Les Protocoles Réseaux

I. La pile de protocoles

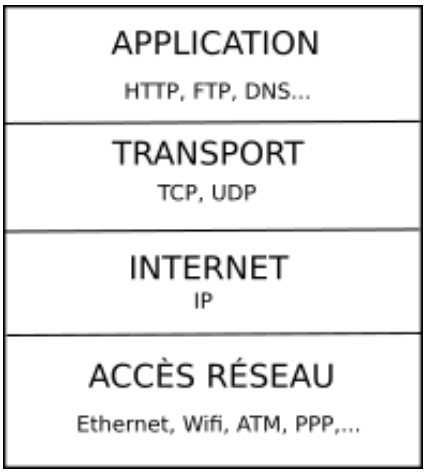
Un réseau informatique est constitué dès que deux machines peuvent échanger des données de façon uniquement logicielle (sans passer par exemple par une mémoire de masse externe manipulée par un humain).

Pour pouvoir communiquer, il a fallu inventer et élaborer des règles, des moyens de se reconnaître, de protéger les messages... Ces règles à suivre s'appellent des protocoles.

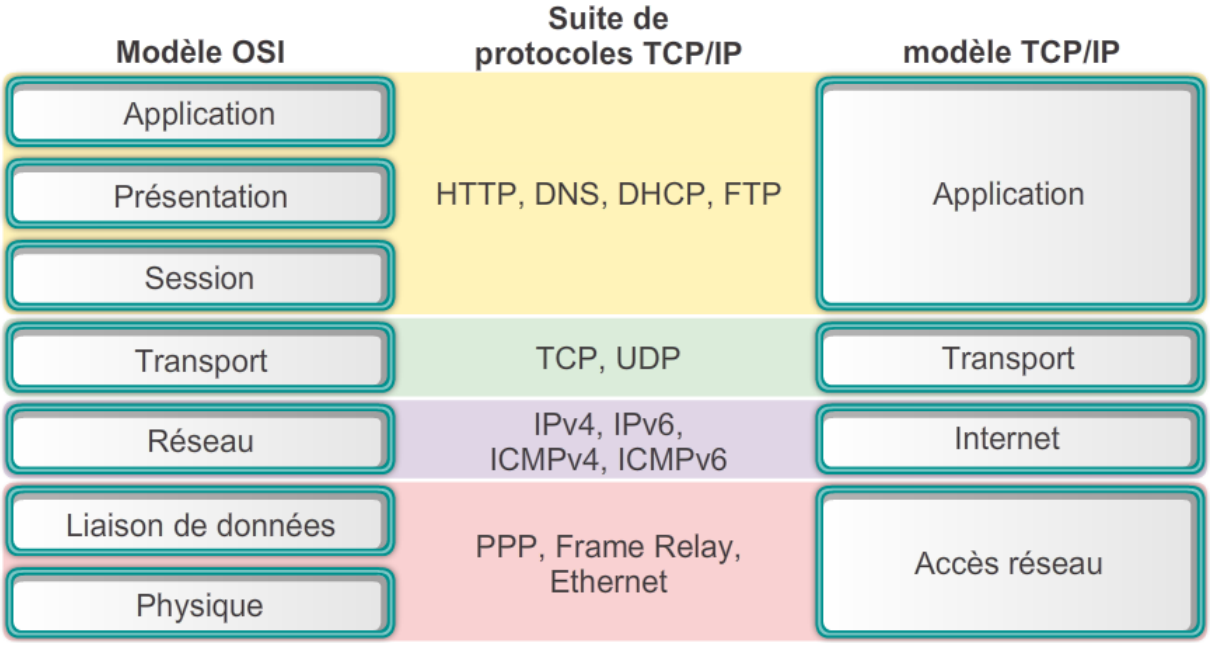
Ces protocoles sont organisés en "couches" imbriquées les unes dans les autres comme des pelures d'oignon ou des poupées russes.

Chaque couche a son propre champ d'action. Chaque couche est mise en œuvre par des concepteurs différents, des ingénieurs de spécialités différentes...

Ces couches sont traditionnellement représentées de façon verticale.
Les couches hautes sont proches du logiciel et de l'interaction avec l'être humain.
Les couches basses sont proches du matériel, de l'électronique, du medium de transmission.



