvent communiquer sans ce soucier de la technique sous-jacente. Pourtant il est intéressant de regarder sous le capot pour comprendre les enjeux de pouvoir mais aussi pour mieux comprendre qui contrôle nos usages et comment.

Dans son principe, la mécanique d'Internet est simple. Elle est basée sur deux notions :

1. un empilement de protocoles de communication avec au milieu une *langue* commune composée des protocoles TCP et IP [1], voir l'encart page 12,

2. la connexion de machines en réseaux et l'interconnexion des réseaux (ce qui a donné le nom Inter-Net).

1. pour simplifier, il existe aussi UDP sur IP utilisé pour la vidéo par exemple et d'autres nettement moins utilisés.

Une langue commune

Le premier point souligne le fait que toutes les machines connectées à Internet parlent la langue informatique commune qu'est TCP/IP [2]. Outre l'aspect d'une langue commune, l'utilisation de TCP/IP impose une numérotation unique des machines, comme il existe une numérotation des téléphones. Cette numérotation est appelée l'adresse IP [3] et se présente sous la forme de 4 nombres inférieurs à 256 séparés par des points comme `134.157.1.12`.

Ainsi deux ordinateurs respectant le protocole TCP/IP peuvent se contacter et communiquer si il existe une liaison entre eux.

TCP/IP, le protocole d'Internet

Les informaticiens ont découpé les communications, entre deux machines ou entre deux programmes, en couches avec le principe que chaque couche communique seulement avec les deux couches l'encadrant. Le modèle de référence des informaticiens fait intervenir 7 couches allant de la couche physique, comment transmettre des 0 et des 1 avec du courant électrique ou des photons, à la couche applicative sur qui définit le protocole de communication d'un programme.

Internet réduit le nombre de couches mais le principe reste le même. Il impose seulement d'utiliser le tronc commun que sont la couche de transport, TCP ou UDP, et la couche réseau qu'est IP. C'est la raison pour laquelle on associe Internet au protocole TCP/IP.

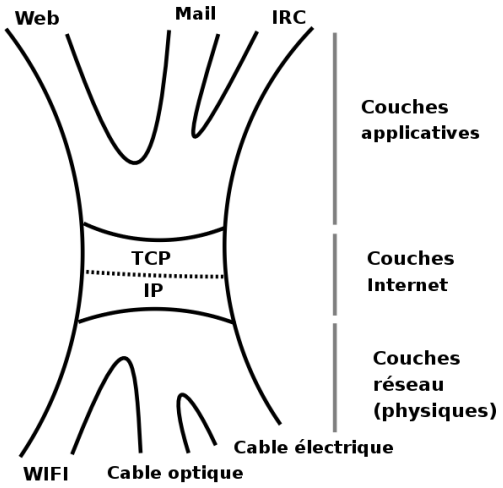


FIGURE 1.1 – TCP/IP au cœur du protocole de communication d'Internet

Ainsi les supports physiques et leur protocole peuvent varier sans avoir d'impact sur la compatibilité Internet. Ce modèle permet aussi de définir tous les protocoles applicatifs désirés tant qu'*in fine* leur couches applicatives peuvent se raccorder à la couche de transport. D'où la possibilité de créer toutes les applications imaginables.

2. en informatique on parle de *protocole*.
3. Ip version 4, la version encore la plus répendue. Pour une brève description de la version suivante, IP version 6, voir l'encart page 14.

Des machines connectées en réseau, des réseaux interconnectés

Le second point souligne la structure d'Internet : Internet est une interconnexion de réseaux indépendants, cf figure 1.2. Que vous soyez chez vous, au travail ou à l'hôtel, votre connexion à Internet passe par un premier réseau qui est le réseau local. Chez vous il est composé de vos appareils connectés et sa limite est la *box* qui vous relie au réseau de votre fournisseur d'accès. Sur le dessin votre réseau local peut être le Bleu et celui de votre fournisseur le Marron. Le réseau Vert étant relié au réseau Marron, vous pouvez vous y connecter depuis votre réseau local.

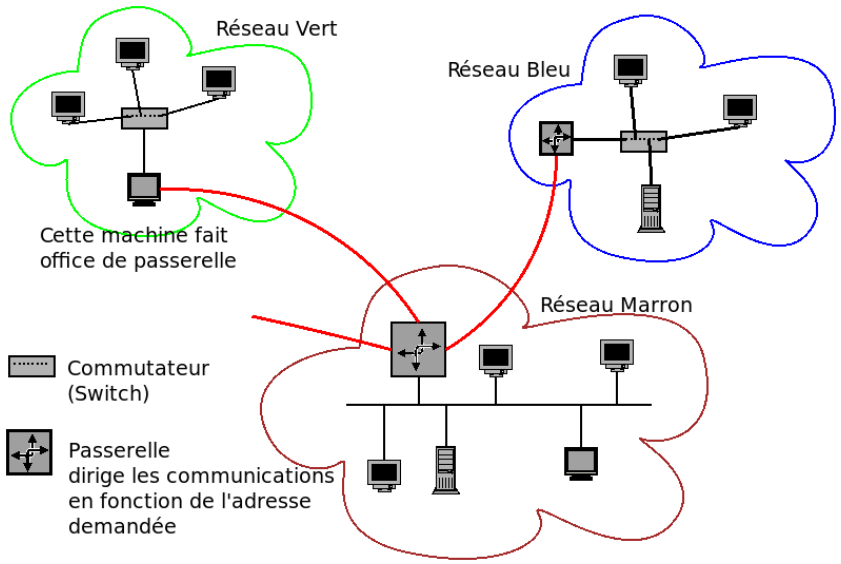


Figure 1.2 – Des réseaux interconnectés.

La connexion entre les réseaux passe par des machines spéciales très importantes puisque permettant l'accès aux autre réseaux donc à Internet. Il s'agit des passerelles qui sont le plus souvent des routeurs (une box est un routeur).

On retrouve ce même schéma avec des grosses organisations, les réseaux Bleu et Vert étant des réseaux de départements et le réseau Marron étant le réseau principal de l'organisation. Cette architecture permet au réseau principal de contrôler ce qu'il laisse passer vers Internet.

Cette notion de sous-réseaux apparaît aussi au niveau des adresses IP dans l'ordre des 4 nombres. Le premier nombre indique une zone, le second une sous-zone… comme `33 1 42 37 xx xx` indique que ce téléphone est en France, dans la région parisienne, à coté de la Croix de Berny (237 étant BER). Mais la comparaison se limite là car l'adressage IP est plus souple, les sous-réseaux n'ayant pas obligatoirement le même préfixe que le réseau auquel ils appartiennent et surtout l'adresse IP n'est pas géographique. D'ailleurs l'attribution des numéros de téléphone a aussi évolué et n'est plus basée sur la position géographique.

Le danger de l'analogie avec le téléphone

```
B1gron : jve te tracé ac ton ip
Nonoeil: Cool.
B1gron : tu va voir
Nonoeil: Oui. Je vais voir, comme tu dis.
B1gron: put1 sa marche pa!!! ta 1 brouyeur????
Nonoeil: Mais qu'est-ce qu'il dit l'autre ? Qu'est-ce qui ne marche pas ?
B1gron: sa sonne mm pa che toi
Nonoeil: ça sonne ? Je suis au boulot là, tu vas tomber sur le Central,
         si t'appelles, mon rigolo
B1gron: ok alor le central a 1 brouilleur
Nonoeil: le central a ce qu'il veut en même temps
B1gron: jvé tosser cher a coze de ses coneries
Nonoeil: Ouais ouais. Si tu le dis !
Myrdène: Mais attends... T'as composé son numéro IP sur ton portable,
         B1gron ?
B1gron: ui pk???
```

source : Les perles d'IRC, `www.danstonchat.com`

Ainsi une entreprise connue possède les adresses IP qui commencent par `129.42` [4]. Il est probable qu'elle a distribué à ses départements des sous-réseaux comme `129.42.2.xxx` pour le département Vert, `129.42.3.xxx` pour le Bleu etc...

Si le département Bleu s'achète une connexion directe vers Internet qui ne passe pas par le réseau Marron, alors cela lui offre deux façons de se connecter à Internet. Il est fort probable que les responsables du réseau n'apprécient guère car ils ne pourront plus filtrer toutes les communications entre l'entreprise et Internet, ce qui rendra d'autant plus difficile la protection du réseau.

IPv6

La nouvelle version d'IP est la version 6, déjà en activité même si l'ancienne version, la version 4, reste la plus courante. La version 6 a été créé afin principalement de répondre au manque d'adresse IPv4 pour tout le monde. Avec 128 bits par adresse, la version 6 offre $2^{128} = 3,4\ 10^{38}$ adresses ce qui fait 670 millions de milliards d'adresses par millimètre carré sur la Terre.

Préfixe (48 bits)	Sous-réseau (16 bits)	Interface (64 bits)
`2001:0db8:0000:`	`85a3:`	`0000:0000:ac1f:8001`

Table 1.1 – Format des adresses d'IPv6

Une adresse s'écrit en hexadécimal (contrairement à IPv4). Ainsi par exemple on a `2001:0db8:0000:85a3:0000:0000:ac1f:8001` ce qui peut aussi être écrit en suprimant les zéros non significatifs `2001:db8:0:85a3:0:0:ac1f:8001` voire `2001:db8::85a3::ac1f:8001`.
On peut faire varier la taille du préfixe et du sous-réseau.

4. on peut trouver son nom avec la commande `whois 129.42.1.1`

Des adresses à usage privé

Comment a-t-on une adresse IP ? En la demandant à celui qui vous fournit la connexion à son réseau. Il vous donnera une adresse parmi celles qui lui ont été attribuées.

Vous pouvez aussi utiliser, sans rien demander, des adresses réservées à usage interne et donc interdites sur Internet. Il s'agit pour la version 4 d'IP de :

- `10.xx.xx.xx` pour se faire un très gros réseau local (16 millions de machines),
- `172.16 à 31.xx.xx` pour un gros réseau (1 million de machines)
- `192.168.xx.xx` pour un réseau moyen (65 000 machines quand même),
- `127.0.0.1` pour désigner votre machine (chaque machine a au moins 2 adresses IP : celle ci qui ne sert qu'à usage interne, l'autre pour communiquer avec l'extérieur.)

Pour IP version 6, les adresses privées appartiennent à l'espace `fc00::/7` (cf RFC4193). En pratique cela revient à choisir comme préfix `fd` puis à choisir de façon aléatoire l'identifiant global et l'identifiant de sous-réseaux. On a ainsi 2^{64} adresses pour soi et très peu de chances qu'une autre personne ait le même réseau privé.

Internet : des milliers de réseaux

D'un point de vue topologique, Internet n'est que la duplication en millions d'exemplaires de la figure 1.2. Pour avoir une image globale il faut détecter quels réseaux sont reliés à quels réseaux, ce qu'a fait sur une partie d'Internet CAIDA en 2001 en analysant 535 000 nœuds d'Internet et plus de 600 000 connexions, cf figure 1.3.

Un point, une adresse IP, est rattachée à un son réseau local et forme un premier groupe de point sur le dessin, une petite tache. Ce groupe du réseau local est le plus souvent rattaché à un groupe qui est celui de son fournisseur d'accès. Ce groupe appartient à un plus grand groupe qui est le réseau du cablo-opérateur [5]. Enfin les cablo-opérateurs ont des interconnexions entre eux.

On voit que le réseau n'est pas totalement distribué mais que chaque groupe a un nœud d'interconnexion qui relie le groupe au groupe père. Ceux qui contrôlent ces nœuds d'interconnexion peuvent limiter les communications, les bloquer ou les espionner. Bien sûr un État peut faire de même avec les nœuds qui sont sur son territoire.

Une autre représentation graphique d'Internet est proposée figure 1.4. Cette fois il s'agit d'une version où chaque point représente un réseau. La taille des points correspond à la taille du réseau et la taille des liens au débit entre les réseaux reliés. Les couleurs des points correspondent au type du réseau (commercial, académique, administratif...). Les auteurs de ce dessin on placé les réseaux les plus importants sur les nœuds d'un maillage grossier plus les nœuds moins important sur un maillage plus fin etc.

5. les cablo-opérateurs sont les entreprises qui posent les câbles d'Internet. Les grandes entreprises des télécommunications sont souvent des cablo-opérateurs, cf section [?].

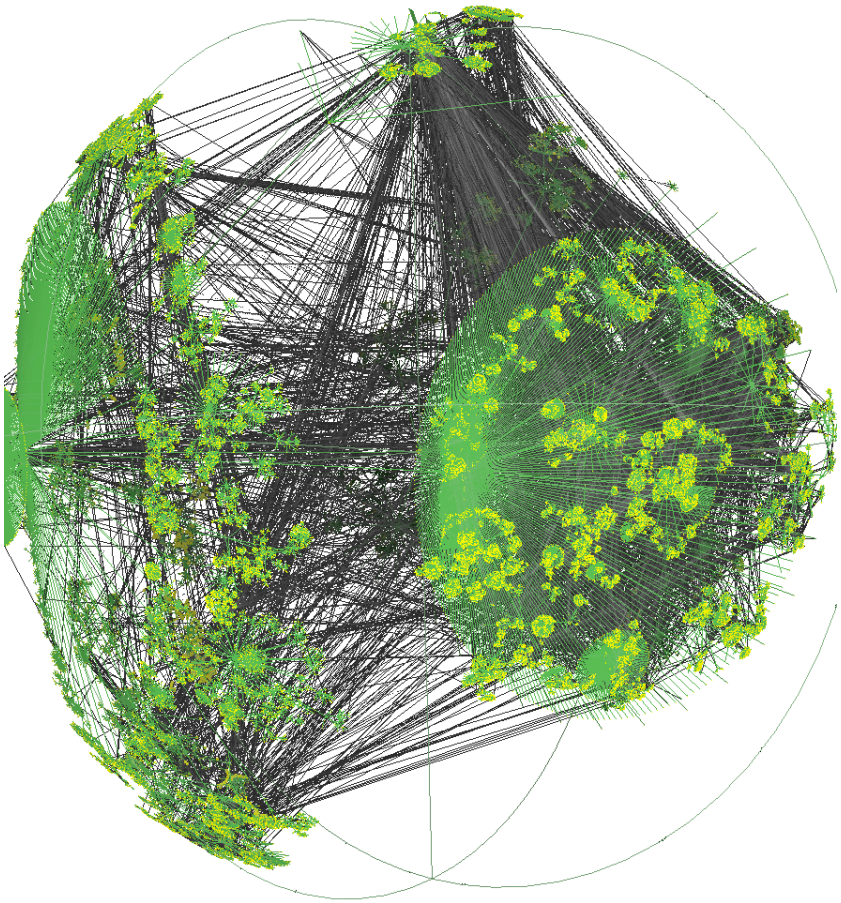


FIGURE 1.3 – Une partie d'Internet vue par le logiciel Walrus
source : CAIDA, mars 2001

1.1.1 L'information relayée de réseaux en réseaux

Que l'on envoie un mail à une machine distante ou que l'on récupère une page web, le principe est le même : l'information est découpée en paquets de données et relayée de réseaux en réseaux.

Traceroute montre le chemin

La détection des réseaux et de leur interconnexion peut se faire simplement à l'aide de la commande `traceroute`, mais aussi par le Web à partir de machines qui offrent ce service, cf http://www.traceroute.org/. Ce programme permet de suivre la route d'un chemin entre deux machines d'Internet. Si l'affichage produit peut sembler abscons au premier abord, il est en fait relativement simple : chaque ligne représente une machine par laquelle passe le message. Ainsi on peut connaître son environnement et la qualité de sa connexion à Internet ou au moins aux nœuds d'Internet que l'on considère le plus important.

En agrégeant les résultats on peut présenter une vue partielle des réseaux d'Internet et de leurs connexions comme le font les figures 1.3 et 1.4.

