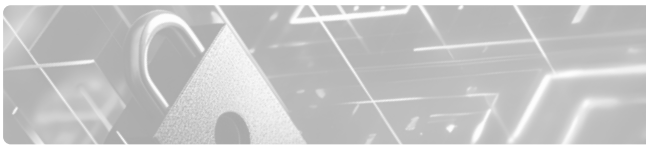
CYBERSÉCURITÉ

Patrick Lallement

Définitions
Concepts
Métiers

ellipses

Chapitre 1



Qu'est-ce que le cyber-espace ?

1.1 Les convergences

1.1.1 Évolution des architectures

Jusqu'aux années 80, les systèmes informatiques sont restés fortement centralisés en terme de ressources de traitement et de stockage. Les ordinateurs dits centraux ou *mainframe* exécutaient l'ensemble des opérations, assuraient le stockage. La centralisation n'est pas en soi une idée obsolète (surtout si on veut établir du contrôle). Ce qui l'était, c'était la position dominante des constructeurs des systèmes informatiques, qui s'appelaient IBM, Digital Equipment, Bull, propres à établir eux-mêmes leurs propres standards de communication entre leurs machines. Pour éviter l'emprise des standards de ces constructeurs, des groupes de travail se sont constitués dans les années 70 pour définir des modèles d'architecture de communication indépendants de ces grands acteurs du marché. C'est le rôle général des groupes de travail que de définir des standards qui permettent l'interopérabilité des systèmes, issus de constructeurs différents. C'est le principe des systèmes ouverts. Les débits de transmission n'étaient pas très élevés, le support était majoritairement le cuivre et l'obsession était alors la fiabilité des communications, l'objectif était de détecter et de corriger les erreurs. Le modèle d'architecture de communication dit modèle OSI (*Open System Interconnection*, standard de l'ISO) est basé sur ce principe. Dans la même période, un autre modèle est à l'étude par des universitaires américains autour des protocoles TCP/IP, standards du futur internet, leur approche était d'abord de concevoir un modèle de communication résilient.

D'un point de vue sécurité, deux facteurs allaient dans le bon sens : un trafic limité en nombre de connexions et en durée, l'approche centralisée qui permettait de contrôler les accès.

À partir des années 80, la décentralisation des ressources va croissante, basée sur l'apparition des mini et micro-ordinateurs, dotés de plus d'autonomie. Ceci a eu un fort impact sur le poste de travail, avec l'apparition des PC (*Personal Computer*) et celle des réseaux locaux : Ethernet, Token Ring. La fibre optique fait son apparition avec par rapport au cuivre, des taux d'erreurs beaucoup plus faibles

et des débits beaucoup plus élevés. L'autonomie (traitement, stockage) a suivi le développement des microprocesseurs. Les nouveaux acteurs sont alors les constructeurs de ces microprocesseurs (Intel, Motorola, NEC, etc.) et l'éditeur de logiciels Microsoft (applications, systèmes d'exploitation). Les réseaux locaux permettent de structurer les échanges entre collaborateurs d'entreprise avec des outils dédiés comme la messagerie, le transfert de fichiers. Le concept de "système d'information" apparaît dans les années 90.

1.1.2 Vers le tout numérique

Dans les années 60, l'information était en grande partie analogique : textes, musique, images, téléphone, TV. Depuis les années 2000, tout est numérisable à des fins de transmission et de stockage. Ceci a pour conséquence que tout processus de communication, de bout en bout, revient à transporter de l'information numérique (des 0 et des 1). La deuxième conséquence est le développement de processeurs spécialisés pour numériser, coder, compresser les signaux analogiques (son et image) avant de les transmettre. De nombreux standards de codage et compression existent dans ce domaine, surtout pour la voix depuis l'apparition des systèmes de communication mobile dans les années 90. Tout signal (son, image) est échantillonné de manière à remplir un buffer (mémoire tampon) de 0 et de 1 pendant un intervalle élémentaire. Chaque buffer constitue potentiellement un paquet à transmettre.