Cette vue à 4 couches est une simplification de la conception purement théorique élaborée dans les années 1980 qui comprend 7 couches. On la nomme le *modèle OSI* :



Ces couches réseaux sont interdépendantes mais totalement modulaires :

- chaque couche utilise la couche inférieure sans se préoccuper de son fonctionnement ; il suffit de suivre les règles établies pour celle-ci et elle peut l'utiliser en lui passant les données à transmettre ;
- chaque couche "encapsule" le message reçu de la couche supérieure en lui rajoutant ses données à elle pour pouvoir jouer son rôle ;
- une couche peut être échangée par une autre couche de même niveau respectant le protocole sans que les autres couches ne doivent être modifiées.

Par analogie, un échange d'informations entre deux êtres humains peut être découpé en plusieurs "couches" :

- la communication peut utiliser un medium
  comme l'air (c'est le cas lorsqu'on utilise la voix pour se parler),
  comme l'eau (pourquoi pas ? comme les dauphins... l'est transmet très bien les vibrations "sonores"),
  comme la lumière (des signaux de fumées, des signaux de lumière, des gestes, ...),
  comme un câble électrique (le "vieux" téléphone),
  comme les ondes électromagnétiques (la radio, la CB),
  comme le réseau internet (la téléphonie actuelle, dite VoIP, les messageries instantanées, les emails...)
  comme la poste (transmettre un message par courrier traditionnel)
  c'est l'équivalent de la couche Accès réseau
- si le medium de la communication ne permet pas de transmettre le message sur une longue distance, par exemple des personnes voulant converser avec la voix, ou avec des signaux visuels, ou des textes écrits sur des dizaines, des centaines, des milliers de kilomètres, il faut prévoir des éléments intermédiaires... Comme une liaison filaire téléphonique, comme des relais (sémaphores), comme les dépôts de poste et les facteurs...
  c'est l'équivalent des couche Transport et Internet
- une fois le message transmis, encore faut-il l'interpréter :
  il peut être parlé ou écrit
  il est peut-être écrit dans une certaine langue qui doit être comprise et qui suit des règles

il peut-être à base d'autre choses que des mots comme un code à base d'images,
il peut reposer sur un codage intermédiaire comme le code morse...
on peut ranger ces aspects dans une couche application..

L'essentiel est de comprendre qu'une simple communication écrite ou parlée entre deux être humains (voir vivants : les animaux communiquent par le son, par des signaux chimiques comme les insectes... peut-être aussi les plantes...) repose sur des protocoles multiples...

II. La couche Accès Réseau

Cette couche est constituée de deux parties (la couche physique et la couche liaison du modèle OSI).

La première consiste en la mise en œuvre de l'exploitation du médium (câble électrique, fibre optique, ondes électromagnétiques) d'un point de vue électronique et de la gestion des collisions inévitables dues à l'utilisation d'un médium commun par plusieurs machines simultanément.

La deuxième permet à deux machines qui sont sur le même réseau local de s'échanger des données.
Par réseau local, on entend un ensemble de machines reliées par des media directement interconnectés.
On parle de *LAN : Local Area Network*.

II 1. l'adresse de la couche accès réseau

Les machines se reconnaissent sur un réseau local grâce à un numéro unique qui s'appelle *l'adresse MAC* ou adresse physique.
MAC signifie : Media Access Control

C'est un nombre de 48 bits, écrit sous la forme de 6 nombres hexadécimaux à deux chiffres.

Exemple : `50:7B:9D:42:D6:07`

Ce numéro est normalement fixe et est écrit en dur dans l'électronique du composant réseau.
Il est unique de façon universelle pour chaque composant.