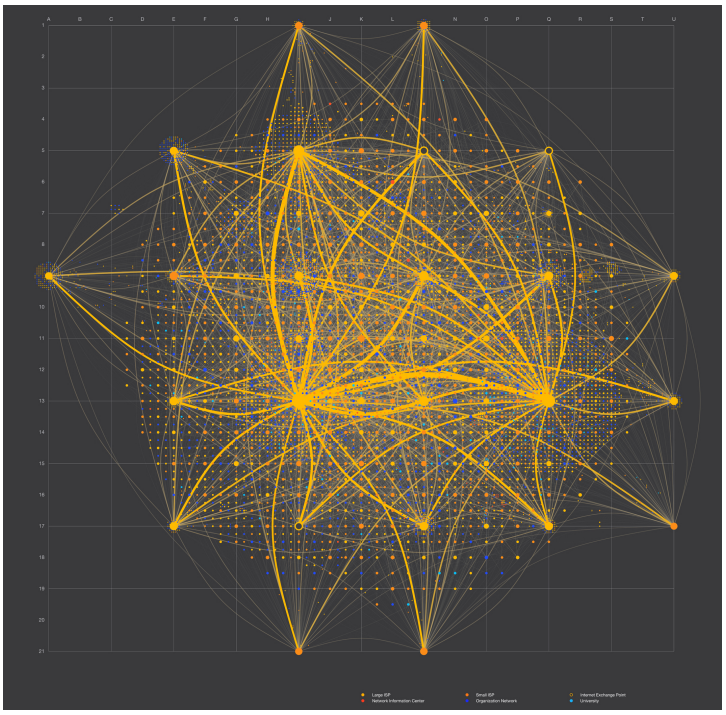


Figure 1.4 – L'interconnexion des réseaux (AS) d'Internet
source : Noosphe.re – 2011

Un exemple : la route entre deux universités

Dans l'exemple qui suit, la connexion entre Jussieu et le MIT n'utilise que des réseaux académiques :

```
(mendel)../home/ricou>traceroute www.mit.edu
traceroute to www.mit.edu (18.7.22.83), 30 hops max, 40 byte packets
 1  134.157.204.126 (134.157.204.126)
 2  cr-jussieu.rap.prd.fr (195.221.126.49)
 3  gw-rap.rap.prd.fr (195.221.126.78)
 4  jussieu-g0-1-165.cssi.renater.fr (193.51.181.102)
 5  nri-c-pos2-0.cssi.renater.fr (193.51.180.158)
 6  nri-d-g6-0-0.cssi.renater.fr (193.51.179.37)
 7  renater-10G.fr1.fr.geant.net (62.40.103.161)
 8  fr.uk1.uk.geant.net (62.40.96.90)
 9  uk.ny1.ny.geant.net (62.40.96.169)
10  esnet-gw.ny1.ny.geant.net (62.40.105.26)
11  198.124.216.158 (198.124.216.158)
12  nox230gw1-PO-9-1-NoX-NOX.nox.org (192.5.89.9)
13  nox230gw1-PEER-NoX-MIT-192-5-89-90.nox.org (192.5.89.90)
14  B24-RTR-3-BACKBONE.MIT.EDU (18.168.0.26)
15  WWW.MIT.EDU (18.7.22.83)
```

Un essai fait d'une machine chez un fournisseur d'accès commercial français vers une université française fera apparaître la machine passerelle `renater.par.franceix.net` qui sert de pas-

serelle entre Renater et les réseaux commerciaux. Elle est située dans le GIX [6] nommé France-IX [7] (anciennement SFINX) qui permet à tous les opérateurs Internet de se relier entre eux suivant leurs accords, dit accords de peering.

Essayons de comprendre le chemin suivi par notre paquet IP entre Jussieu et le MIT. Le premier intermédiaire que notre message va rencontrer est la passerelle de notre réseau. Son adresse IP est 134.157.204.126 comme on le voit sur la ligne numérotée 1. De là on rejoint l'interconnexion entre Jussieu et le RAP, réseau académique parisien, en 2, pour entrer sur le réseau universitaire français, Renater, en 4, cf figure 1.5.

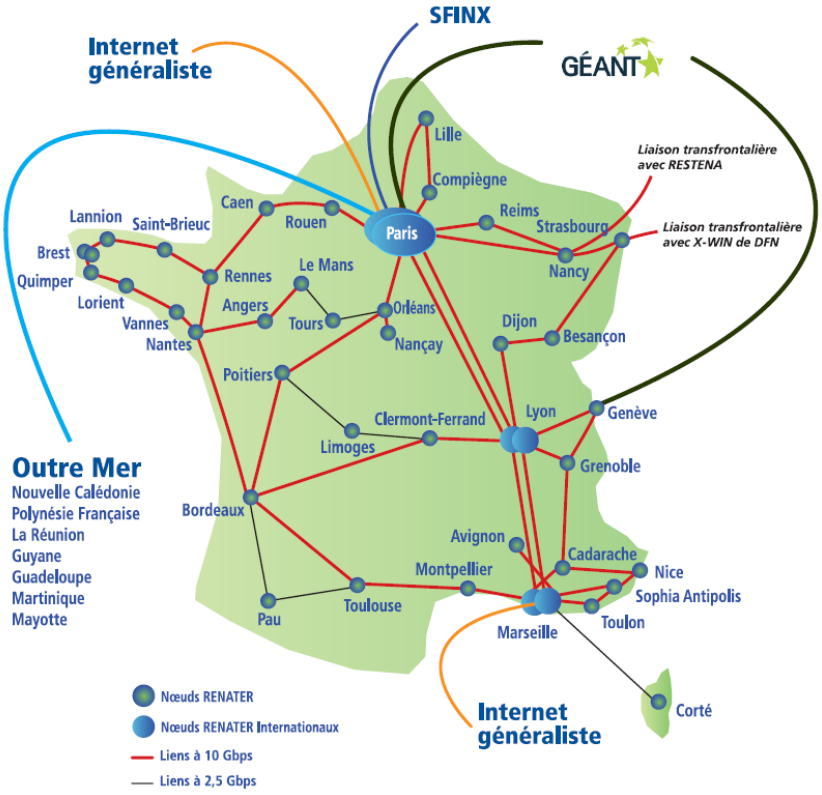


FIGURE 1.5 – Renater, le réseau universitaire français
source : Renater 2011

On passe de Renater à Géant, le réseau universitaire européen, en 7, qui nous envoie en Angleterre, en 8, d'où on va à New-York rejoindre le réseau académique d'Amérique du Nord, Internet 2, en 9 et 10, cf figures 1.6 et 1.7.

De là on passe sur NOX, le réseau de la Nouvelle Angleterre, en 12 et 13, pour atteindre le réseau du MIT, en 14, et enfin le serveur web `www.mit.edu`, en 15, cf figure 1.8.

6. Global Internet eXchange point ou IXP, Internet eXchange Point.
7. la liste de membres parisien de ce GIX est sur https://www.franceix.net/en/france-ix-paris/members-in-paris/

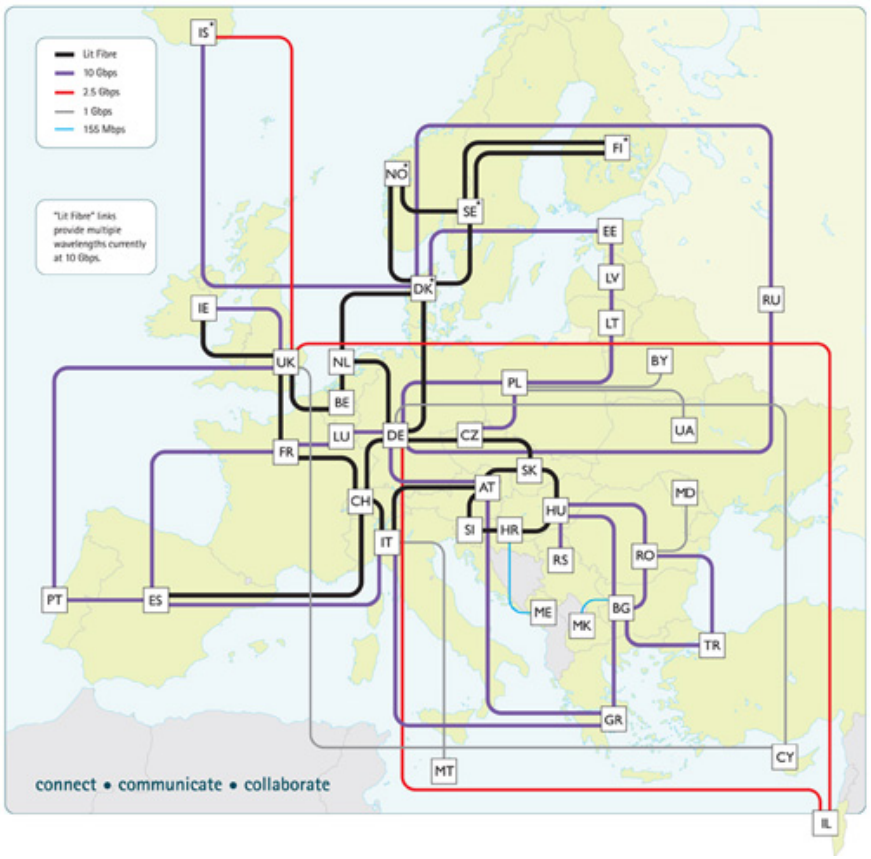


Figure 1.6 – Géant, le réseau universitaire européen
source : Géant, 2012

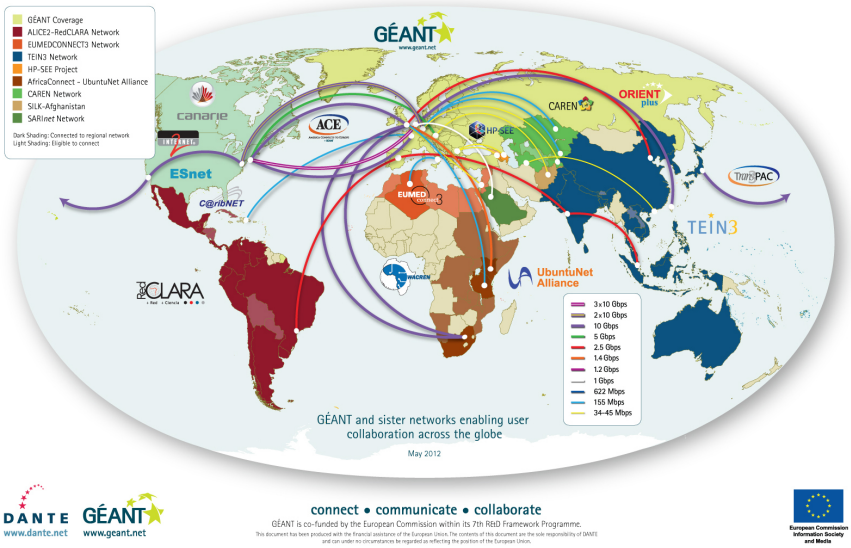


Figure 1.7 – Interconnection entre Géant et les autres réseaux académiques
source : Géant, 2012

m.à.j. sur http://www.ricou.eu.org/e-politique.html

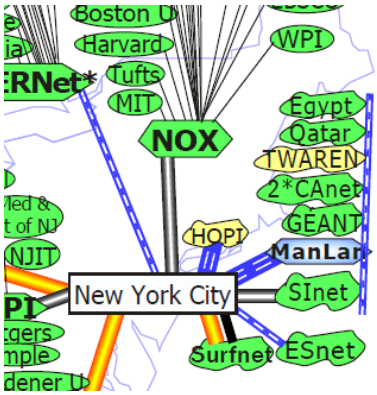


FIGURE 1.8 – Internet 2 et NOX pour arriver au MIT
source : Internet 2, 2005

Le calcul du débit

Sachant que le débit entre deux machines est celui du nœud le plus faible, si un réseau a un goulot d'étranglement en un point, cela se ressent directement. Aussi il est toujours bon de savoir quels seront vos partenaires principaux et de savoir par quels opérateurs vous devrez passer. En pratique il faut savoir quels accords d'interconnexion a votre hébergeur, avec quels opérateurs, à quel débit et quelle est l'occupation moyenne du réseau.

Certains opérateurs proposent de pouvoir suivre en direct la *météo* de leur réseau, malheureusement cette information est devenue rare en France. On peut néanmoins avoir quelques informations :

— L'état du réseau chez Free est visible sur http://www.free-reseau.fr/
— Le tableau de bord de Nerim permet de voir vers quels réseaux vont les données des clients de Nerim et le débit

Il est aussi possible de faire le travail à la main avec des outils comme `iperf3` qui mesurent le débit entre deux machines dont on a le contrôle ou entre sa machine et un serveur ouvert comme ceux indiqué sur https://iperf.cc/.

1.1.2 Des domaines et des noms

Des noms Au commencement les machines avaient des numéros et rapidement des noms tant pour faciliter la vie des humains que pour permettre de changer l'adresse numérique de la machine sans en changer son nom [8]. Dès 1973 la correspondance entre les noms et les adresses numériques des machines reposait sur un fichier avec les noms et adresses IP de toutes les machines d'Internet. Cela impliquait de télécharger ce fichier régulièrement pour connaitre les nouvelles machines reliées à Internet et les changements d'adresse. Puis le nombre de machine est devenu

8. L'adresse IP est liée au réseau ce qui implique de changer d'adresse IP lorsqu'on change une machine de réseau. Avoir un nom qui redirige vers une adresse IP permet de changer l'IP tout en garantissant la continuité des service basés sur le nom (mail, web…).

trop important et variait trop vite pour garder ce fichier à jour sur toutes les machines. Aussi en 1984 on a créé un service appelé Domain Name System, DNS, qu'on interroge pour connaître l'adresse IP d'une machine dont on connaît le nom.

L'archéologie des noms de machines

Un ami qui aime consulter les textes de référence de l'Internet que sont les RFC, Request For Comments, a fait cette constatation :

- RFC 1, avril 1969 : aucune mention des noms des machines, juste les adresses (sur 5 bits)
- RFC 33, février 1970 (remplace RFC 1) : toujours pas de noms, mais les adresses passent à 8 bits
- RFC 229, septembre 1971 : première mention des noms. Aucun mécanisme de résolution n'est envisagé (même pas un simple fichier de correspondances) mais il y a une table des noms officiels et de l'adresse correspondante.
- RFC 606, décembre 1973 : première mention d'un mécanisme de résolution, un fichier, avec une syntaxe formelle, placé à un endroit bien connu, le futur HOSTS.TXT

Le DNS tel qu'il est aujourd'hui arrivera seulement en 1984.

Des domaines Internet étant un ensemble de réseaux, il semble naturel que chaque réseau ait un nom de domaine dans lequel il peut ranger ses sous-réseaux et machines. Mais se pose alors la question de l'organisation globale et comment faire pour que chaque réseau connaisse les noms de tous les autres réseaux. La réponse retenue a éte un système d'arborescence dont chacun connait la racine et qui permet de retrouver tout le monde à partir de la racine.

Regardons la figure 1.9 page 22. La terminaison la plus à droite est la machine `whois.eu.org.`. Pour comprendre ce nom il est plus simple de le lire de droite à gauche avec au début la racine que l'on nomme "`.`" [9]. En continuant de droite à gauche on trouve le domaine terminal [10] `org` et ensuite le domaine `eu.org` auquel appartient la machine `whois`.

On imagine bien qu'il ne serait pas gérable que chaque machine obtienne son nom d'une seule autorité tant pour des raisons de performance que de praticité (sans parler du contrôle absolu qu'aurait ainsi cette autorité sur Internet). Aussi le nommage d'Internet se base sur un système de délégation de zone. Ainsi `.org` à délégué la gestion de `.eu.org` ce qui fait que `.eu.org` est une zone indépendante de la zone `.org` et qu'elle peut faire ce qu'elle veut en "dessous" de `eu.org`.

La figure 1.10 montre que le domaine `eu.org` délègue les sous domaines `gr`, `dk` et `uk.eu.org` mais gère le sous domaine `fr.eu.org` et les machines `www` et `whois.eu.org`.

9. .org est un raccourci accepté pour .org. où le point final qui correspond à la racine est oublié.
10. *Top Level Domain* ou TLD

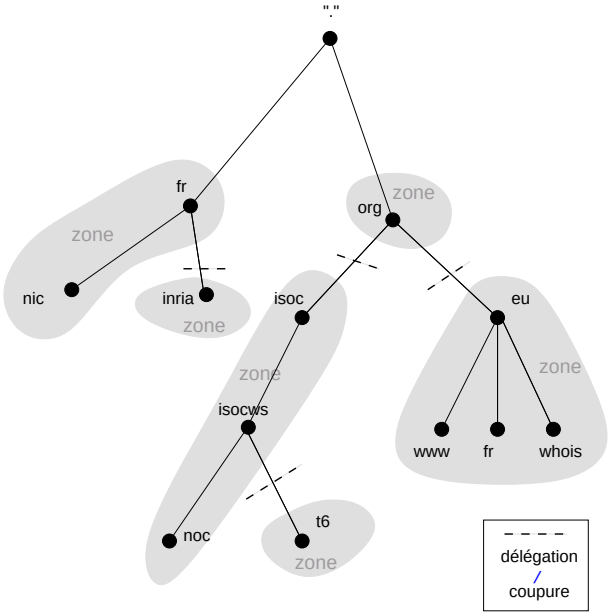


FIGURE 1.9 – Une toute petite partie de l'arborescence des noms de domaines

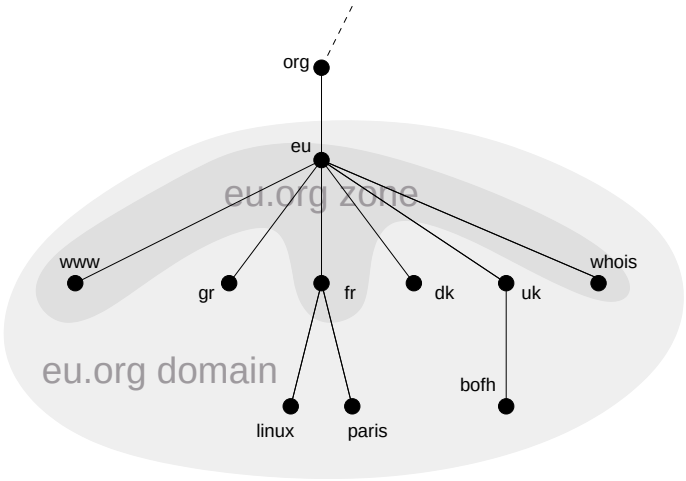


FIGURE 1.10 – La différence entre une zone et un domaine

Il existe donc :

— les *domaines* qui comprennent tout ce qui finit par le nom de domaine,
— les *zones* formées de l'ensemble des machines et sous-réseaux contrôlés par le propriétaire du nom de domaine.