1.1.3 Le multimédia

Dans les années 90, la notion de données s'enrichit en se complexifiant. On parle de contenu multimédia pour désigner la nature hétérogène de l'information (à traiter, à transmettre, à stocker) : texte, son, image, vidéo. La messagerie voit l'extension MIME (*Multi-purpose Internet Mail Extension*) qui permet d'inclure dans un message des éléments de toute nature (voix, vidéo, photo).
Le web est inventé en 1990 et connaît un succès rapide, moteur du développement et de la démocratisation d'internet. Les sources d'information se multiplient, ce qui créé un nouveau problème : comment en retrouver 1 parmi N, ou comment retrouver une aiguille donnée dans une botte de foin qui ne cesse de grandir ? C'est le rôle des moteurs de recherche. De nouveaux acteurs apparaissent alors qui proposent des outils, Google est le plus connu.

1.2 Les technologies de l'internet

1.2.1 Le réseau

Le projet ARPANET (1964) avait pour but de concevoir une technologie résiliente de communication par paquets. Les principaux standards qui définissent les bases du modèle dit TCP/IP sont apparus dans les années 70 (IP, TCP, UDP,

DNS, etc.), un peu avant les concepts du modèle OSI (de l'ISO). Le standard TCP/IP s'est imposé comme réseau mondial à partir des années 90, *boosté* par l'apparition de l'application web. Le réseau n'est effectivement qu'un outil, au service des applications qui communiquent. Le web en est une, qui va dépasser toutes les autres. Le modèle TCP/IP est d'abord une technologie d'interconnexion (des réseaux existants) autour du protocole de transfert de paquets : IP (*Internet Protocol*) [1]. L'équipement clé de cette interconnexion est une passerelle appelée routeur (figure 1.1).

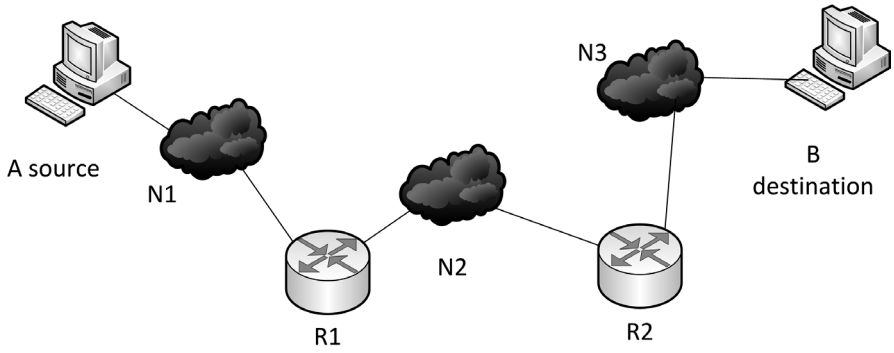


FIGURE 1.1 – Réseau internet

Internet est en fait un réseau de réseaux existants, interconnectés par des routeurs. Il peut même être simplement vu comme un réseau de routeurs (en faisant abstraction de la nature hétérogène des réseaux inter-connectés). Le routeur est d'abord une passerelle, pour passer d'un réseau à un autre, via une interface d'accès à une autre interface d'accès). Il est ensuite un aiguilleur de paquets. Les deux fonctions contribuent à l'acheminement des paquets de bout en bout, de la source jusqu'au destinataire final. Le réseau internet est l'aboutissement d'un projet né en pleine guerre froide dont la figure 1.2 rappelle les jalons.