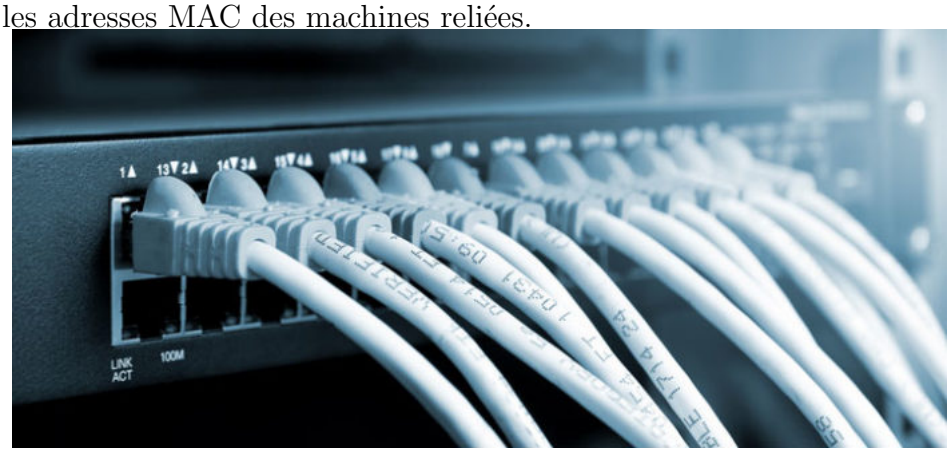
Remarque : Attention : l'adresse MAC n'est pas associée basiquement à une machine mais à chaque composant d'accès réseau de cette machine !
Cela signifie qu'on ordinateur portable a, le plus souvent, une adresse MAC pour son port Ethernet, une adresse MAC pour sa carte WIFI, une adresse MAC pour son composant Bluetooth.
Un smartphone, lui, a une adresse MAC wifi, une adresse MAC "datas" fournisseur (4G, 5G...), une adresse MAC Bluetooth...
Tout objet connecté possède au moins une adresse MAC.

II 2. les différents types de liaisons

Les acronymes de types de liaisons à connaître sont :
- l'Ethernet : c'est la liaison filaire la plus usitée dans les réseaux locaux.
  Il est utilisé sur les box internet, les réseaux d'entreprises, ...
  Il repose sur des câbles en cuivre torsadés et de prises dont la norme s'appelle RJ45.
- le Wifi : une liaison sans fil par ondes électro-magnétiques
- le Bluetooth : une liaison sans fil par ondes électro-magnétiques à courte distance

II 3. Les appareils fonctionnant dans la couche Accès

- Dans un réseau filaire, dit Ethernet, les machines utilisent des câbles qui sont interconnectés grâce à des concentrateurs efficaces appelés *switches*. Ces switches fonctionnent au niveau Accès avec les adresses MAC des machines reliées.



- Dans un réseau sans fil, dit Wifi, les machines utilisent les ondes électromagnétiques pour acheminer les données mais elles ne communiquent as directement l'une avec l'autre. Les messages transitent tous par un appareil appelé *borne d'accès wifi* et qui fait office de switch pour le réseau sans fil.

III. La couche Réseau

Cette couche a pour mission d'acheminer les messages d'une machine à une autre, même (et surtout !) si elles ne sont pas dans le même réseau local (LAN).

On interconnecte donc ici les réseaux locaux aux autres réseaux locaux et aux différents appareils et serveurs constituant l'internet.

On est au cœur d'internet !

C'est pourquoi cette couche réseau s'appelle aussi la couche Internet.

Dans la pile TCP/IP, cette couche porte le nom de *IP : Internet Protocol*.

Les appareils qui ont pour tâche de faire passer les données d'un LAN à un autre sont les *routeurs*. Ceux-ci travaillent uniquement dans la couche IP.

Ils sont interconnectés, sont capables de faire prendre la route la plus efficace aux données, de changer de route lors de l'apparition d'incidents sur l'une d'entre elles...

Comme pour la couche Accès Réseau, on a besoin d'identifier par un numéro les éléments, ici des machines. C'est ce qu'on appelle *l'adresse IP*.

Il existe deux versions de l'adresse IP :

- l'adresse IP v4 : elle a été rigoureusement définie en 1981 ;
- l'adresse IP v6 : elle a été définie en 1998.

III 1. l'adresse IP v4

Ce nombre de 32 bits est traditionnellement représenté par 4 octets écrits en décimal.

Exemple : `172.17.34.101`

III 1. a. Notion de sous-réseau

Le fonctionnement d'IP repose sur une différenciation "interne-externe" du réseau.

Une machine finale donnée est dans un réseau local (LAN) et est capable, grâce à la couche Accès, de communiquer avec ses voisines dans le même LAN.

Aussi, la règle est claire et nette :

- Si la machine envoie un message à l'une de ses voisines, elle le lui adresse directement, la couche Accès permet d'effectuer ce transfert ; une traduction entre l'adresse IP et l'adresse MAC (donc entre les identifiants de la couche IP et ceux de la couche Accès) est prévu, c'est le ARP dont on parle plus loin.
- Si la machine envoie un message à une machine "extérieure" à son LAN (on dit "si le mesaage doit sortir"), alors les données sont transmises à une machine particulière du LAN appelée *passerelle* et qui est en fait un routeur en bordure du LAN qui est relié à l'extérieur.