On comprend ainsi pourquoi le propriétaire d'un domaine, comme `.fr`, ne peut être tenu responsable de ce qu'on trouve sur un serveur web hors de sa zone, comme `www.tf1.fr` par exemple.

Par contre, techniquement parlant, il peut toujours retirer la délégation de zone et donc fermer le domaine `tf1.fr`. De même le gestionnaire du point final peut fermer `.com` ou `.fr`.

Confessions d'un voleur ou l'argent des noms de domaines

par Laurent Chemla, co-fondateur de la société Gandi

Je vends des noms de domaines sur Internet.

Un peu d'histoire et de technique sont nécessaires pour comprendre à quel point je suis un voleur.

Un nom de domaine, c'est ce qui sert à identifier un ordinateur sur Internet. Quand on vous propose d'aller visiter www.machinchose.org on vous indique un nom d'ordinateur (www) qui se trouve dans le domaine « machinchose.org » et qui contient ces informations que vous pouvez consulter sur le Web.

Sans un nom de ce genre, un ordinateur ne peut être consulté qu'en utilisant un numéro, tel que par exemple 212.73.209.251. C'est nettement moins parlant et beaucoup plus difficile à mémoriser. Alors pour simplifier on donne des noms aux ordinateurs qui contiennent de l'information publique. Ce qui nécessite, bien sûr, une base de données qui soit capable de retrouver un numéro à partir d'un nom. Et que cette base soit unique et accessible de n'importe où.

Pendant des années, ce système a fonctionné grace à un organisme de droit public financé par le gouvernement américain. L'Internic (c'était le nom de cet organisme) se chargeait de faire fonctionner la base de donnée, et chacun pouvait y ajouter le nom de domaine de son choix, gratuitement, selon la règle du « 1er arrivé 1er servi ».

Puis vint le temps de l'ouverture d'Internet au grand public (1994), et la fin des subventions gouvernementales au profit du seul marché. Et là, surprise : une agence publique (qui gérait gratuitement ce qu'il faut bien appeler une ressource mondiale unique) fut transformée en entreprise commerciale (Network Solutions Inc, ou NSI), sans que quiconque s'en émeuve particulièrement, et se mit à vendre 50$ par an (puis 35$ par an dans un fantastique élan de générosité) ce qui était totalement gratuit peu de temps avant. Et pour son seul profit.

Je dois vous livrer un chiffre qui, s'il n'est pas confidentiel, mérite cependant le détour : le coût réel de l'enregistrement d'un nom dans la base de données mondiale, y compris le coût de fonctionnement d'une telle base, a été évalué il y a deux ans à 0,30$.

Des chiffres comme ça, je pourrais en donner beaucoup. Je pourrais dire par exemple qu'en estimant le nombre de domaines enregistrés par NSI à une moyenne mensuelle de 40.000, son bénéfice sur les 5 dernières années tourne autour des 80 millions de dollars. Et encore ce chiffre est-il une estimation basse, quand on sait que NSI vient d'être racheté par une autre Net-Entreprise pour la modique somme de 21 milliards de dollars.

Et pourtant, NSI vend du vent, tout comme moi. En fait, nous vendons le même vent.

source : Extrait d'un article publié dans le journal Le Monde en avril 2000 et disponible dans son intégralité sur http://www.chemla.org/textes/voleur.html.

`eu.org`, des domaines gratuits

L'exemple `eu.org` est d'autant plus intéressant que ce domaine délègue gratuitement des sous domaines c.a.d. que si vous désirez avoir un sous domaine comme `ricou.eu.org` [a], il suffit de le demander sur le site web `www.eu.org`. Cela demande bien sûr de savoir gérer un sous domaine.

a. `ricou.eu.org` est déjà pris et pour longtemps puisque c'est gratuit.

Trouver l'adresse IP d'un nom, le fonctionnement du DNS

La recherche d'une adresse IP est l'opération initiale pour chaque connexion dès lors que l'on initie la connexion avec le nom de la machine et non son adresse IP. Pour faire la correspondance nom/adresse IP, vous devez avoir indiqué à votre machine l'adresse d'un "Serveur de nom", ou serveur DNS. Si tel n'est pas le cas vous ne pourrez plus vous connecter aux autres machines d'Internet, sauf en donnant directement leur adresse IP bien sûr.

Pour trouver l'adresse IP d'une machine à partir de son nom, votre serveur de nom va lire le nom de la machine de droite à gauche pour savoir à quel autre serveur de nom il pourra demander l'adresse IP s'il ne la connaît pas.

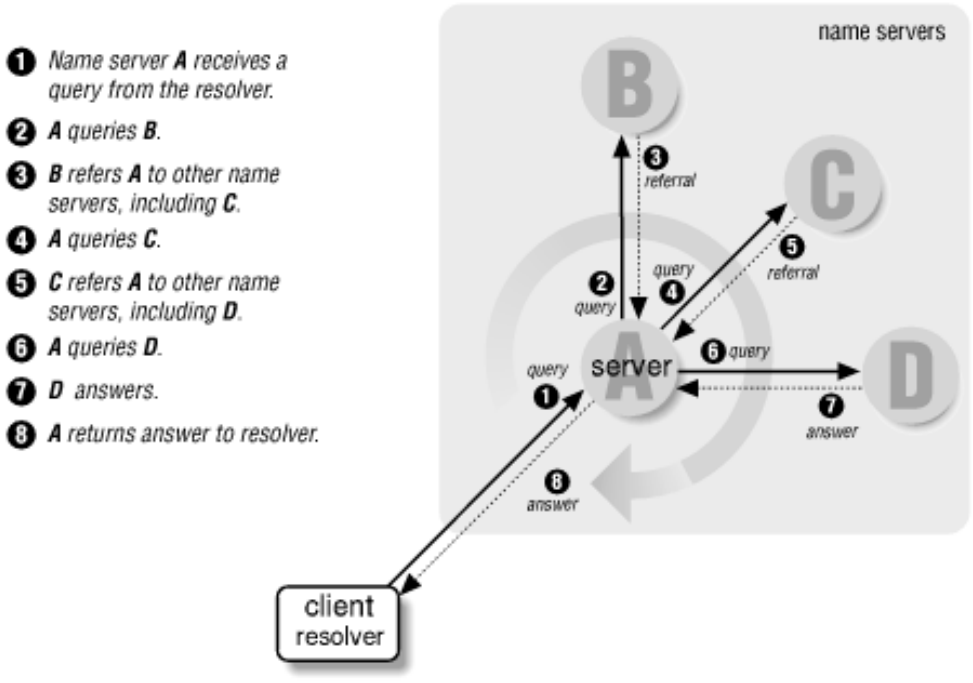


Figure 1.11 – Fonctionnement récursif du DNS
(illustration extraite du livre DNS & Bind chez O'Reilly)

Supposons que l'on cherche à se connecter sur le serveur web `www.jussieu.fr`. Notre serveur

de nom, A sur la figure 1.11, n'a rien en mémoire et donc pas l'adresse IP [11] de cette machine. Aussi il demande a un des serveur racine du DNS [12] dont l'adresse est stockée dans chaque serveur de nom. Le serveur racine, B sur la figure, qui ne connait que les TLD le renverra sur le serveur C qui gère `.fr.`, lequel renverra au serveur D qui gère `jussieu.fr.` et qui donnera l'adresse IP de son serveur web à savoir `134.157.250.59`.

1.2 La sécurité

Internet n'est pas sûr.

On voit que si un serveur DNS nous ment, on ira à une mauvaise adresse IP. Si on désirait consulter son compte bancaire, cela peut être très facheux car notre mot de passe va tomber entre de mauvaises mains. On a vu aussi qu'un message passe de réseau en réseau ce qui laisse entendre que les réseaux intermédiaires peuvent le lire voire le détourner. Mais ces failles structurelles ne sont qu'un petit morceau de possibilités d'agression sur Internet. En fait les possibilités sont immenses pour les agresseurs.

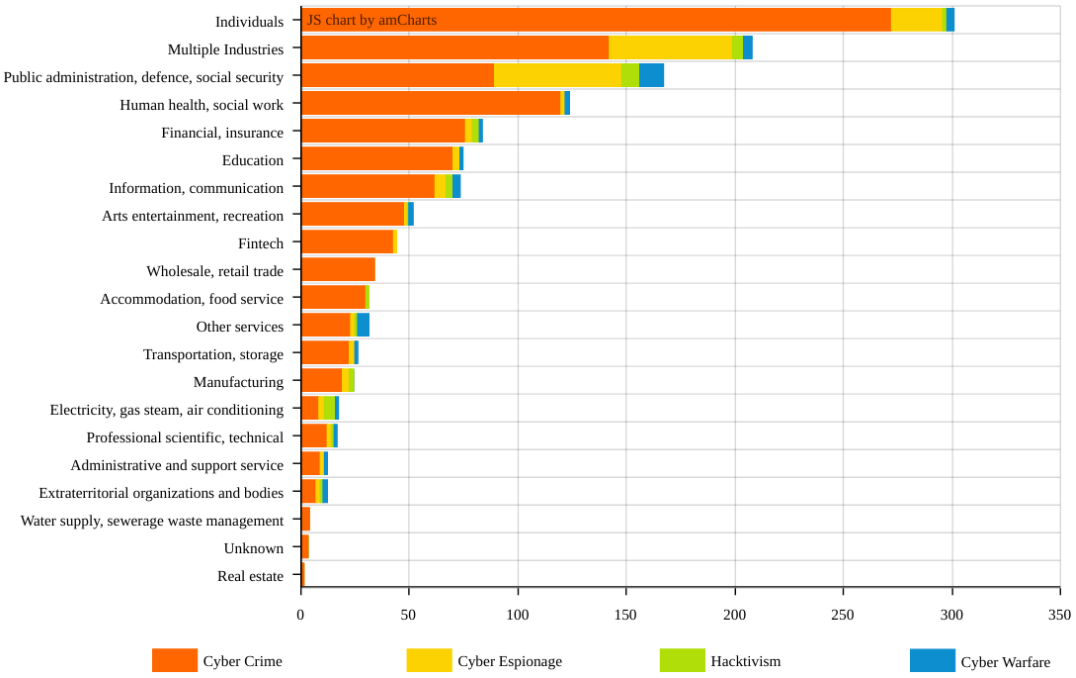


Figure 1.12 – Source des attaques sur Internet et leur cibles
source : Hackmageddon – 2018

À l'origine de l'Internet, club fermé, ce manque de sécurité n'était pas un problème. Mais maintenant qu'Internet est un outil économique et stratégique de première importance, il est naturel-

11. il la conservera un certain temps une fois la demande faite ce qui évite de réitérer le processus à chaque connexion.

12. ce serveur racine est tellement important qu'il est dupliqué en 13 exemplaires. Si ces 13 serveurs (plus en fait, cf la section **??**) tombent tous en panne, Internet s'arrêtera doucement, le temps que les mémoires des serveurs de nom de la planète s'effacent.

lement devenu une cible privilégiée pour tout type d'agression, criminelle, politique, étatique... Comme de plus le risque est quasiment nul pour l'agresseur, on comprend que le cyber crime se porte bien. Si pour un individu le risque peut sembler lointain car virtuel, il est ne faut surtout pas le sous-estimer. Perdre ses économies, ses données ou son identité n'est pas une expérience agréable. La figure 1.12 montre que le cyber-crime est l'activité la plus importante dans les agressions sur Internet et que sa cible numéro 1 est les invidus. Il s'agit de la population la plus fragile car peu de personnes ont des notions de sécurité informatique ou peuvent s'offrir un expert pour les protéger.

Aussi regardons les problèmes de sécurité sur Internet et comment les personnes mal intentionnées en profitent.

1.2.1 Les failles sur Internet

Pour simplifier, rangeons les sources de failles sur Internet en trois types :

- l'architecture d'Internet,
- les bugs logiciels,
- les utilisateurs (utilisateurs finaux mais aussi opérateurs du réseau).

En tant qu'utilisateur on est bien sûr responsable de nos bêtises mais on subit aussi celles des autres. Aussi il est important de prévenir et de se protéger.

Les failles architecturales d'Internet

D'un point de vue technique, Internet actuel a deux vulnérabilités fondamentales :

- les messages sont transmis sans protection sur le réseau
- l'identification de l'interlocuteur qu'il soit individu ou machine est peu fiable.

Ces failles date de la création d'Internet, ou d'IPv4, à une époque où le réseau était universitaire et sans que la sécurité soit considérée comme utile. Il était alors bien agréable de pouvoir savoir que tel collègue à l'autre bout du monde est debout connecté à telle machine et donc de pouvoir lui afficher une image sur son écran afin de travailler dessus ensemble. Tout cela utilisait des protocoles bloqués aujourd'hui pour des raisons de sécurité. Par exemple ce qui permettait d'afficher sur un écran distant une image permettait aussi de lire le clavier distant et donc de voir tout ce tapait le collègue, y compris ses mots de passe.

Donc Internet a été créé sans penser à la sécurité mais fort heureusement il est tout à fait possible d'ajouter une couche de sécurité. Aujourd'hui un utilisateur averti ne risque plus grand chose au quotidien à cause des protocoles mal sécurisés d'Internet. Il peut surfer en mode HTTPS, il peut chiffrer ses mails, se connecter à distance et transférer des fichiers via des canaux sécurisés.

Les communications en clair Le protocole de transport des données sur Internet, TCP/IP, ne prévoit pas de protéger les données transportées. Tous les paquets sont transmis en clair. Ainsi toute personne qui contrôle un des ordinateurs par lequel passent les données peut les lire.

Par exemple lorsqu'on surfe sur le web, on utilse souvent le mode non sécurisé HTTP et non HTTPS [13], ce qui permet à notre fournisseur d'accès de voir toutes les pages qu'on regarde.

Autre exemple, au niveau d'un réseau local, à la maison, tous les paquets sortant vers Internet doivent passer par une passerelle. Le contrôle de cette machine permet la lecture de tout ce qui va et vient. Toujours sur un réseau local une personne qui est physiquement sur le même fil Ethernet qu'une autre [14] peut y détecter le courant qui y passe et donc lire les données.

Voici ce qu'un renifleur de paquets IP comme le programme `tcpdump` permet voir passer si on est sur le chemin [15] pour écouter :

```
18:12:23.988 IP (tos 0x0, ttl  64, id 24337, offset 0, flags [DF], proto:
TCP (6), length: 1019) po8.pmmh.espci.fr.3192 > mg-in-f147.google.com.www:
P 1:968(967) ack 1 win 1460 <nop,nop,timestamp 7090623 2265318920>
E..._.@.@.i..6Q..U...x.P.Yn....B....r......
.l1.....GET /search?hl=fr&q=piratage+i
```

On voit ici un paquet destiné à Google avec une demande de recherche contenant le mot "piratage".

Il est donc important de garder à l'esprit que les données ne sont pas protégées par le réseau et que le travail de protection doit être fait au niveau des applications [16] afin que les données ne quittent votre machine que chiffrées.

Ainsi depuis l'affaire Snowden et l'espionnage de plus en plus actif des États, les applications WhatsApp et Telegram ont intégré la cryptographie depuis l'émetteur jusqu'au destinataire. Cela veut aussi dire qu'avant 2014, les messages envoyés étaient lisibles par votre opérateur, l'État et tous les pirates sur le chemin.

Si l'application n'a pas de mode de chiffrement intégré, il est possible d'établir un canal sécurisé entre deux machines à l'aide d'un tunnel ou un VPN. Dans ce cas tout ce qui sort de la machine par ce canal est protégé jusqu'à l'autre machine. C'est une bonne solution pour relier son ordinateur portable à son serveur et pouvoir utiliser les réseaux wifi mis à disposition à l'hôtel ou en visite dans une entreprise sans être espionné par le propriétaire du wifi voire par toute personne connectée si le protocole du wifi est trop faible.

On regardera plus en détail les façons de se protéger dans la section sur la cryptographie.

L'identité de l'interlocuteur Comme indiqué au début de cette section, le DNS peut mentir et donner la mauvaise adresse IP lorsqu'on lui demande `www.machin.com`. Lorsqu'on envoie un mail, là aussi le destinataire peut être différent de celui espéré. Le mail peut être intercepté en chemin. Inversement cela peut être le destinataire qui est trompé sur l'identité de l'émetteur ce qui a généré, par exemple, l'arnaque au faux virement qu'on verra.

13. C'est le serveur qui choisi le protocole. On peut voir dans l'URL si on utilise HTTPS. Les navigateurs mettent souvent un cadenas lorsqu'on est en mode sécurisé.
14. Toutes les personnes branchées sur un même *hub* sont sur le même fil Ethernet. Avec un *switch* il est plus difficile d'intercepter les communications mais cela reste possible (voir l'ARP Spoofing).
15. chemin que révèle `traceroute`
16. Cela demande à ce qu'il existe un protocole chiffré pour les applications concernées.

Aussi il est important d'avoir la preuve qu'on communique avec la bonne machine ou la bonne personne et pour cela, là encore, la cryptographie apporte une solution et en particulier le système de certification [17].