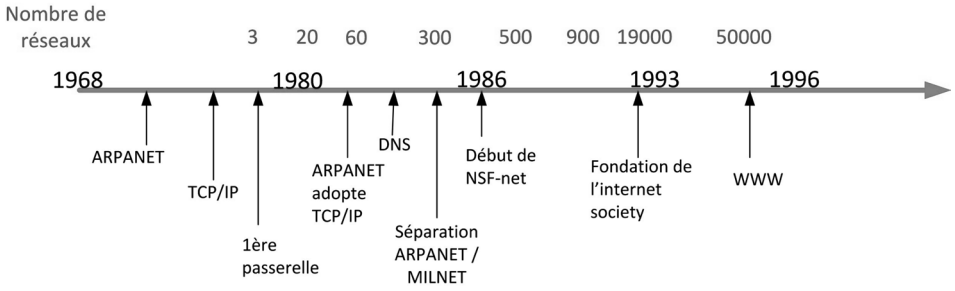


FIGURE 1.2 – Internet : les jalons

1.2.2 L'adressage

Principe

IP est le protocole d'interconnexion utilisé par l'internet pour l'acheminement des paquets. Ces paquets transitent donc de routeur en routeur avec l'adresse réseau (adresse IP) du destinataire.
Internet est aujourd'hui un réseau mondial, accessible à n'importe qui, ce qui lui confère un statut de réseau public. Par conséquence, joindre n'importe quel destinataire suppose que l'adresse de celui-ci soit publique (on doit pouvoir la trouver) et qu'elle soit unique.
Une adresse IP a un format unique structuré en deux parties : une partie publique (attribuée par les autorités de l'internet, ce qui garantit son unicité) et une partie privée (gérée par l'administration locale qui l'utilise), comme le montre la figure 1.3. Ce concept existait déjà avec les numéros de téléphone : par ex. dans une numérotation à 10 chiffres, les 6 premiers sont attribués à l'entreprise par l'opérateur (adresse de base visible dans un annuaire) et les 4 derniers sont laissés à la libre gestion et attribution par l'entreprise, ligne par ligne, pour la distribution finale, via son auto-commutateur privé.

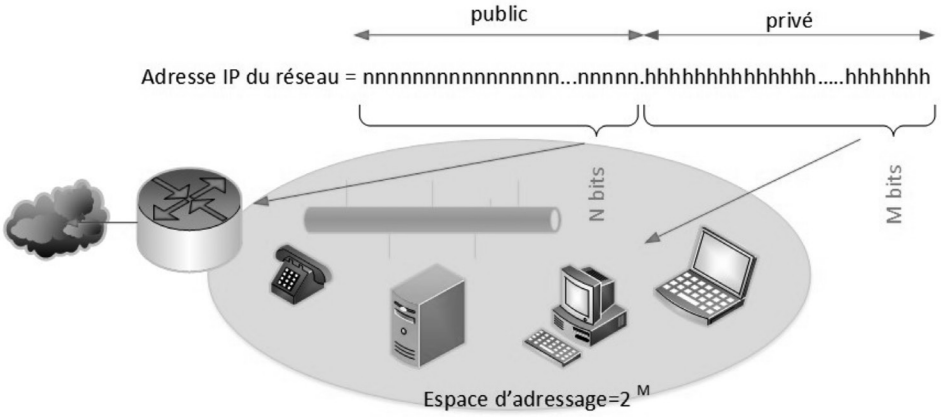


FIGURE 1.3 – Adressage IP

Format IPv4

L'adresse IP (notée ici @IP) est utilisée au format binaire par les machines et en décimal par les personnes (pour des raisons pratiques évidentes). Les chiffres (ou digits) sont séparés par des points.
Par exemple @IP=192.60.10.25 est une adresse IPv4 de 4 digits au format décimal, ce qui donne en binaire 1100000.00111100.00001010.00011001, c'est-à-dire 32 bits (4*8). La version décimale est plus facile à retenir.