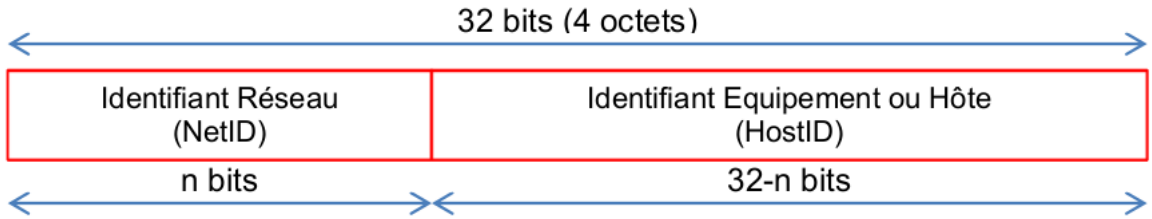
Dans ces conditions, la machine doit être capable de savoir si sa destinataire est une voisine ou pas.

Cela est possible avec le principe de *sous réseau* de l'adresse IP.

Une adresse IP comporte en fait deux parties.

Le numéro de réseau *NetId* est la partie de gauche de l'adresse, le numéro de la machine *HostId* est à droite.

La délimitation entre les deux parties est au choix de l'administrateur réseau.



On peut décrire la taille de chaque partie de deux façons :

- avec le *masque de sous réseau* : une sorte d'adresse IP particulière où tous les bits de la partie netId (à gauche) sont à 1 et tous les bits de la partie hostId sont à 0 ; cela peut donner par exemple `255.255.255.0` si la partie netId est constituée par les trois octets de gauche.
- avec un signe / (notation dite CIDR) qui suit l'adresse IP avec le nombre de bits dans la partie network. Cela donnerait `192.168.0.34/24` pour indiquer que la partie réseau est constituée des trois premiers octets.

De cette façon, les machines sont capables de reconnaître une adresse "sortante", c'est à dire qui n'est pas dans le même LAN (sous réseau) qu'elle.

III 1. b. Adresses publiques - adresses privées

Le nombre de valeurs possibles d'adresses IP v4, sur 32 bits, n'étant que d'environ 4 milliards, les adresses IP v4 sont rapidement venues à être rares.

Ce nombre est bien trop insuffisant pour que les adresses IP soient uniques de façon universelle.

Pour cette raison, des principes de fonctionnement ont dès les premiers temps été inventés pour palier à cette pénurie.

L'idée générale est que les machines à l'intérieur de leur LAN peuvent avoir des adresses qui sont juste uniques au sein de leur réseau mais pas d'un point de vue global.

Des machines dans des réseaux locaux différents peuvent donc très bien avoir des adresses ip identiques.

Ce n'est pas grave car elles ne communiquent pas directement l'une avec l'autre.

Elles n'échangent pas de données directement de l'une à l'autre car la communication passe par des routeurs... qui, eux, feront en sorte que tout arrive à la bonne machine.

On réserve donc des séries de numéros (on parle de *plages d'adresses*) à des adresses ip qui sont utilisées uniquement dans les réseaux locaux et qui ne sont pas "routées" à l'extérieur de ces LAN.

On appelle ces adresses des *adresses privées*.

Ces adresses privées sont libres d'utilisation.

Exemple : `172.17.34.101` , `192.168.34.101` , `10.54.174.231` sont des adresses privées.

Exemple : Il est très probable que dans quasiment tous les lycées de France il existe une machine ayant comme adresse IP `172.17.1.10` ! Cela ne pose aucun problème... Lorsqu'une de ces machines communique avec une machine qui n'est pas dans son réseau local, ce n'est pas son adresse IP qui est utilisée... (voir plus loin le mécanisme du NAT)

Exemple : Comme dans l'exemple précédent, le nombre de machines qui ont une adresse IP valant `192.168.0.1` est très important : c'est l'adresse par défaut de la box internet sur le réseau local (la box internet, comme tout routeur, possède plusieurs interfaces réseau et donc plusieurs adresses IP).