Les bugs logiciels

Un bug logiciel est une erreur de programmation que le pirate peut exploiter pour obtenir un accès privilégié à une machine ou au moins pour y exécuter des commandes. Les bugs existent partout. Ils sont le plus souvent référencés car connus mais parfois ils sont nouveaux. Un bug nouveau qui permet de prendre le contrôle d'une machine à distance vaut très cher sur le marché noir.

Un exemple classique de bugs que peut utiliser un pirate consiste à donner à un programme une valeur à laquelle il ne s'attend pas.

Par exemple lorsqu'on envoie des données sur Internet, elles sont découpées en paquets. Chaque paquet a des méta-données qui décrivent le paquet dont une qui est la taille du paquet. Dans les années 90 on a découvert que si on indique que la taille du paquet est de -1 octet alors l'ordinateur qui reçoit le paquet se fige. Vous pouviez ainsi très facilement figer n'importe quel serveur sur Internet.

Un autre exemple s'appelle les injections SQL. Il s'agit d'introduire sa requête dans une base de données. De nombreux services dont des serveurs web s'appuient sur des bases de données. Cela permet par exemple de faire une requête pour avoir le prix d'un produit. En regardant comment le serveur web soumet sa requête à la base de donnée, on peut la modifier pour effectuer notre requête. Elle peut aussi bien être la destruction de la base de l'exportation d'information privée [18].

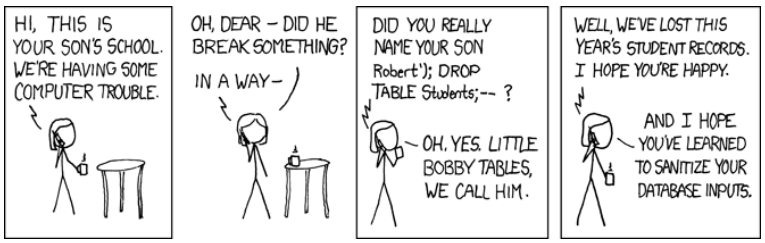


FIGURE 1.13 – Exploit [19] d'une mère
source : https://xkcd.com/327/

Enfin regardons les débordements de mémoire. Sachant que les variables d'un programme sont contigües au code du programme dans la mémoire tampon, il est possible de donner une valeur à la variable qui, dans certains cas, va déborder de son espace alloué et modifier d'autres choses. Le plus souvent cela va casser le programme et l'arrêter mais un bon pirate pourra en profiter pour faire faire ce qu'il veut alors même que le programme tourne sur une machine distante.

17. La machine produit un document signé par une autorité de certification que notre navigateur connait ce qui lui permet de vérifier que la signature est valide, cf section 1.3.3
18. il est heureusement facile de se protéger de ce type d'attaque mais il faut y penser.
19. L'exploitation d'un bug s'appelle un exploit.

Il existe bien sûr de nombreuses autres façons de profiter d'un bug ou d'une faiblesse de conception d'un programme. Pour l'instant il n'est malheureusement pas possible actuellement de garantir qu'un programme n'en contiennent pas [20].

La meilleure façon de lutter contre les bugs est de veiller à ce que sa machine soit régulièrement mise à jour pour y appliquer les correctifs.

L'erreur humaine

Quels que soient les outils de sécurité mis en place, il est difficile voire impossible de protéger un système si un utilisateur autorisé aide le pirate. Cette aide peut aller du mot de passe trop simple à l'installation sur sa machine d'un programme comprenant un logiciel malveillant (*malware*).

Une étude lors de la conférence DEFCON 2016 [21] indique que 84% des pirates utilisent les réseaux sociaux pour mener leurs attaques c.a.d. qu'ils cherchent la faille humaine. Et s'ils le font, c'est que ça marche…

Le premier type d'attaque consiste à profiter de la naïveté humaine ou simplement au manque de compréhension des bases de l'informatique. Les attaques de ce type sont innombrables et parfois très bêtes mais l'idée est qu'en l'envoyant à des millions de personnes, si 1% se fait avoir c'est gagné. Dans ce domaine deux types d'attaques ont fait leurs preuves : l'arnaque et l'hameçonnage ou filoutage (*phishing* en anglais). Le FBI a évalué le coût de ces attaques à plus de 10 milliards de dollars entre 2013 et 2018.

La seconde faille humaine consiste à pousser l'utilisateur à installer un programme malveillant sur sa machine. Cela peut être fait en mettant en libre téléchargement un logiciel merveilleux qui comprend le *malware* ou en l'incluant dans une pièce attachée à un mail.

Une troisième faille humaine est l'erreur de ceux qui sont au contrôle. Si un administrateur d'un système informatique configure mal un logiciel ou un appareil alors des pirates pourront en profiter. Ainsi des numéros de cartes bleues de clients de Tati étaient disponibles sur le web et référencés par Google simplement parce que Tati n'avait pas configuré correctement son serveur web [22].

Mais la liste des erreurs possibles n'est malheureusement pas fermée, l'imagination étant sa limite. En 2018 la Nasa s'est fait piratée car un employé a mis sur le réseau interne un petit ordinateur mal protégé qui a servi de cheval de Troie.

Les failles qui n'en sont pas

Un ordinateur, un composant du réseau peut tomber en panne. La justice peut demander l'accès à un ordinateur et à son contenu. On peut perdre son ordiphone. Un cambrioleur peut voler

20. C'est possible sur des tout petit programme comme celui qui contrôle une machine à café, mais pas sur les programmes usuels.

21. https://www.esecurityplanet.com/hackers/fully-84-percent-of-hackers-leverage-social-engineering-in-attacks.html

22. cf affaire Tati versus Kitetoa, http://www.kitetoa.com/Pages/Textes/Les_Dossiers/Tati_versus_Kitetoa/historique.shtml

un ordinateur portable. Il existe bien des façons de perdre le contrôle ou l'accès à ses données, d'avoir un serveur coupé de l'Internet sans pour autant que l'on puisse parler de cyber-attaque. C'est évident lorsqu'on le dit mais c'est souvent une problématique sous-estimée.

Pour lutter contre ces désagréments aux conséquences vraiment graves parfois, il existe des stratégies qui ont fait leurs preuves :

- faire des sauvegardes dans des lieux différents [23],
- chiffrer ses données (voire tout le disque dur),
- installer un mouchard qui permet de reprendre le contrôle de sa machine, de son ordiphone, à distance,
- comprendre où sont stockées les données et comment elles sont stockées, voir l'encart FBI/CIA.

Enfin dans la liste des failles qui n'en sont pas mais qui pourraient en être, il y a les programmes écrits en JavaScript qui s'exécutent en arrière plan lorsque vous regardez une page web. Ils ne vont que consommer de l'énergie, par exemple pour miner des bitcoins à leur bénéfice, mais cela peut mettre un ordiphone à genoux rapidement. Notons que souvent le site web qui vous envoie le programme Javascript ne le fait pas volontairement, il a été lui même piraté.

Le FBI lit les mails du patron de la CIA

Même chez les espions on ne comprend pas toujours très bien que le mail n'est pas protégé s'il est hébergé par un fournisseur de service. L'affaire Petraeus (2012) en a été la preuve.

Paula, la maîtresse cachée du chef de la CIA, David Petraeus, est jalouse de Jill, une copine de ce dernier. Elle lui envoie donc des mails malveillants mais en se protégeant (faux compte Gmail, mails envoyés que depuis des lieux publics et hôtels avec wifi gratuit). Jill porte plainte contre X.

Le FBI récupère auprès de Google les adresses IP des machines qui ont envoyé les mails puis en regardant les registres des hôtels, il s'avère qu'une seule personne était dans ces différents hotêls à ces différents moment : la maîtresse secrète. Une fois Paula identifiée, le FBI obtient de Google l'accès à son compte Gmail officiel. Il y découvre la correspondance avec le chef de la CIA. Le FBI en profite pour regarder aussi le compte Gmail de Jill et découvre une relation avec un général.

Résultat, le patron de la CIA démissionne et le général perd le poste de chef de l'OTAN qui lui tendait les bras.

1.2.2 Les cyber-attaques

Maintenant que l'on a fait le tour des principales failles, regardons comment elles sont exploitées. La figure 1.14 présente la liste des cyber-attaques les plus utilisées en 2017 et 2018.

23. cela protège aussi du chantage aux données chiffrées (*ransomware*)

Top Threats 2017	Assessed Trends 2017	Top Threats 2018	Assessed Trends 2018	Change in ranking
1. Malware	➲	1. Malware	➲	→
2. Web Based Attacks	↑	2. Web Based Attacks	↑	→
3. Web Application Attacks	↑	3. Web Application Attacks	➲	→
4. Phishing	↑	4. Phishing	↑	→
5. Spam	↑	5. Denial of Service	↑	↑
6. Denial of Service	↑	6. Spam	➲	↓
7. Ransomware	↑	7. Botnets	↑	↑
8. Botnets	↑	8. Data Breaches	↑	↑
9. Insider threat	➲	9. Insider Threat	↓	→
10. Physical manipulation/ damage/ theft/loss	➲	10. Physical manipulation/ damage/ theft/loss	➲	→
11. Data Breaches	↑	11. Information Leakage	↑	↑
12. Identity Theft	↑	12. Identity Theft	↑	→
13. Information Leakage	↑	13. Cryptojacking	↑	NEW
14. Exploit Kits	↓	14. Ransomware	↓	↓
15. Cyber Espionage	↑	15. Cyber Espionage	↓	→

Legend: Trends: ↓ Declining, ➲ Stable, ↑ Increasing
Ranking: ↑ Going up, → Same, ↓ Going down

Figure 1.14 – Évolution des agressions sur Internet
source : ENISA Threat Landscape Report 2018

On peut ranger ces agressions en fonction du type de faille :

Malware	Phishing	Denial of Service
Web Based Attack	Insider Threat	Spam
Web Application Attacks	Information Leakage	
Data Breaches	Physical manipulation...	
Cryptojacking		Botnets
Ransomware	Identity Theft	Cyber Espionage

Table 1.2 – Classement des agressions par type de faille
: Bug : Utilisateur : Architecture d'Internet

Les États-Unis ont estimés le coût de ces attaques sur leur territoire entre 57 et 109 milliards de dollars en 2016. La marge d'erreur est liée à la difficulté d'estimer un coût comme une perte de réputation. La figure 1.15 présente les différents types de coûts avec leur incertitude et importance.

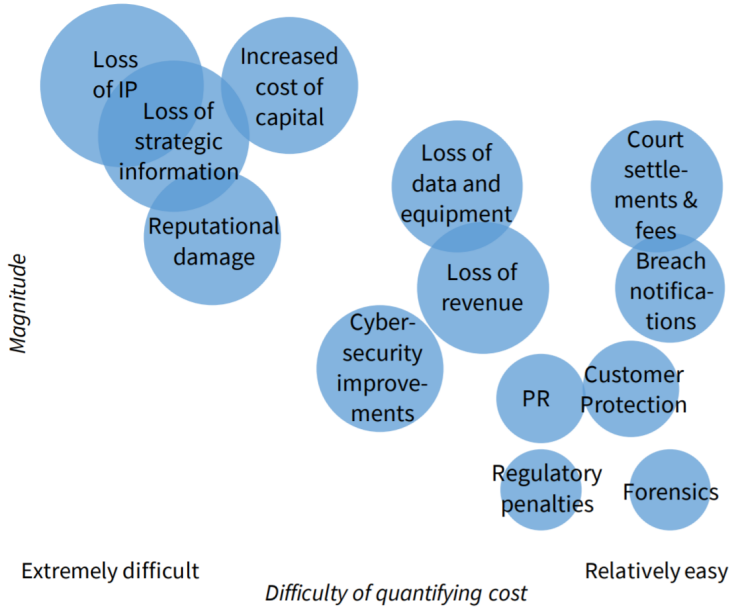


FIGURE 1.15 – Coûts possibles suites à une cyber-attaque
source : The Cost of Malicious Cyber Activity to the U.S. Economy – 2018

Regardons quelques unes des attaques les plus usuelles.

Les logiciels malveillants (*malwares*)

De nombreux programmes, programmes craqués, greffons et autres extensions sont librement téléchargeables et très attractifs mais malheureusement ils contiennent parfois des virus, cheval de Troie ou autre méchanceté qui agira seul ou permettra au pirate d'agir à distance sur la machine infectée, ordinateur ou ordiphone. La machine infectée peut aussi devenir un *zombie* que le pirate utilisera pour attaquer ailleurs ou envoyer du spam.

Lorsque le *malware* est inclus dans une pièces attachées on parle encore de *phishing*. Un exemple simple consiste à envoyer un fichier Excel ou Word en se faisant passer pour un collègue. Le simple fait d'ouvrir la pièce attachée peut suffire à infecter sa machine.

Parmis les plus célèbres de ces malwares citons :

- NotPetya le destructeur conçu pour affaiblir l'Ukraine. Ses dégâts ont été évalué à 10 G$. Si l'Ukraine a subit la majorité des dégats, d'autres ont aussi été touchés comme l'entreprise de transport danoise Maersk qui a déclaré 300 M€ de pertes ou Saint-Gobain en France qui a évalué ses dégats à 80 M€.
- ILOVEYOU qui a touché des dizaines de millions de machines sous Windows en se propageant par le mail. Son coût a été estimé entre 5 et 9 G$ de dommages et 15 G$ de correctifs pour s'en protéger.
- MyDoom qui, comme ILOVEYOU, s'est propagé par le mail sur les machines Windows. En 2004 entre 16 et 25% des mails étaient infectés par ce virus. Le but de ce virus semble avoir été de créer des zombies pour relayer du spam. Son coût global a été estimé à 38 G$,

— Stuxnet un virus israélo-américain conçu pour détruire du matériel du programme nucléaire iranien (cf chapitre sur la cyber-guerre).

Pour ce protéger de ces logiciels malveillants il faut développer une bonne hygiène informatique.

La première règle est de maintenir sa machine à jour en installant toutes les mises à jour au fur et à mesure qu'elles sortent. Même si ainsi la sécurité n'est pas totale, elle est souvent suffisante, les pirates allant vers les proies les plus faciles à savoir les machines pas à jour. Bien sûr lorsqu'un virus utilise une faille pas connue, tout le monde est nu jusqu'à l'arrivée du correctif.

La seconde règle consiste à faire attention aux logiciels qu'on installe et à supprimer ceux qu'on utilise plus. Ceci est particulièrement vrai sur les ordiphones. Ainsi il n'est pas normal qu'une application de type minuteur demande le droit d'accéder au carnet d'adresse, aux paramètres du réseau ou je ne sais quoi. Une telle application n'a besoin d'aucun droit spécifique. Si elle en demande il y a danger. Le danger ne peut être que commercial et toucher la vie privée de l'utilisateur mais il peut aussi être bien plus grave.

Il est aussi possible d'utiliser un système d'exploitation et des logiciels qui n'intéressent pas les cyber-criminels car trop peu utilisés. Ainsi un système comme Linux ou FreeBSD est bien moins attaqué que Windows ou MacOS. Pour de nombreux experts Linux et FreeBSD sont aussi intrinsèquement plus sûr car ouverts ce qui permet un audit permanent et une correction plus rapide des failles. L'agence de la sécurité française, l'ANSSI va dans ce sens et propose un système d'exploitation très sécurisé basé sur Linux : Clip OS.

Les demandes de rançon (*ransomware*)

Certain virus chiffrent tous les fichiers de la machine infectées et demande ensuite au propriétaire de payer une rançon pour que ses données soient déchiffrées.

L'un des plus connus, WannaCry, a touché plus de 300 000 ordinateurs en 2017. Il a utilisé la faille de sécurité de Windows en s'appuyant sur l'exploit EternalBlue développée par la NSA [24] pour son usage personnel mais qui a fuité ! Notons qu'il n'a touché que les machines pas à jour ou les vieux systèmes d'exploitation, comme Windows XP, que Microsoft ne mainteanait plus. Des hôpitaux, des ministères, des villes, des entreprises ont été affectés par WannaCry.

Le principe de chiffrer les données pour demander une rançon ne nécessite pas obligatoirement l'utilisation d'un virus mais c'est quand même bien pratique.

Les attaques web

L'importance du Web est telle que pour certains Internet est le Web. Cela implique que toute organisation a un site web qui va d'une simple présentation à un site marchand qui est le cœur de l'entreprise. Aussi pouvoir pénétrer un serveur web intéresse de nombreuses personnes. Des activistes cassent des sites web d'ennemis ou les détournent pour y afficher leur message. Les Anonymous sont coutumiers du fait, ISIS a fait de même contre TV5Monde via des hacker russes

24. L'agence de sécurité informatique américaine, cf chapitre sur la démocratie

a priori. Parfois ce sont des États qui vont bloquer des sites [25]. Le plus souvent ce sont des cybercriminels qui attaquent pour faire payer les organisations victimes.

Les injections sont la méthode la plus utilisée pour casser un serveur web.