Une adresse IP désigne en fait une interface d'accès à un réseau (et non pas un terminal !), ce qui veut dire que le routeur qui est une passerelle a au moins deux interfaces et a autant d'adresses que d'interfaces. Chaque adresse doit être conforme au plan d'adressage du réseau de rattachement.
À l'adresse IP on associe un masque de même longueur. À tous les bits contigus qui correspondent à la partie publique de l'adresse, correspond un 1 binaire dans le masque associé. Le masque permet par une simple opération (un ET logique) de retrouver l'adresse de base du réseau de rattachement. Par ex. si une entreprise se voit attribuer l'adresse @IP publique (en décimal) 195.60.10.0/24 cela veut dire que le masque associé a 24 bits à 1 consécutifs, soit (en décimal) : Masque = 255.255.255.0 = 11111111.11111111.111111111.00000000
Si on reprend l'adresse précédente @IP = 192.60.10.25 alors l'opération @IP ET Masque = 192.60.10.0, c'est-à-dire l'adresse de base du réseau d'appartenance de l'interface. Précision importante : une adresse de base de réseau ne peut pas être attribuée à une interface.
Tous les routeurs de transit dans l'internet ne connaissent (directement ou indirectement) que les adresses de base des réseaux. Seuls les routeurs de départ et d'arrivée connaissent l'adresse complète respectivement de la machine source et de la machine destinataire.
L'épuisement de l'espace d'adressage IPv4 est une des raisons qui ont conduit au développement d'une nouvelle version, IPv6, apparue dans les années 90 ([2]). Les principes restent les mêmes mais les deux protocoles sont incompatibles et les formats d'adressage différents. Aujourd'hui les routeurs doivent gérer les deux systèmes (IPv6 n'a pas remplacé IPv4) et les deux formats qui sont :

— IPv4 : adresses codées sur 32 bits soit 4 octets (espace d'adressage épuisé depuis 2014) ;

— IPv6 : adresses codées sur 128 bits soit 16 octets (espace d'adressage immense).

Les prévisions sur l'épuisement de l'adressage IPv4 se sont révélées exactes mais le scénario optimiste du remplacement de IPv4 par IPv6 ne s'est pas réalisé, même si la communauté mobile a fortement poussé à ce remplacement (les smartphones, apparus en masse dans les années 90, ont besoin d'une adresse IP).

Translation d'adresse (NAT)

Pour dépasser les limites de leur espace d'adressage avec l'adresse IPv4 dont elles disposaient déjà, les entreprises ont massivement utilisé un dispositif décrit dans le standard RFC 1918 [3]. Ce dispositif réserve des plages d'adresses utilisables dans la sphère privée, ce qui permet d'utiliser en privé un espace d'adressage beaucoup plus grand que l'espace publique qui a été attribué. Ces plages sont :

— 10.0.0.0/8 soit un espace d'adressage de $2^{32-8} = 2^{24}$;

— 172.16.0.0/12 soit un espace d'adressage de 2^{20} ;

— 192.168.0.0/16 soit un espace d'adressage de 2^{16}.

Tant que le trafic reste intra-entreprise (donc privé) cela fonctionne. Si la connexion doit sortir et donc passer dans le réseau public, cela n'est pas possible en l'état. Les adresses utilisées ne sont pas réservées exclusivement à des fins privées, elles ont aussi une existence publique quelque part (déjà attribuées dans le passé) et surtout elles ne sont pas uniques (elles peuvent être attribuées à plusieurs endroits en utilisant le standard RFC 1918). Pour réaliser la connectivité de bout en bout et passer de l'espace privé à l'espace public, il est alors nécessaire de translater l'adresse source privée en adresse source publique (l'adresse destination est forcément publique pour se connecter à un serveur externe publique).
Cette translation est réalisée au niveau du routeur par la fonction NAT (*Network Address Translation*), comme le résume la figure 1.4. Aujourd'hui presque tous les systèmes IPv4 utilisent le dispositif NAT pour adresser leurs machines dans l'espace privé, sans souci de croissance des besoins.
Ce dispositif prévu à des fins utilitaires a aussi un impact en terme de sécurité car vu de l'extérieur, l'adressage réel (privé) d'une machine est masqué. Il trouve sa place dans un schéma de connexion de type client (privé) vers serveur (public). Il pose problème par contre dans un schéma client privé vers client privé, ce qui est le cas de la téléphonie, car ne peut appeler de l'extérieur que des client dont on connaît l'adresse publique.