Voici les plages habituelles d'adresses privées :

- `10.0.0.0/8` donc de `10.0.0.1` à `10.255.255.254`
- `172.16.0.0/12` donc de `172.16.0.1` à `172.31.255.254`
- `192.168.0.0/16` donc de `192.168.0.1` à `192.168.255.254`

Par contre, les machines qui constituent le réseau des réseaux, les interconnexions des LANs, les routeurs, les serveurs principaux (DNS par exemple) doivent, eux, avoir des adresses IP uniques dans le monde.

Ce sont les *adresses publiques*.

Exemple : `83.196.123.73` , `201.23.45.67` , `180.54.4.23` sont des adresses IP publiques.

Les adresses IP publiques sont attribuées par les organismes officiels qui organisent internet et ne peuvent pas être utilisées de façon libres.

Le fournisseur d'accès internet fournit, à un foyer qui le rémunère, une adresse IP publique.

Elle est la même pour toutes les machines du foyer qui utilisent internet et qui, elles, sur le réseau local, ont toutes des adresses IP différentes (mais privées bien entendues).

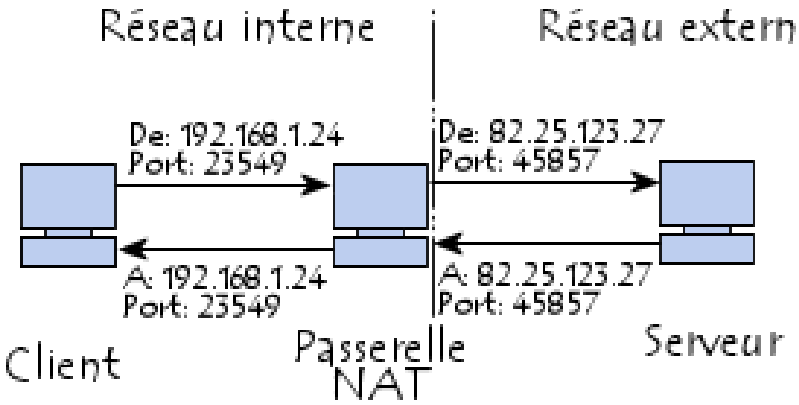
C'est la box internet, qui fonctionne comme un routeur, qui va raccorder le réseau local au reste du grand réseau de réseaux.

III 1. c. le principe du NAT

Mais comment est-ce possible, que des machines ayant des adresses IP privées, les mêmes que des milliers d'autres dans le monde, des adresses qui ne sont pas (ne doivent pas) être utilisées dans le réseau externe puissent échanger des données avec toute machine dans le monde ?

Le principe utilisé est le *NAT* : Network Adress Translation.

Le principe est le suivant :
Lorsque qu'une machine envoie des données "à l'extérieur" de son réseau local, elle les passe au routeur (à la maison c'est la box internet).
Le routeur, lui, va remplacer l'adresse IP de la machine émettrice par son adresse IP publique à lui pour l'envoyer vers l'extérieur.
Il va mémoriser (grâce au numéro de port TCP/UDP dont l'usage est ici quelque peu détourné, voir plus loin) quelle machine et quel port a été l'émetteur du message.



Ainsi, sur le réseau internet global, les échanges de données sont uniquement faits avec des adresses publiques, uniques au monde.

Ce sont les routeurs qui sont passerelles des LANs qui ont alors pour fonction de renvoyer le message à la machine destinatrice dans le LAN.

III 1. d. Adresses particulières réservées

Tout ensemble de 4 octets ne peut pas forcément être une adresse IP valide (publique ou privée).

- `127.0.0.1` est réservé au routage interne au sein de la machine ; on l'appelle *boucle locale* ou encore *loopback* ou encore *localhost* ; elle sert à la communication inter-processus internes à la machine ; le message n'est envoyé à aucune interface réseau et reste en mémoire vive ;
- si tous les bits à droite sont nuls (dans la partie host voir paragraphe "sous réseau") l'adresse ne peut désigner une machine en particulier mais désigne justement le sous réseau lui-même ; exemple `172.17.0.0` désigne le sous réseau 172.17.x.x/16 ;
- si tous les bits à droite valent 1 (dans la partie host) l'adresse ne peut désigner une machine en particulier mais permet d'effectuer un adressage *de diffusion* (broadcast), c'est à dire à envoyer le message à toutes les machines du réseau ; exemple `172.17.255.255` toutes les machines du sous réseau 172.17.x.x/16 vont recevoir le message ;

III 2. IP v6

Pour pallier au manque d'adresses IP v4, on a élaboré un principe reposant sur des nombres bien plus larges et donc bien plus nombreux.