Vol de données (*data-breach*)

Nous laissons nos données partout, dans les banques, les hôtels, les transports, les réseaux sociaux, les sites spécialisés… et ces données ont de la valeur. Elles permettent d'utiliser une identité pour créer de faux vrais papiers, elles permettent aussi d'agir sur Internet en notre nom, elles permettent parfois d'accéder à des comptes bancaires pour se servir, de connaitre notre historique, nos relations pour faire du chantage…

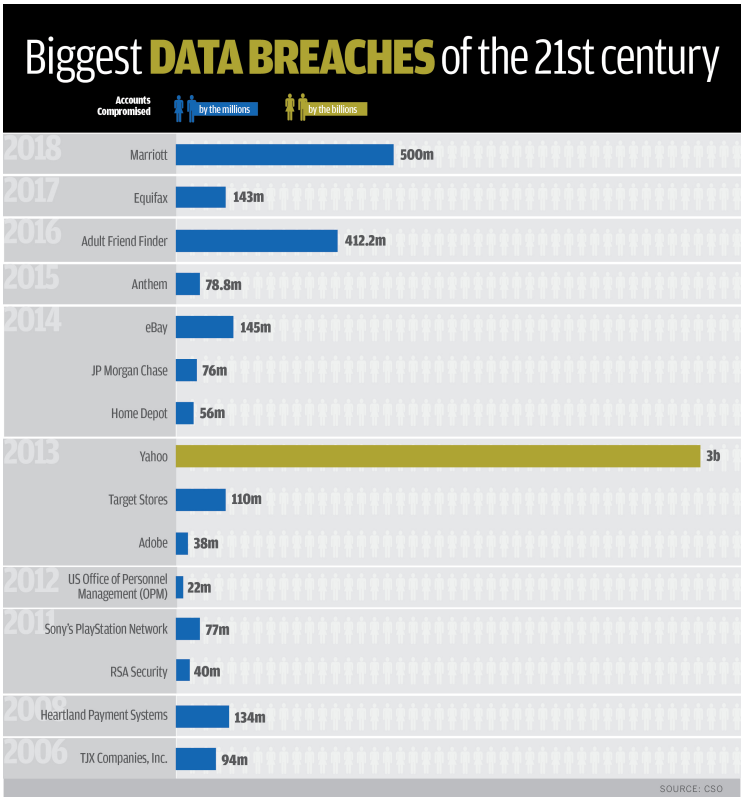


Figure 1.16 – Nombres connus de données volés entre 2006 et 2018

La liste des vols les plus importants, figure 1.16, ne doit pas faire oublier les moins connus nettement plus nombreux.

Le premier de la liste concerne les hôtels Marriott. D'après la dernière annonce en 2019, 383 millions d'identité ont été volées avec 25 millions de numéros de passeport (dont 80% était chiffrés), 9 millions de carte de crédit chiffrées mais 385 000 cartes de crédit pas chiffrées. Ces chiffres sont à prendre avec précaution car l'intrusion dans le système informatique de Marriott date de 2014.

25. La Russie l'a fait en Estonie et en Géorgie.

Aussi pendant plus de 4 ans les pirates voyaient tout.

Le dénis de service

Il s'agit d'une attaque spéciale puisqu'elle ne nécessite pas de casser la sécurité d'un autre. Il s'agit simplement de faire une tentative de connexion à une autre machine et de la répéter. Cela peut être charger une page web et la recharger toutes les secondes. Bien sûr si seulement une personne fait cela, cela n'a aucun impact sur la machine visée, mais si des millions voire des milliards de requêtes ont lieu en même temps [26], la machine visée passe en surcharge et ne peut plus répondre. Elle devient inaccessible, coupée de l'Internet, ce qui est le but.

On comprend que le succès dépend du le nombre d'attaques simultanés. Aussi pour mettre toutes leurs chances de leur coté, les attaquants utilisent des machines zombies dont ils ont pris le contrôle par le passé (via des virus par exemple) afin d'avoir une force de frappe conséquente.

En 2016 des jeunes joueurs de Minecraft ont piratés des caméras IP et d'autres objets de l'Internet pour lancer des dénis de service distribués contre des serveurs Minecraft concurrents (attaque nomée Mirai). Le piratage a été très simple puisqu'ils ont simplement utilisé les login et mot de passe par défaut de ces objets de l'Internet [27]. Ainsi ils ont pu lancer des attaques d'1 Tbits/s contre le réseau d'OVH qui hébergeait des serveurs Minecraft concurents ainsi que sur d'autres réseaux. L'importance de l'attaque, nettement plus forte que le pic de l'attaque de 2007 contre l'Estonie [28], a laissé penser à une attaque étatique initialement.

Le filoutage (*Phishing*)

Le filoutage ou phishing consiste à récupérer des informations personnelles que l'attaquant pourra exploiter ensuite. Il peut s'agir d'un login/mot de passe, d'un numéro de carte bleue ou même des données a priori moins sensibles qui permettront une usurpation d'identité. Le processus consiste le plus souvent à envoyer un mail alarmant demandant au destinataire de suivre un lien pour se protéger. Cela peut être votre soi-disant banque qui vous demande de changer votre mot de passe mais le lien envoie sur une copie du site de la banque. En 2019, le directeur exécutif du groupe Orange a estimé qu'environ deux millions de français sont victimes chaque année du phishing.

L'exemple qui suit demande aux propriétaires d'un nom de domaine chez Enom de se connecter sur leur compte pour valider les informations les concernant sous peine de perdre leur nom de domaine.

```
Date: Sat, 1 Nov 2008 10:56:39 +0100
From: eNomCentral Team <support@enom.com>
To: olivier@ricou.eu.org
Subject: Inaccurate whois information.

Dear user,
```

26. on parle alors de DDOS pour Distributed Deny of Service
27. Comme quoi ne pas changer un mot de passe par défaut peut rendre complice d'une cyber-attaque…
28. cf chapitre sur la cyber-guerre

```
On Sat, 1 Nov 2008 10:56:39 +0100 we received a third party complaint of
invalid domain contact information in the Whois database for this domain.
Whenever we receive a complaint, we are required by ICANN regulations to
initiate an investigation as to whether the contact data displaying in the
Whois database is valid data or not. If we find that there is invalid or
missing data, we contact both the registrant and the account holder and
inform them to update the information.

...
PLEASE VERIFY YOUR CONTACT INFORMATION - http://www.enom.com.ssl48.mobi
LINK TO CHANGE INFORMATION - http://www.enom.com.ssl42.mobi

Thank you,
Domain Services
```

Bien sûr, le lien donné est un faux qui ne renvoit pas chez Enom, `www.enom.com`, mais sur `www.enom.com.ssl48.mobi`, site qui appartient à celui qui contrôle `ss148.info`. Si l'on suit ce faux lien, on tombe sur une page identique d'aspect à la page d'authentification du site d'Enom et si l'on entre son login/mot de passe, on s'est fait avoir. Ainsi le pirate récupère le contrôle du nom de domaine ce qui lui permet d'intercepter de l'information et rediriger des requêtes. S'il le fait discrètement, le propriétaire du domaine ne s'en rendra pas compte.

Lorsqu'on craint d'être la victime d'une telle attaque, il est conseillé de contacter directement et par la voie usuelle l'entreprise concernée. Ainsi, dans notre cas, en allant sur la page d'accueil d'Enom, la véritable : `www.enom.com`, on sait immédiatement à quoi s'en tenir, le message suivant confirmant l'arnaque :

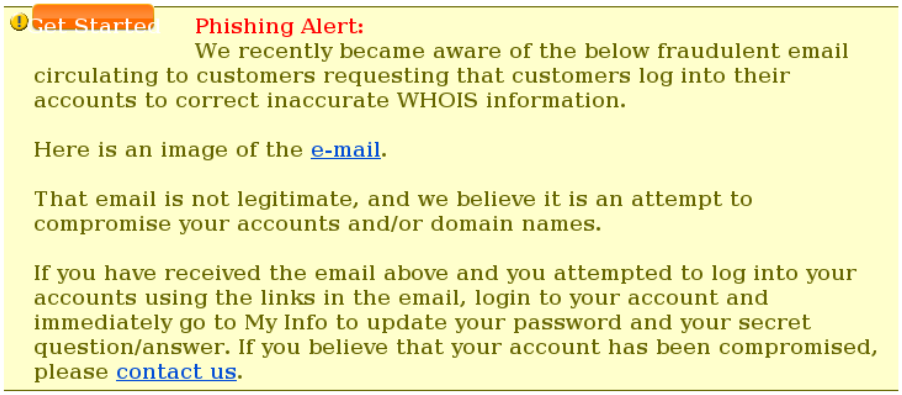


FIGURE 1.17 – Message d'avertissement d'Enom contre une arnaque

La règle numéro 1 pour se protéger du filoutage est de ne jamais cliquer sur un lien compris dans un mail. Si on veut quand même cliquer, c'est de vérifier précautionneusement l'URL de la page web une fois qu'on a cliqué.

m.à.j. sur http://www.ricou.eu.org/e-politique.html

L'arnaque

Si cette attaque n'est pas dans la liste elle n'en est pas moins un type d'attaque très utilisé. Il s'agit de convaincre la victime de verser de l'argent à l'arnaqueur.

L'arnaque nigériane a été et reste un grand classique. Elle consiste le plus souvent à demander de l'aide pour sortir des sommes considérables de son pays en échange d'un pourcentage conséquent. Pour cela il va falloir ouvrir un compte bancaire au Nigéria, y déposer une somme minimale puis d'autres excuses permettront de demander d'autres sommes supplémentaires. Les quelques personnes arnaquées qui ont été sur place pour réclamer leurs biens ont souvent fini à la morgue. Notons que parfois c'est le gros lot pour l'arnaqueur puisque le FBI avait arrêté la comptable d'un cabinet d'avocats américains qui avait versé plus d'un millions à ses arnaqueurs.

Mais les arnaques ne se font pas que par mail. De nombreuses personnes ont perdu des sommes importantes en pensant avoir rencontré l'amour de leur vie sur des sites de rencontre puis en les aidant financièrement à sortir d'une mauvaise passe pour les rejoindre. Ce type d'arnaque effectué le plus souvent par des étrangers (c'est moins dangeureux) s'est sophistiqué avec temps et aujourd'hui l'arnacœur dira vivre dans une ville française dont il sait tout grâce à Internet. Mais il aura toujours des problèmes qui coûtent bien cher.

Les réseaux sociaux sont aussi des endroits parfait pour les arnaqueurs. Enfin d'autres types d'arnaques existent, cf figure 1.18.

Investment	$8,648	Fake Invoice	$441
Romance	$6,003	Credit Repair/Debt Relief	$388
Moving	$3,993	Online Purchase	$365
Cryptocurrency*	$3,147	Fake Check/Money Order	$341
Home Improvement	$2,895	Tech Support	$255
Nigerian/Foreign Money Exchange	$2,133	Credit Card	$231
Business Email Compromise	$1,717	Government Grant	$218
Family/Friend Emergency	$1,219	Health Care/Medicaid/Medicare	$170
Counterfeit Product	$1,210	Scholarship	$155
Travel/Vacation	$887	Utility	$106
Advance Fee Loan	$716	Debt Collection	$98
Charity	$708	Yellow Pages/Directory	$91
Identity Theft	$683	Phishing	$44
Rental	$662	Tax Collection	$31
Employment	$598		
Sweepstakes/Lottery/Prize	$547	Other	$746

*Denotes a category first tracked in 2018

Figure 1.18 – Somme perdue en moyenne par type d'arnaque
source : BBB Scam Tracker – 2015-2018

Les particuliers ne sont pas les seules victimes des arnaqueurs. Les entreprises subissent depuis quelques années l'arnaque au faux virement. Elle consiste à envoyer un faux mail au nom du patron avec une demande de virement urgente pour conclure une affaire. Bien sûr de nombreux comptables ont fait le virement sans vérifier l'authenticité de l'émetteur du message. Notons que souvent l'affaire est bien préparée avec une bonne connaissance de l'entreprise de la part des arnaqueurs, et pour cause, les sommes en jeu sont nettement plus importantes. Le préjudice a été estimé à 485 M€ entre 2010 et 2018 pour les entreprises françaises et 2.3 G$ pour les États-Unis [29].

29. https://www.lesechos.fr/idees-debats/cercle/les-arnaques-au-virement-concernent-toutes-les-entreprises-130970

Pour éviter les arnaques il faut prendre le temps de la réflexion (la moindre chose bizarre dans le message est un indice d'arnaque possible), en parler à des proches et regarder des sites qui référencent les arnaques comme https://info.signal-arnaques.com/.

1.3 La cryptographie

La cryptographie protège les communications dès lors que votre machine n'est pas infectée, que votre logiciel n'a pas de bug, que vous ne donnez pas vos clés ou mot de passe au pirate... Elle permet

- de chiffrer les données,
- d'en garantir l'intégrité,
- de signer le message.

Le premier point implique

- la confidentialité des communications (transactions bancaires, connexions à distance, téléphone, mail...),
- la protection de données informatique stockées (secrets militaires, industriels, commerciaux, médicaux, personnels...)

Le second point, la garantie de l'intégrité, offre la certitude qu'un document est complet (contrat, mail, logiciel...) et que personne n'a pu le modifier.

Enfin le dernier point, la signature, permet

- de savoir avec certitude qui est l'origine d'un document item et inversement de prouver qu'on est l'auteur du document
- la non-répudiation,
- de protéger des systèmes informatiques contre les intrusions en vérifiant l'identité des machines et utilisateurs,
- de vérifier l'authenticité d'un site Web

Avec la combinaison des trois, on peut envoyer un mail en étant certain que personne d'autre

que mon destinataire ne pourra le lire (chiffrage). Le destinataire aura la certitude que mail vient bien de l'émetteur grace à la signature et qu'il n'a pas été modifié (intégrité). Ainsi l'émetteur ne pourra pas contester le fait d'avoir écrit le message (non-répudiation).

Avant de regarder l'utilisation de la cryptographie pour se protéger sur Internet, essayons de comprendre les principes de la cryptographie.

1.3.1 La théorie

Les clés symétriques ou secrètes

La façon la plus simple de chiffrer un message est de lui appliquer une fonction mathématique. Ainsi Jules César chiffrait ses messages en décalant les lettres de N, ainsi avec N=3, le A devient D. Pour le déchiffrer il suffit d'appliquer la fonction inverse avec la même clé. Bien sûr un bon système de cryptographie propose une fonction inverse assez compliqué pour qu'on ne puisse pas deviner le message sans la clé (N dans le cas de Jules César). Ce système est celui de la clé symétrique.

ATTAQUEZ GERGOVIE	DWWDTXHC JHUJRYLH
↓ +3	↓ -3
DWWDTXHC JHUJRYLH	ATTAQUEZ GERGOVIE

Figure 1.19 – Un message secret de Jules César

Pour communiquer entre 2 ou 3 personnes il suffit d'avoir une clé commune pour pouvoir communiquer de façon protégée par la suite. Bien sûr plus il y a de personnes qui partagent la clé, plus les risque de fuite sont importants. Pour éviter cela, on peut choisir de créer une clé par paire de personnes, soit $N^2/2$ clés pour un groupe de N personnes ce qui est rapidement ingérable.

Aussi pour un grand groupe on peut préférer le système dit de tiers de confiance, TDC, (Trusted Third Party en anglais, ou TTP) qui propose de définir une seule clé K_i pour chaque utilisateur qui lui permet de communiquer avec le tiers de confiance.

Lorsque deux personnes, i et j, veulent communiquer, le TDC génère une clé de session k qu'il transmet chiffrée à i avec la clé K_i et à j avec la clé K_j. Puis les utilisateurs utilisent la clé de session k pour communiquer, cf figure 1.20.

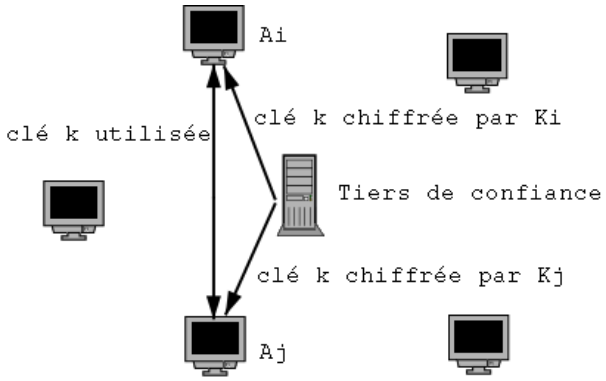


FIGURE 1.20 – Tiers de confiance pour chiffrement symétrique

Ce système de clés symétriques a les avantages suivants :

- il est facile d'ajouter un nouvel entrant dans le réseau,
- chaque individu ne stocke que sa clé de communication avec le TDC,
- le chiffrement et déchiffrement sont rapides.

Les inconvénients sont :

- le besoin du TDC pour initier toute communication,
- le TDC peut lire tous les messages.

On comprend que la présence du tiers de confiance peut être jugée problématique.

Les clés asymétriques ou publiques

Le système de cryptographie par clé symétrique a été le seul disponible jusqu'après la seconde guerre mondiale, ce qui veut dire que durant la seconde guerre mondiale les clés utilisées devaient être transmises physiquement à travers les théatres d'opération avec tous les risques d'interception possibles lorsqu'on doit traverser les lignes ennemies. Lorsqu'on veut renouveler les clés régulièrement au cas où l'ennemi aurait réussi à les avoir, on en veut à la technologie qui impose cet exercice délicat.

La clé asymétrique corrige ce défaut en permettant de transmettre une clé publiquement pour chiffrer tout en gardant une clé privée pour déchiffrer. Les messages qu'on reçoit et que tout le monde peut intercepter, sont chiffrés avec la clé diffusée publiquement mais seule la clé privée peut les déchiffrer.