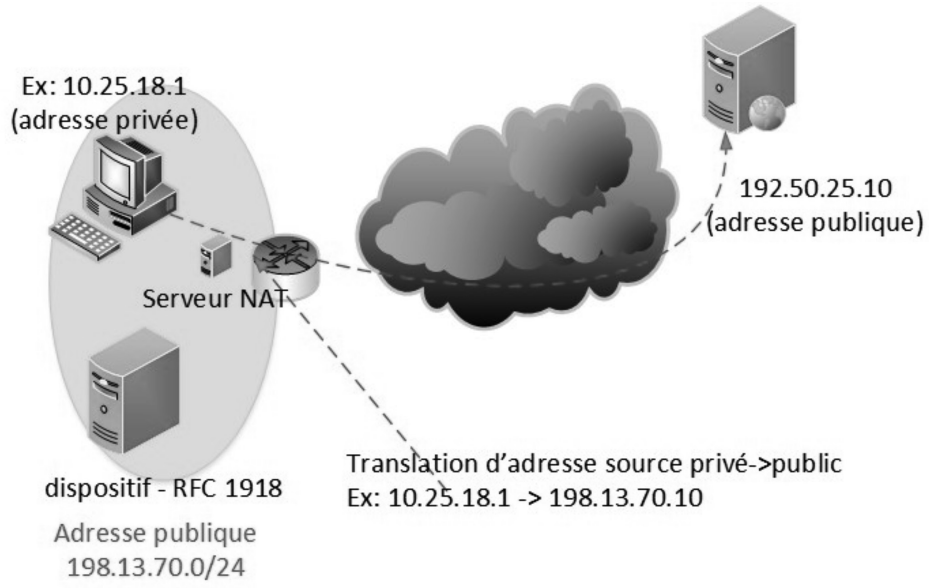


Figure 1.4 – Serveur NAT

1.2.3 Le protocole de transport

Le protocole IP ne contrôle ni la perte de paquet ni les erreurs. C'est le rôle du protocole de transport TCP (*Transport Control Protocol*) qui n'est exécuté qu'au

niveau des machines source et destinataire. Entre les deux, on ne voit plus qu'un tuyau virtuel. Comme plusieurs applications peuvent communiquer en même temps et utiliser le même interface IP, il est important de les distinguer par une adresse au niveau transport : le numéro de port. Il existe des ports statiques (référencés pour toutes les applications connues) et aussi des ports dynamiques, non référencés.
Il existe un protocole de transport non fiable (pas de contrôle des erreurs, ni de connexion préalable) qui est surtout utilisé pour des communications de 1 vers N ou des sessions multimédia : téléphonie, flux vidéo, vidéo conférence). C'est le rôle de UDP (*User Datagram Protocol*). Le système d'adressage est commun pour TCP et UDP.

1.2.4 IP et la sécurité

La technologie IP s'est massivement déployée au point de devenir le standard de transport de bout en bout, pour les données, la voix, l'image.
La convergence des technologies de communication vers un même standard protocolaire facilite les aspects fonctionnels, permet de réduire les coûts mais d'un point de vue sécurité, cela simplifie aussi la tâche de l'attaquant potentiel, surtout que les aspects sécurité ne figuraient pas dans le cahier des charges de ces standards de communication. Il faut attendre le protocole IPv6 dans les années 90 pour des mécanismes de sécurité soient prévus, en option : authentification, chiffrement. Mais IPv6 n'a pas remplacé IPv4. Il s'est ajouté. À mesure que les problèmes et les besoins sont apparus, il a fallu rajouter aux standards basés sur IPv4 des mécanismes plus ou moins complexes, selon le chemin à sécuriser, entre deux nœuds (au niveau IP) ou de bout en bout (au niveau des applications).