En pratique un utilisateur génére sa clé privée d_i et publique e_i puis diffuse cette dernière ce qui permet à quiconque de lui envoyer un message sans risque d'interception. On peut imaginer un répertoire public où chacun dépose sa clé publique. Ainsi le message m est chiffré par la fonction E qui utilise la clé publique du destinataire, puis déchiffré par la fonction D à l'aide de la clé privée du destinataire, cf figure 1.21.

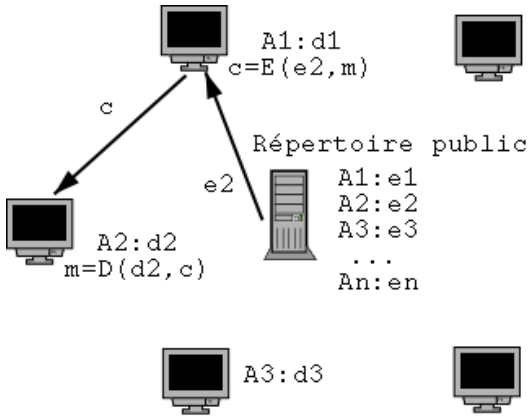


Figure 1.21 – Utilisation de clés asymétriques

Les avantages de la méthode sont

- l'absence d'intermédiaire, pas de TDC,
- le fichier des clés publiques peut être largement diffusé.

Les inconvénients sont :

- un pirate peut diffuser une fausse clé publique (cf ci-dessous),
- le chiffrement est plus lent qu'avec une clé symétrique.

L'attaque de l'homme au milieu L'attaque la plus simple contre ce système est de substituer la clé publique d'un utilisateur par celle du pirate et d'intercepter tous les messages. Une fois le message intercepté, le pirate, l'homme au milieu, le déchiffre, le note, puis le chiffre avec la véritable clé publique du destinataire pour lui envoyer afin qu'il ne détecte pas l'interception.

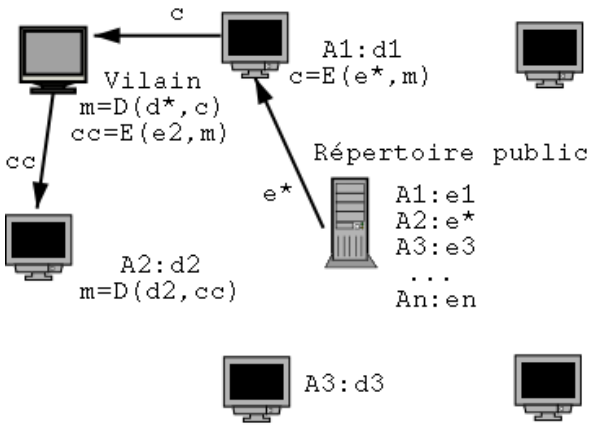


Figure 1.22 – Attaque de l'homme au milieu

La parade, pour ne pas voir son message intercepté, réside dans la fiabilité de la clé publique de

son destinataire. Une clé publique est sûre, soit parce que le destinataire vous l'a remise en main propre, soit parce qu'une personne en qui vous avez entièrement confiance vous garantit cette clé publique. Cette personne de confiance peut être une autorité de certification (cf section 1.3.3) ou une personne dont vous êtes sûr car elle est dans votre liste des personnes de confiance. Dans ce dernier cas on parle de votre toile de confiance ou *Web of trust* [30].

Différents algorithmes de cryptographie

Sans remonter jusqu'à Jules César, il existe de nombreux algorithmes de cryptographie. Certain sont plus connus que d'autres et leur célébrité est la garantie de leur fiabilité. En effet il est difficile de créer un algorithme de cryptographie solide et seule sa vérification par le plus grand nombre possible de mathématiciens et d'utilisateurs peut offrir une garantie de sécurité.

DES, Triple DES et AES Historiquement DES, Data Encryption Standard, est le premier standard officiel des États-Unis à destination des entreprises. Il s'agit d'un algorithme de chiffrement à clé symétrique développé par IBM dans les années 70. DES utilise une clé de 56 bits qui, de nos jours, est bien trop faible pour résister aux attaques. Aussi DES ne doit plus être utilisé.

Son premier remplaçant a été Triple DES qui n'est que l'application de DES trois fois avec des clés différentes. Cela permet en effet d'amener la sécurité à un niveau correct mais pour un coût élevé en temps de calcul.

Aussi à la fin des années 90, le gouvernement américain à lancé un concours pour trouver le remplaçant idéal, sûr et peu gourmand en CPU afin de pouvoir l'exécuter sur le processeur d'une carte à puce. En 2001 le vainqueur a été déclaré, l'algorithme symétrique Rijndael [31] a été choisi pour être l'Advanced Encryption Standard (AES).

Les Rivest Cipher et RSA Ronald Rivest est un cryptologue qui a conçu de nombreux algorithmes de chiffrements symétriques dit à la volée ("stream cipher" – RC4) et par bloc ("block cipher" – RC2 / RC5 / RC6). Parmis ces algorithmes, RC4 est le seul de la famille à être propriétaire ("trade secret") mais son code a été largement diffusé. RC6 était un des 5 candidats retenus à AES.

Mais l'heure de gloire [32] est arrivée avec RSA [33]. Cet algorithme conçut en 1977 avec Adi Shamir et Len Adleman est le premier algorithme publié [34] à clé publique/clé privée (ou asymétrique). Il est toujours très utilisé. Son principe mathématique est expliqué dans l'encart page 43.

Les condensats ou empreintes Un condensat permet de garantir l'intégrité d'un document. Il s'agit du résultat d'une fonction à sens unique, dite de hachage, qui résume un document en

30. cf https://fr.wikipedia.org/wiki/Toile_de_confiance
31. cf la BD qui présente Rijndael, http://www.moserware.com/2009/09/stick-figure-guide-to-advanced.html
32. Prix Turing 2002, le Nobel des informaticiens
33. Les initiales de ses inventeurs, Riverst, Shamir et Adleman.
34. L'armée anglaise avait trouvé quelques années auparavant un algorithme asymétrique mais bien sûr, elle s'était bien gardée de l'annoncer

Les mathématiques de RSA

L'algorithme RSA est un algorithme de chiffrement asymétrique.

L'idée d'un algorithme asymétrique a été proposée par Whitfield Diffie et Martin Hellman dans un article en 1975 et mise en pratique en 1977 par Ronald Rivest, Adi Shamir et Leonard Adleman. James Ellis et Clifford Cocks des services de communication de l'armée anglaise, avaient trouvé cet algorithme quelques années plus tôt mais ne purent le dévoiler pour cause de secret militaire (cf l'histoire présentée par Ellis).

Son principe est relativement simple mais totalement révolutionnaire. On n'imaginait pas jusque là qu'il puisse être possible de décoder un message sans avoir la clé ayant permis de l'encoder.

Pour cela chaque utilisateur a

- une clé publique (e, n)
- une clé privée (d, n)

avec les nombre e, d et n construits ainsi :

		Exemple
1	Choisir 2 nombres premiers distincts p et q. Plus ils sont grands et meilleure sera la sécurité.	Prenons $p = 11$ et $q = 19$.
2	Soit $n = p\,q$.	$n = 11 \times 19 = 209$
3	Choisir un nombre d premier avec $(p-1)\,(q-1)$.	Dans notre cas $(p-1)\,(q-1) = 180 = 2^2 \times 3^2 \times 5$ donc $d = 7$ marche.
4	Choisir e tel que $(e\,d)\,\%\,((p-1)(q-1)) = 1$.	$e = 103$ vérifie $7e\,\%\,180 = 1$ puique $7 \times 103 = 721 = 4 \times 180 + 1$.

Ces choix impliquent que $e\,d\,\%\,J(n) = 1$ où $J(n)$ est l'indicatrice d'Euler sachant que $J(n) = (p-1)(q-1)$ lorsque p et q sont premiers. C'est la propriété magique qui permet à RSA de fonctionner.

Chiffrer un message On chiffre le message M à l'aide de la clé publique (e, n) du destinataire ainsi (tout n'est que 0 et 1 sur un ordinateur donc tout message est un nombre) :

	Exemple avec $M = 123$
$M' = M^e\,\%\,n$	$M' = 123^{103}\,\%\,209 = 63$

Message chiffré qu'il peut déchiffrer avec sa clé privée (d, n) car

$M'^d\,\%\,n = (M^e\,\%\,n)^d\,\%\,n$ $= M^{ed}\,\%\,n = M^{ed\,\%\,J(n)} = M$	$M'^d\,\%\,n = 63^7\,\%\,209 = 123 = M$

Prouver son identité L'émetteur chiffre le condensat $C(M)$ d'un message M avec sa clé privée (d, n) et l'envoie avec le message :

$$C' = C(M)^d\,\%\,n$$

Pour être certain que le message vient bien de l'émetteur il suffit de comparer $C(M)$ et $C'^e\,\%\,n$ avec (e, n) la clé publique de l'émetteur. S'ils sont égaux c'est bon.

Casser RSA Si on peut décomposer n en p et q alors trouver d est simple sachant que l'on connait e. Heureusement décomposer un très grand nombre en nombres premiers est une opération très lourde qui peut prendre des siècles pour un n de bonne taille (sauf pour les futurs ordinateurs quantiques).

une ligne. Cette fonction est telle que si l'on modifie quoi que ce soit dans le document, alors le condensat devient totalement différent. Les condensats MD5 et SHA-1 ayant été cassés, ce qui rend possible la génération d'un autre document qui produit le même condensat, seule la famille des SHA-2 restait sûre. Aussi le concours SHA-3, a été lancé pour définir une nouvelle fonction de hachage sûre. En 2012 l'algorithme Keccak a été choisi comme nouvelle norme.

```
md5("Le condensat garantit l'intégrité") = 9fb6e5c02fd664892271ca02e0266457
md5("Le condensat garantit l'intégrite") = d80c680cf92d64cb7830c86fbb2350f7
```

Seul le é final a changé mais le condensat est totalement différent.

FIGURE 1.23 – Utilisation d'un condensat

1.3.2 Utilisation de la cryptographie

Protéger son courrier avec GPG

Comme on l'a vu, le courrier est particulièrement vulnérable et la seule façon de le protéger nécessite l'usage de la cryptographie. Actuellement il existe deux principaux logiciels pour chiffrer les mails : GPG, GNU Privacy Guard, et S/Mime. Tous les deux utilisent différents algorithmes de cryptographie pour remplir toutes les conditions nécessaires à la protection du courrier :

- un algorithme de chiffrement symétrique de type AES par défaut pour chiffrer la session,
- un algorithme de chiffrement asymétrique de type RSA, DH ou DSA pour chiffrer la clé de session et signer,
- un condensat comme SHA-256 ou SHA-3 pour vérifier l'intégrité.

Pour des raisons de performance, les messages sont donc chiffrés à l'aide d'un système à clé symétrique dite clé de session. Cette clé est elle-même chiffrée avec la clé publique du destinataire, ainsi lui seul pourra la récupérer avec sa clé privée et donc lire le message.

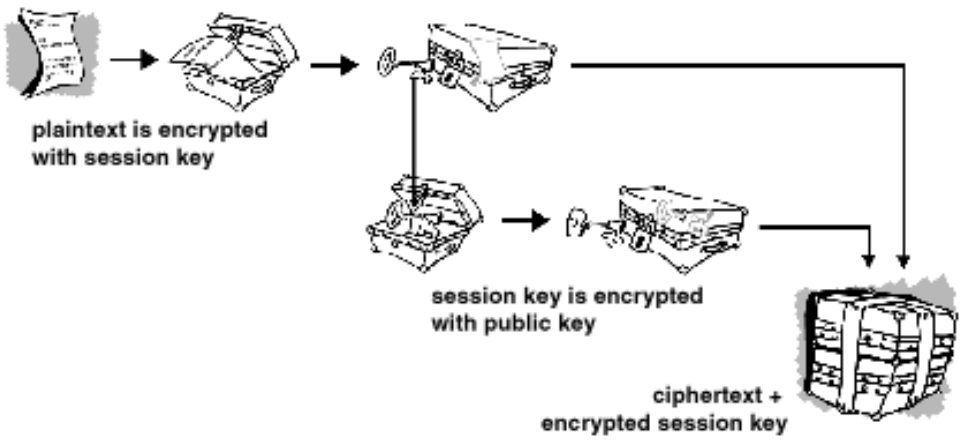


FIGURE 1.24 – Encodage d'un message à l'aide de GPG

La carte bleue cassée

Le 4 mars 2000, le texte suivant tombait dans le forum Usenet `fr.misc.cryptologie` :

```
Petite feuille Maple

> pub:=2^320+convert(`90b8aaa8de358e7782e81c7723653be644f7dcc6f816daf46e532b91e84f`,
                     decimal,hex);
pub := 213598703592091008239502270499962879705109534182641740644252416500858395774644508841...

> facteur1:=convert(`c31f7084b75c502caa4d19eb137482aa4cd57aab`, decimal, hex);
facteur1 := 1113954325148827987925490175477024844070922844843

> facteur2:=convert(`14fdeda70ce801d9a43289fb8b2e3b447fa4e08ed`, decimal, hex);
facteur2 := 191748170252450443937578626823086218069693418929 3

> produit:=facteur1*facteur2;
produit := 213598703592091008239502270499962879705109534182641740644252416500858395774644 5...

> exposant_public:=3;
exposant_public := 3

> modulo_div_eucl:=(facteur1-1)*(facteur2-1);
modulo_div_eucl := 213598703592091008239502270499962879705109534182338597041485083258128 26...

> essai_rate_exposant_prive:=expand((1+modulo_div_eucl)/3);
essai_rate_exposant_prive := 213598703592091008239502270499962879705109534182338597041485 0...

> exposant_prive:=expand((1+2*modulo_div_eucl)/3);
exposant_prive := 142399135728060672159668180333308586470073022788225731360990055505418845...

> testnb:=1234;
testnb := 1234

> testsignnb:=testnb &^ exposant_prive mod produit;
testsignnb := 223593814777518377564104232545040455789953214462648171523669429097480691923 4...

> testverifsignnb:=testsignnb &^exposant_public mod produit;
testverifsignnb := 1234
```

On y trouve les nombres premiers p et q, ici `facteur 1` et `facteur 2` qui permettent de connaître le module n, ici `produit`. On voit que l'exposant publique, e, est 3 et après un premier test raté on trouve l'exposant privé d. Pour être sûr que tous ces chiffres sont bons, on chiffre 1234 et on le déchiffre. Ça marche.

Ce jour là le grand public voyait en clair la clé RSA à 320 bits qui permet de vérifier l'authenticité d'une carte bleue (voir l'article de Louis Guillou) . Cela indique seulement qu'une carte est authentique et non que l'on connaît le code secret de l'utilisateur, mais cela permet de faire des fausses cartes [a] qui tromperont un lecteur non relié aux banques comme celui qu'on présente souvent dans les restaurants.

C'est cette faiblesse connue des milieux de la cryptographie qu'a utilisé Serge Humpich [b]. La trouvaille n'est pas extraordinaire car casser une clé de 320 bits n'était plus un exploit depuis le début des années 90. L'exploit réside surtout dans la légèreté du groupement des cartes bleues qui a pris 10 ans pour corriger une faille connue.

a. faire une fausse carte bleue est assimilé à faire de la fausse monnaie. Le tarif est 30 ans de prison.
b. cf http://fr.wikipedia.org/wiki/Serge_Humpich

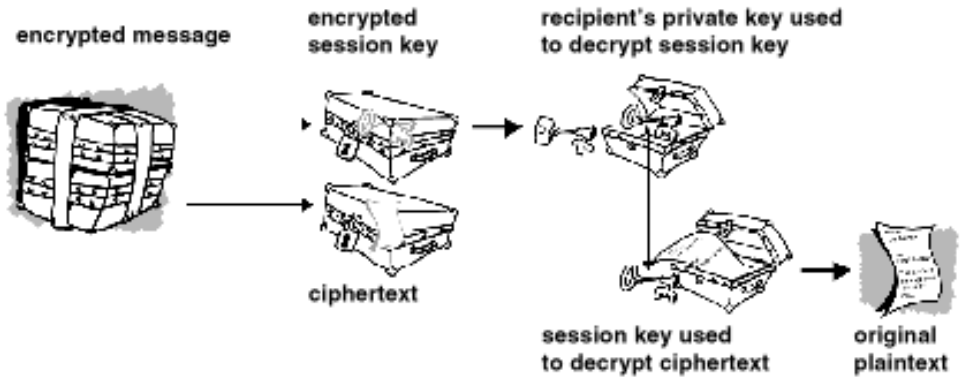


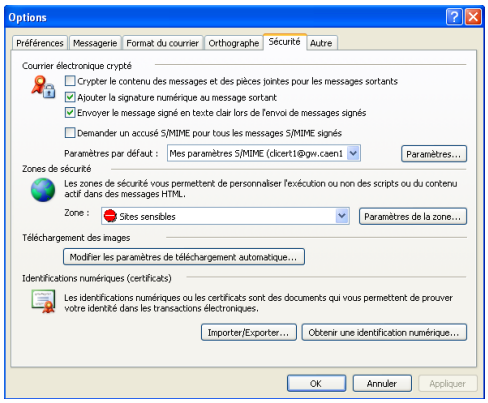
FIGURE 1.25 – Décodage d'un message à l'aide de GPG

Pour signer et vérifier l'intégrité du courrier, l'émetteur fait un condensat du courrier et le chiffre avec sa clé publique. Ainsi le destinataire peut générer le condensat du courrier déchiffré et le comparer avec le condensat que lui a envoyé l'émetteur après l'avoir déchiffré avec la clé publique de l'émetteur.

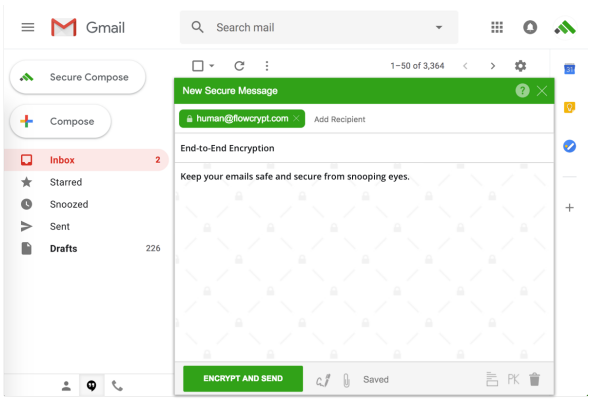
Lorsque GPG est inclu dans votre logiciel de mail, son utilisation est transparente. Son initialisation peut faire peur pour celui qui ne connait rien à la cryptographie puisque qu'on va lui demander de protéger sa clé privée avec un mot de passe et de publier sa clé publique. La publication de la clé publique est la partie la plus sensible puisque mal faite, elle peut permettre l'attaque de l'"homme au milieu", cf page 41. Il est donc soit nécessaire de la transmettre main dans la main [35], soit de la faire signer par une connaissance dont on a déjà la clé publique de façon sûre et ainsi agrandir sont réseau de confiance. On pourrait envisager de faire signer sa clé par une autorité de certification mais je ne l'ai jamais vu faire (trop lourd, trop cher ?).

En pratique S/MIME ou GPG sont de plus en plus intégrés dans les lecteurs de courrier mais il peut être quand même nécessaire de créer ses clefs soi même, cf ce site d'autodéfense courriel. Pour les webmails [36] c'est plus rare mais il existe des greffons pour Gmail comme FlowCrypt ou Mailvelope. Notez que le mode confidentiel de Gmail (*Gmail Confidential Mode*) n'est pas une protection acceptable car elle ne protège pas contre Google ni contre les états qui ont autorité sur Google ou d'un piratage des serveurs de Gmail [37].

35. transmettre le condensat de la clé publique est souvent plus simple et permet ensuite de récupérer la clé sur Internet puis de vérifier qu'elle est la bonne.
36. Le mail qu'on lit avec son navigateur.
37. lire aussi la réaction de l'EFF à ce sujet



Outlook et S/MIME



Gmail et FlowCrypt

Figure 1.26 – Lecteurs de mail et leur outil de cryptographie

Le Web sécurisé

Le Web est protégé par l'algorithme de chiffrage SSL [38] qui est présent sur les navigateurs les plus courants. Par contre, comme pour tout algorithme de chiffrage, son efficacité est directement liée à son utilisation et à la taille de la clé de codage utilisée.

Ainsi la majorité des pages web ne sont pas chiffrées et donc passent en clair sur le réseau avant d'arriver sur votre ordinateur. Cela veut dire que toute personne qui contrôle les machines intermédiaires peut savoir quelles sont les pages web que vous regardez.

Lorsque vous arrivez sur une page sécurisée, ce qui est visible par une icône en forme de clé ou de cadenas ainsi que dans l'URL qui commence par `https`, personne ne peut intercepter le contenu si la clé de chiffrement utilisée est dite forte, à savoir contient 256 bits [39]. Si par contre la clé est trop courte, comme 40 bits largement utilisé dans les années 90, alors la sécurité est illusoire car trop faible pour résister aux attaques brutales, type d'attaques qui essayent toutes les clés possibles. Avec un niveau de sécurité entre les deux comme 128 bits, encore très utilisé en 2018, vos communication sont protégées pour quelques années à savoir il faudra des années pour déchiffre le message. Si vous ne changez votre mot de passe pour vous connecter à votre banque que tous les 10 ans, vous prenez peut-être des risques.

Aussi n'hésitez pas à cliquer sur le petit cadenas pour vérifier si la protection utilisée est SSL 256 ou 128 bits (si vous trouvez du 40 bits, veuillez me l'indiquer svp !).

Il existe aussi un autre risque qui est celui de ne pas être connecté au véritable serveur mais à une copie comme dans le cas d'hameçonnage. Aussi il existe un système d'authentification du site web visité.

38. renomé TLS depuis 2001
39. un bit est 0 ou 1, 1001 est un nombre binaire à 4 bits qui vaut 9 en décimal.

1.3.3 Authentification et autorité de certification

L'authentification consiste à vérifier l'identité du correspondant.

Dans le cas d'un mail le champs `From` n'est pas suffisant car facilement falsifiable. Aussi il est nécessaire que le courrier soit signé par la clé privée de votre correspondant et que vous ayez sa clé publique. Bien sûr il faut être certain qu'il s'agit de sa clé publique et non pas d'une fausse. Comme il n'est pas toujours aisé de donner main dans la main cette clé ou son condensat, un autre système a été conçu : la certification.

La certification consiste à demander à un organisme reconnu d'offrir la garantie que le document [40] récupéré sur Internet est bien celui de notre correspondant. Pour cela l'organisme ajoute au document sa signature à l'aide de sa clé privée. Ainsi toute personne qui a la clé publique de l'organisme peut vérifier que la signature est bonne. On voit qu'on a seulement repoussé le problème puisque maintenant pour savoir si un document est le bon, il faut récupérer la clé publique de l'organisme.

La bonne nouvelle est que les autorités de certification sont des organismes reconnus aussi leur clés publiques sont présentes par défaut dans tous les ordinateurs. Ainsi la personne qui désire falsifier une clé publique doit maintenant commencer par trafiquer le système d'exploitation ou le navigatuer utilisé pour y mettre de fausses clés publiques d'autorité de certification. La tâche est nettement plus ardue.

Ce système est surtout utilisé pour le Web afin de garantir qu'un site appartient bien à celui qu'il déclare être. Dans ce cas il existe différents niveaux de certification. L'autorité peut vérifier seulement que le domaine appartient bien à celui qui demande le certificat ou aller plus loin en demandant des documents officiels. Cela se retrouve graphiquement dans certains navigateurs :

FIGURE 1.27 – Certificat basé sur le nom de domaine et certificat basé sur des papiers officiels

Les autorités de certification commerciales

De nombreuses entreprises sont des autorités de certification [41]. Elles bénéficient d'un marché très lucratif puisque signer la clé d'une personne est une opération dont le seul coût est la vérification de son identité. On retrouve l'une des nombreuses "poules aux œufs d'or" qui se promènent sur Internet [42]. Cela étant un trublion en la personne de Let's Encrypt perturbe sérieusement le marché depuis 2016 en offrant des certificats gratuits.

L'autorité de certification la plus importante a longtemps été Vérisign, la même entreprise que celle qui gère les `.com` et `.net`. Elle a acheté de nombreux concurrents comme Thawte et GeoTrust avant de céder en 2010 sa partie autorité de certification à Symantec [43] pour 1,28 milliards de

40. clé publique, certificats SSL ou autre
41. Pour se décréter autorité de certification, il suffit d'avoir une clé publique et de se faire connaitre
42. Dans la même veine que la gestion des noms de domaine.
43. L'entreprise d'anti-virus

dollars. Depuis c'est le déclin et plus aucune autorité n'a pu dépasser les 50% de part de marché.

Nom	Nombre de certificats	Part de marché (%)	Variation mensuelle (%)
COMODO CA Limited	1 158 223	31,65	0,85
DigiCert	541 108	14,79	0,60
Let's Encrypt	510 360	13,95	27,51
GlobalSign nv-sa	278 362	7,61	-1,79
GoDaddy,com Inc,	271 728	7,42	-1,00
cPanel, Inc	91 346	2,50	3,68
Unknown	84 514	2,31	-5,84
Amazon	69 770	1,91	-2,17
Google Trust Services	51 826	1,42	3,36
Starfield Technologies, Inc,	35 063	0,96	2,58
GeoTrust Inc,	34 682	0,95	-44,28

Table 1.3 – Classement des autorités de certification
source : Security space – nov. 2018

Les autorités de certification gouvernementales

La signature électronique étant reconnue par la loi en France, il semblerait normal que l'État certifie les signatures des citoyens après les avoir duement vérifiées comme il le fait pour les cartes d'identité. Malheureusement ce n'est pas le cas. L'État délaisse l'identité numérique au secteur privé et il n'est pas possible d'aller au commissariat de police avec sa clé publique et demander qu'elle soit certifiée.

Cela mène à des situations problématiques. Ainsi l'État a demandé aux entreprises de payer la TVA par Internet. Pour cela il leur demandait de justifier leur identité en présentant leur certificat numérique certifié par une autorité de certification. Et pour être bien clair, le ministère des finances indiquait dans sa FAQ sur la TéléTVA que

> *Les autorités de certification font autorité pour certifier les identités et principales caractéristiques des personnes à qui elles délivrent des certificats numériques. Elles jouent un peu le même rôle que les mairies lorsque vous faites une demande de passeport.*

et ajoute

> *(le) Ministère de l'Economie, des Finances et de l'Industrie qui en les référençant, reconnaît la qualité des procédures mises en œuvre dans l'identification des demandeurs, l'enregistrement et la délivrance des certificats. C'est la raison pour laquelle elles sont amenées à vous demander de nombreux justificatifs.*

On voit que dans ces deux cas où l'internaute veut ou doit justifier son identité, il ne peut le faire qu'en passant par une entreprise privée *qui sera amenée à demander de nombreux justificatifs*, justificatifs qu'un citoyen n'a peut-être pas envie de donner à une entreprise privée. Ce point est d'autant plus triste que la carte d'identité électronique nationale serait très utile sur Internet pour réduire les risques d'arnaques, diminuer le nombre de spam, communiquer avec l'administration,

Qui certifie ce site web ?

L'équivalent du champs `From:` pour identifier les sites web est leur adresse ou URL. On imagine que `www.lcl.fr` appartient à au Crédit Lyonnais (presque vrai, à sa maison mère) mais là encore il s'agit d'une information qui peut être trompeuse. Ainsi que penser de `www.lcl.net` ou `particuliers.secure-lcl.fr` ? Aussi le web dispose avec SSL d'un outil qui permet ce certifier qui est derrière un site web.

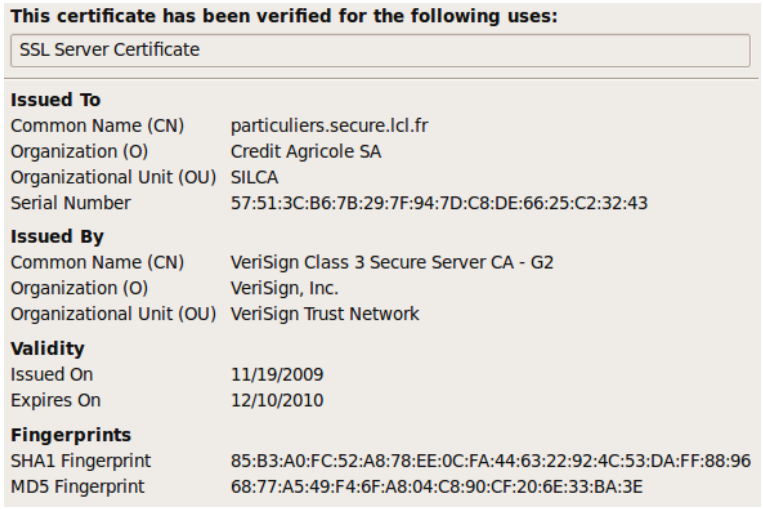


FIGURE 1.28 – Le certificat de `particuliers.secure.lcl.fr`

Là encore on se base sur la signature de la clé publique par une autorité de certification supérieure. Ainsi on peut voir dans le certificat du site du Crédit Lyonnais (lcl) qu'il est certifié par le Crédit Agricole, sa maison mère, qui elle-même est certifiée par Vérisign *Class 3 Secure Server* laquelle est certifiée par Vérisign *Class 3 Primary*. Enfin cette dernière est certifiée par elle-même, il faut bien s'arrêter quelque part.

```
% openssl s_client -connect particuliers.secure.lcl.fr:443
CONNECTED(00000003)
depth=2 /C=US/O=VeriSign, Inc./OU=Class 3 Public Primary Certification Authority
verify error:num=19:self signed certificate in certificate chain
verify return:0
---
Certificate chain
 0 s:/C=FR/ST=Hauts de Seine/L=La Defense/O=Credit Agricole SA/OU=SILCA/\
     CN=particuliers.secure.lcl.fr
   i:/C=US/O=VeriSign, Inc./OU=VeriSign Trust Network/OU=Terms of use at\
     https://www.verisign.com/rpa (c)05/CN=VeriSign Class 3 Secure Server CA
 1 s:/C=US/O=VeriSign, Inc./OU=VeriSign Trust Network/OU=Terms of use at\
     https://www.verisign.com/rpa (c)05/CN=VeriSign Class 3 Secure Server CA
   i:/C=US/O=VeriSign, Inc./OU=Class 3 Public Primary Certification Authority
 2 s:/C=US/O=VeriSign, Inc./OU=Class 3 Public Primary Certification Authority
   i:/C=US/O=VeriSign, Inc./OU=Class 3 Public Primary Certification Authority
---
```

vérifier l'âge des Internautes et peut-être un jour faire des débats en ligne à visage ouvert ou inventer une nouvelle forme de démocratie.

L'auto certification

Puisque la certification est nécessaire pour avoir un site web chiffré et que les autorités de certification étaient payantes avant l'arrivée de Let's Encrypt en 2015, de nombreuses personnes s'auto-certifiaient à savoir qu'elles signaient leur propre clé privée via leur autorité de certification crée pour l'occasion. Dans ce cas la clé ne sera pas reconnue puisque l'autorité n'est pas référencée mais cela permet néanmoins de chiffrer la communication entre le serveur et le navigateur. Malheureusement (ou heureusement) lorsque que le navigateur arrive sur une page web chiffrée par une clé auto-certifiée, il va bloquer la connexion tant que l'utilisateur ne lui indique pas explicitement de passer outre.

Le risque du certificat auto-certifié est qu'il est simple de lui appliquer l'attaque de l'homme au milieu. Ainsi votre FAI pourrait tout à fait générer un certificat auto-certifié qu'il vous présente chaque fois que vous vous connecté sur un site auto-certifié. Il pourra ainsi intercepter vos communications y compris si vous remplissez des formulaires et ce malgré le protocole HTTPS rassurant.

1.3.4 La sûreté de la cryptographie

Pour finir ce chapitre, regardons comment on casse un algorithme de cryptographie :

- il existe une faille mathématique ou une découverte mathématique casse l'algorithme,
- il existe une faille de programmation [44],
- la clé est trop courte et il est possible de tester toutes les clés possibles en un temps raisonnable,
- une faille ou découverte mathématique permet d'éliminer assez de clés pour que l'on puisse tester toutes les autres en un temps raisonnable.

Il y a donc deux catégories : les failles et la force brute qui teste toutes les clés possibles.

La force brute

La longueur d'une clé est la seule protection contre cette attaque. Ainsi suivant les caractères que vous utilisez, l'alphabet, et la longueur de votre mot de passe, tester toutes les clés possibles est raisonnable ou non. Le tableau ci-dessous en donne une idée :

44. Il est très difficile de programmer un logiciel de cryptographie même si l'algorithme est simple. Il est plus prudent d'utiliser une bibliothèque qui comprend les algorithmes dont on a besoin.

Alphabet	4 caractères	8 caractères	12 caractères
Lettres minuscules	$26^4 = 456\,976$	$208{\times}10^9$	$954{\times}10^{15}$
Lettres minuscules et chiffres	$36^4 = 1,6{\times}10^6$	$2{\times}10^{12}$	$4{\times}10^{18}$
Minuscules, majuscules et chiffres	$62^4 = 14{\times}10^6$	$218{\times}10^{12}$	$3{\times}10^{21}$

TABLE 1.4 – Nombre de clés possibles suivant l'alphabet et la longueur

Si on suppose qu'on a un ou des ordinateurs qui peuvent tester un million de clés par seconde (chiffre très raisonnable) alors on voit qu'une clé de 4 caractères résiste au mieux 14 secondes. Par contre la clé de 12 caractères avec minuscules, majuscules et chiffres résistera un million de siècles…

La destruction de DES DES a une clé de 56 bits [45]. Il a été l'une des premières victimes cassées par la force brute :

- **En juin 97** Rocke Verser de Loveland, Colorado, le casse avec des machines d'autres internautes en 90 jours.
- **En janv 98** `distributed.net` le casse en 39 jours avec 10 000 ordinateurs et une moyenne de 28.1 milliards de clés testées par jour.
- **En juillet 98** `Electronic Frontier Foundation, EFF` le casse avec une machine à 250 000 $ fabriquée pour en 3 jours,
- **En janvier 99** DES est cassé en 22 heures par la machine de l'EFF couplée aux 100 000 machines réunies par le `distributed.net`.