1.2.5 Le pare-feu

Définition

Le pare-feu (ou *firewall* FW) est un équipement placé en coupure entre deux réseaux, en premier lieu entre réseau public (internet) et réseau privé (voir figure 1.5).

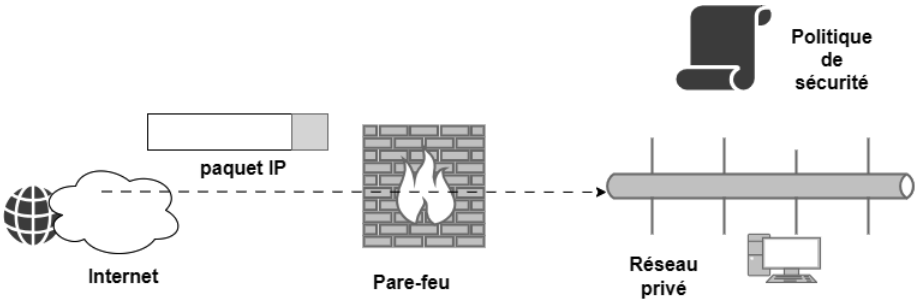


FIGURE 1.5 – Pare-feu (*firewall*)

Le FW implémente un mécanisme de filtrage basé sur des règles. Pour configurer un FW, il est nécessaire de définir au préalable une politique de sécurité.

Configuration

Pour filtrer un flux, deux stratégies sont possibles :

— tout ce qui n'est pas explicitement interdit est autorisé (liste noire) ;

— tout ce qui n'est pas explicitement autorisé est interdit (liste blanche).

Une règle de filtrage est une relation logique formalisée comme suit : SI <condition> ALORS <décision>. La décision est : laisser passer ou bloquer. Un filtre de paquets analyse chaque paquet IP et décide s'il est conforme ou non à la politique de sécurité du système où il veut entrer.
Les règles de filtrage sont couramment appelées des ACL (*Access Control List*).

Profondeur de filtrage

Pour décider de la conformité d'un flux, le FW utilise :

— des données de l'entête IP (niveau 3) ;

— des données de l'entête TCP (niveau 4).

Un pare-feu en coupure de l'internet inspecte tous les paquets entrants, ce qui constitue une charge de traitement importante. Pour cette raison les règles de filtrage utilisent des champs exploitables des entêtes IP et TCP.

— entête IP : adresses IP (source, destination) champ *Protocol* qui indique si le niveau 4 est TCP ou UDP ;

— entête TCP : numéros de port (source, destination).

On pourra se référer aux recommandations de l'Agence nationale de la SSI (ANSSI) [4] pour le paramétrage d'un pare-feu.

Pare-feu et DMZ

La fonction de FW est une fonction purement logicielle, comme le routage. Associées dans un même équipement, le filtrage/ routage permet de router les paquets acceptés vers un port physique associé à une zone démilitarisée ou DMZ. Une DMZ est une zone dite neutre mais qu'il faut sécuriser. C'est un segment de réseau séparé par un FW via un port spécifique dédié. La figure 1.6 en montre le principe qui suit une logique de segmentation des flux à des fins de sécurité. Il est conseillé d'utiliser ce mécanisme pour regrouper dans une même DMZ tous les serveurs publics (web, messagerie, DNS), car ils sont les plus exposés. Ainsi les paquets à destination de ces serveurs seront routés vers un port spécifique distinct des autres machines.
On peut aussi en suivant la même logique définir une DMZ pour regrouper les serveurs internes (bases de données, serveur intranet, serveurs métier etc.), qui ne doivent pas être vus par des personnes non autorisées.