Dans le dernier cas, près de mille milliards de clés étaient testées par secondes. A ce rythme, la clé de 12 caractères avec minuscules, majuscules et chiffres n'aurait tenu qu'un siècle. Sachant que la puissance des ordinateurs double tous les deux ans [46], cela veut dire que dix ans plus tard, la même clé ne résisterait plus que 3 ans.

Cela étant, tester une clé de type DES peut prendre moins de temps que de tester une clé de taille égale d'un autre algorithme, aussi il est important de faire attention aux comparaisons.

Le calcul distribué La force brute est une méthode qui se répartie très bien sur un ensemble d'ordinateurs, chacun testant une partie des clés. Aussi des internautes ont créé l'organisation distributed.net [47] afin de répartir le travail parmi les ordinateurs mis à leur disposition.

Avec cette méthode, le RC5 a été régulièrement cassé avec des clés de plus en plus longues :

- **En octobre 1997, RC5-56** est cassé en 212 jours de travail. Le pourcentage de clés vérifiées est de 47,03%, vitesse moyenne : 5,3 G clés/s. Au rythme final, il aurait fallu 83 jours pour vérifier l'ensemble des clés restantes.
- **En juillet 2002, RC5-64** trouvé en

45. il faut 6 bits pour stocker un caractère qui soit une minuscule ou une majuscule ou un chiffre, donc par rapport au tableau ci-dessus, 56 bits représente moins de 10 caractères.
46. Loi de Moore interprétée assez librement
47. depuis son dernier succès sur le RC5-64, ce site ne travaille plus sur les algorithmes de cryptographie.

```
1 757 jours de calcul, environ 4 ans et 10 mois
331 252 participants
15 769 938 165 961 326 592, 15 milliards de milliards de clés testées
  soit 81% des clés possibles

vitesse maximale : 270 147 024 000 clés/seconde
  - soit 32 000 de Apple PowerBook G4 800MHz ou
    46 000 PC AMD Athlon XP 2Ghz travaillant en parallèle
  - à cette vitesse il suffirait de 790 jours pour tester
    l'ensemble des clés
```

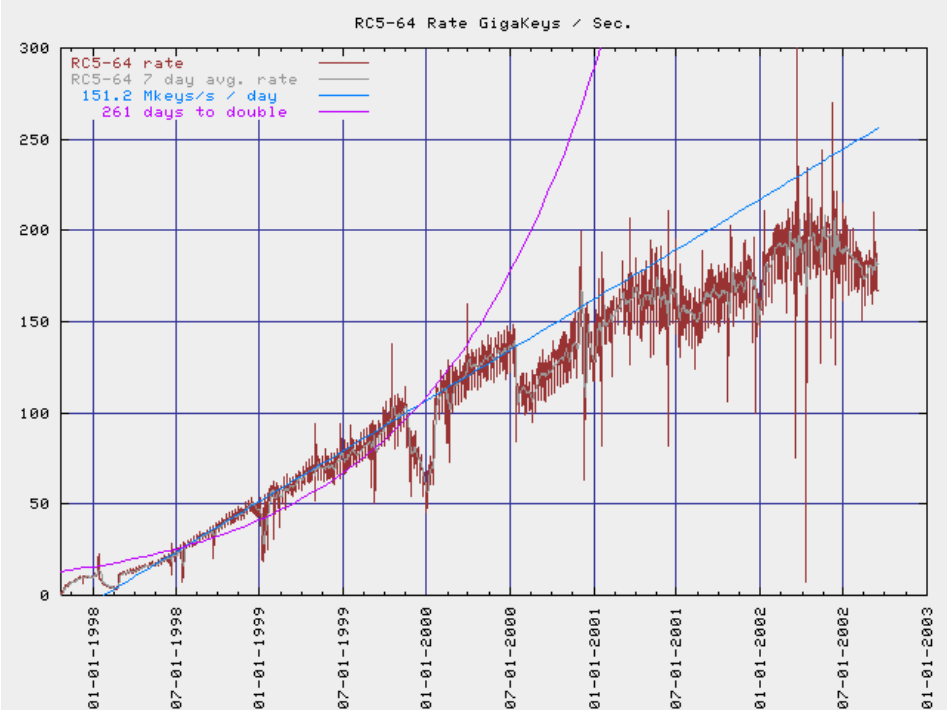


Figure 1.29 – Vitesse de test des clés RC5-64 durant le calcul

Casser le Web L'algorithme de cryptographie du Web est SSL. Dans les années 90 et encore au début des années 2000, il se conjuguait en 3 variantes de longueur de clés différentes : SSL 40 bits, SSL 56 bits et SSL 128 bits. Dès l'été 1995, SSL 40 bits a été cassé en 32 heures à l'INRIA et en 3h30 durant l'été 1997 à Berkeley et pourtant des banques l'utilisaient toujours en 2000. Aujourd'hui quelques quelques secondes suffiraient aussi il est indispensable d'utiliser la méthode SSL 128 voire 256 bits.

L'intelligence contre la cryptographie

Il existe peu de cas où des avancées mathématiques cassent des algorithmes de cryptographie, en voici néanmoins deux exemples.

RSA mis à l'épreuve Afin d'avoir une estimation de la sécurité de RSA, l'entreprise RSA Security organise un concours ouvert dont le but est de casser un message chiffré avec l'algorithme RSA d'une longeur de clef déterminée. Le but est de trouver les deux nombres premiers p et q qui génèrent le module de l'algorithme de RSA ce qui permet d'avoir la clé privée. Pour venir à bout de ce défi, la méthode mathématique utilisée est celle du "crible algébrique" qui permet de ramener le problème à un calcul matriciel dont la résolution nécessite un super ordinateur [48]. Ainsi

- **En février 1999, RSA-140 chiffres** a été cassé. Le crible a nécessité environ 125 stations SGI et Sun à 175 MHz et environ 60 PCs à 300 MHz pendant 1 mois. Le système matriciel a demandé 100 heures CPU et 810 MO de mémoire vive sur un Cray C916.
- **En août 1999,RSA-155 (512 bits)** tombe. Le crible a nécessité 160 stations SGI et Sun à 175-400 MHz, 8 SGI Origin 2000 processeurs à 250MHz, 120 Pentium II PCs à 300-450 MHz et 4 500 Digital 500 Mhz pendant **3**.7 mois. La matrice à résoudre avait 6 699 191 lignes et 6 711 336 colonnes pleines à 62.27%. Il a fallu 224 heures CPU et 3.2 GO de mémoire vive sur le même Cray pour résoudre le système.
- **En novembre 2005, RSA 640 bits** est tombé après 5 mois de calcul.
- **En décembre 2009, RSA 768 bits** est le dernier défi tombé.

Depuis 2010 RSA 1024 bits n'est plus considéré comme sûr. En 2017 RSA Security a indiqué que les clefs de 2048 bits devraient tenir jusqu'en 2030. L'organisme NIST suggère d'utiliser des clefs de 3072 bits si on désire que la sécurité dépasse 2030.

MD5 cassé affaiblit le Web Depuis 2004 on sait qu'il est possible de faire deux messages qui ont le même condensat MD5. En 2008, une équipe de chercheurs [49] a appliqué cette possibilité théorique à un cas bien pratique : la génération de faux certificats Web.

En temps normal un site web sécurisé envoie au navigateur un certificat qui prouve qu'il est bien le site web qu'il prétend être, cf figure 1.30. Le navigateur vérifie l'identité du site Web en vérifiant que le certificat qu'on lui envoie est bien signé par une autorité de certification connue (c.a.d. dont la clé publique est dans le navigateur). Si c'est le cas, il ne reste plus qu'à vérifier que les données écrites sur le certificat, comme l'URL, correspondent à celles du site web qu'on est en train de visiter. Tout ce travail est invisible pour l'utilisateur si tout se passe bien.

48. Une présentation sur la factorisation et donc sur la façon de casser RSA est présentée sur ce site : http://pauillac.inria.fr/algo/banderier/Facto/

49. Alexander Sotirov, Marc Stevens, Jacob Appelbaum, Arjen Lenstra, David Molnar, Dag Arne Osvik, Benne de Weger, cf http://www.win.tue.nl/hashclash/rogue-ca/

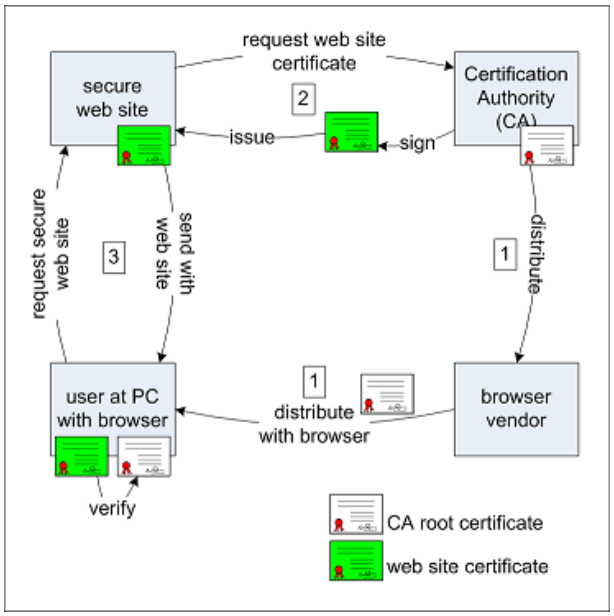


Figure 1.30 – Signature et utilisation normale des certificats SSL

Dans le cas normal, l'utilisateur (en bas à gauche) vérifie le certificat du site web (en haut à gauche) avec la clé publique de l'autorité de certification (en haut à droit) qui lui a été fournie avec son navigateur (en bas à droite).

L'attaque, cf figure 1.31, consiste à demander à l'autorité de certification de nous signer un certificat (le bleu). La signature étant faite sur le condensat MD5 du certificat, elle sera aussi valable si elle est attachée à un autre document qui a le même condensat que le certificat qu'on a envoyé. Cet autre est ici la clé publique de notre fausse autorité de certification (la noire). Avec cette fausse autorité, on peut signer le certificat de notre faux site web (le rouge). Maintenant il ne reste plus qu'à intercepter les requêtes vers le site web d'origine (en haut à gauche) et à lui présenter le certificat rouge du faux site accompagné de celui de la fausse autorité de certification (le noir). Ainsi le navigateur constate que le site a un certificat (le rouge), que ce certificat est signé par le noir lequel est signé par le certificat officiel de l'autorité de certification (puisque le noir a le même condensat que le bleu). Donc tout va bien et aucun avertissement ne sera envoyé à l'utilisateur qui se connectera au faux site web en toute confiance puisque la connexion est sûre grâce à SSL.

Depuis l'annonce de cette faille, les autorités de certification sérieuses n'utilisent plus le condensat MD5. Cela peut être vérifié en regardant l'algorithme de signature utilisé dans la description du certificat.

La bêtise contre la cryptographie

SSH cassé par ignorance En mai 2008 la distribution Debian de Linux doit annoncer que toute la sécurité basée sur OpenSSL est compromise. Quelques années auparavant, une personne en charge de faire marcher le logiciel OpenSSL sur Debian a retiré du code source des lignes qui semblaient ne servir à rien. Et si le fait de retirer ces lignes n'a rien modifié au fonctionnement

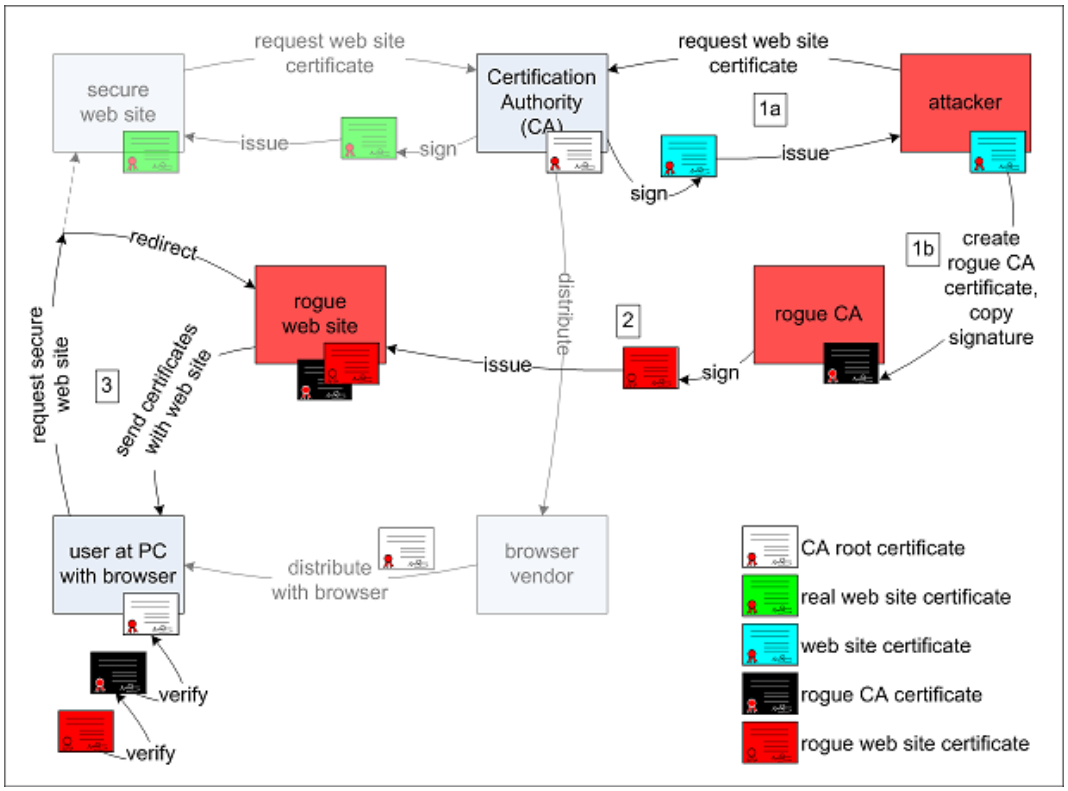


FIGURE 1.31 – Introduction d'un faux certificat SSL

du logiciel, cela a détruit la fonction aléatoire en charge de fournir les nombres de base pour générer les clés. Or si on peut deviner ces nombres de base, on peut aussi deviner les clés, donc toute la sécurité s'effondre.

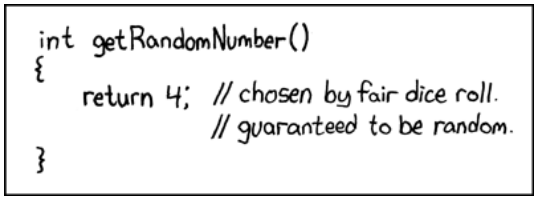


FIGURE 1.32 – XKCD se moque

Moralité : la programmation de la cryptographie est réservée aux spécialistes. Il est illusoire d'espérer programmer un algorithme de cryptographie sans générer des failles de sécurité si on n'a pas une longue expérience dans le domaine.

1.4 Plus

CircleID réunit des articles sur le fonctionnement de l'Internet, ce qui couvre plus que ce simple chapitre.

À propos de l'architecture d'Internet

- Quesques articles de l'encyclopédie Wikipédia :
  - le protocole d'Internet : http://fr.wikipedia.org/wiki/Internet_Protocol,
  - le DNS : http://fr.wikipedia.org/wiki/DNS
- Architectural Principles of the Internet, RFC 1958 par B. Carpenter, IAB, Juin 1996, https://www.rfc-editor.org/info/rfc1958

À propos de la sécurité

Pour lutter contre la faille humaine, une grande faiblesse de la sécurité informatique :

- Signal Arnaque, https://info.signal-arnaques.com/ lorsqu'on sent l'arnaque possible,
- INFO ESCROQUERIES au 0805 805 817 ou https://www.internet-signalement.gouv.fr/ pour signaler un mail ou site qui semble être une tentative d'escroquerie.
- Cybermalveillance https://www.cybermalveillance.gouv.fr, tant pour les victimes que pour se prévenir des cyber-attaques.

Pour les informaticiens ou passionnés, quelques sources hétéroclites :

- l'observatoire de la sécurité des systèmes, http://www.ossir.org/,
- le blog de Bruce Schneier, http://www.schneier.com/
- Black Hat, https://www.blackhat.com/, l'une des plus grande conférence sur la sécurité,
- le site de l'ANSSI, en charge de la cyber-sécurite en France, https://www.ssi.gouv.fr/
- l'ENISA, en charge de la cyber-sécurité en Europe, https://www.enisa.europa.eu/

À propos de la cryptographie

En ce qui concerne la cryptographie, on pourra aussi consulter les ouvrages suivants : *The Code-breakers* de David Kahn et l'*Histoire des codes secrets* de Simon Sing.

Certains manuels (livres) sont disponibles en ligne dont *The Handbook of Applied Cryptography* et les *Frequently Asked Questions About Today's Cryptography* des RSA Labs.

Enfin voici quelques sites sur le sujet :

- le cours de crypto de la Khan Academy pour débuter en cryptographie
- Learn cryptography présente de nombreux algorithmes et des informations générales
- la section crypto de Reddit
- le blog de Matthiew Green avec ses pensées sur l'actualité de la crypto mais aussi des liens vers d'autres ressources et ses cours.