Une adresse IP v6 est un nombre de 128 bits ! C'est à dire 16 octets !

Ce nombre est gigantesque et peut être considéré comme largement inatteignable pour l'instant. En effet : $2^{128} \simeq 3 \times 10^{38}$.
Cela représente plus de 2 milliards de milliards de machines possibles pour chaque millimètre carré de surface de la planète Terre !

On écrit habituellement une adresse IP v6 sous la forme de 8 mots de 16 bits en hexadécimal, c'est à dire 8 fois 4 caractères.

Exemple : `2001:0db8:0000:85a3:0000:0000:ac1f:8001`
qui peut se raccourcir en : `2001:db8:0:85a3:0:0:ac1f:8001`

III 3. Service DHCP

Dans un réseau local, chaque machine doit posséder une adresse IP (a priori privée).

L'administrateur réseau peut configurer ses machines une par une en entrant à la main cette adresse dans la configuration du système d'exploitation.
Cela peut rapidement être fastidieux s'il y a de nombreuses machines. Il peut aussi y avoir des erreurs (adresses non uniques par exemple...)
Par ailleurs, il ne suffit pas d'entrer une adresse IP pour que la configuration soit complète : il faut aussi renseigner la taille du sous réseau, l'adresse du routeur, l'adresse du serveur DNS, ...

Aussi, un service automatique existe sur les réseaux pour que les machines puissent obtenir de façon automatique une adresse IP valide et unique dans le réseau ainsi que les autres paramètres indispensables.

Ce service s'appelle *DHCP* : Dynamic Host Configuration Protocol.

Habituellement, dans un foyer, c'est encore une fois la box internet qui fournit ce service.

III 4. Interface entre les couches Accès et IP

La couche IP reconnaît les machines par leur adresse IP.
La couche Accès reconnaît les interfaces par leur adresse MAC.

Pour que les couches puissent travailler de concert (l'envoi d'un datagramme IP vers la couche Ethernet, c'est à dire Accès), il faut que la machine connaisse les adresses MAC des machines proches auxquelles elle doit envoyer des données.

Concrètement, lors de l'envoi de données par IP, le système analyse déjà (grâce à la donnée de la taille de la partie NetID de l'adresse IP) si la machine destinataire est dans le même réseau local.
Si oui, la couche Accès a besoin de l'adresse MAC de la machine destinataire pour envoyer la trame.
Si non, le protocole impose d'envoyer les données au routeur du réseau local qui est relié à l'extérieur (ce qu'on appelle la passerelle). Dans ce cas, la couche Accès a besoin de l'adresse MAC

de l'interface locale du routeur.

Ici intervient le protocole *ARP*, à la limite entre les deux couches. Ce protocole permet de mémoriser une table de correspondance entre adresses IP et adresses MAC, avec un système de demande-réponse lorsque l'information n'est pas connue.

IV. La couche Transport

La couche transport, reposant évidemment sur la couche Réseau, assure la communication non plus entre machines mais, de façon plus fine, entre processus.

Bien entendu, en réalité ce sont les programmes qui tournent sur la machine qui ont besoin du réseau.

Typiquement le navigateur Web a besoin du réseau pour effectuer ses requêtes au serveur Web distant, recevoir ses données permettant d'afficher la page web avec son code HTML, ses images, son code CSS, ses programmes Javascript...
Mais simultanément, un autre programme a aussi besoin d'accéder au réseau... Le client de messagerie par exemple.
Et un autre encore, par exemple le service de surveillance de mises à jours qui tourne en arrière plan.

Les programmes sont en fait constitués d'un ensemble de processus qui s'exécutent. C'est au niveau du processus que la communication se fait dans la couche transport.
Pour se différencier dans cet accès concurrent du réseau, le processus va passer un identifiant à l'adresse IP de la couche inférieure... cet identifiant s'appelle *le port*.

Ainsi, un processus, indiquant son port, échange des données avec un autre processus distant situé sur une machine qui possède une adresse ip et qui doit lui aussi être identifié grâce à son port.
Il y a donc un port pour le processus émetteur et un port pour le processus destinataire.

Le port pour le processus émetteur est important pour que la réponse qui reviendra sans doute du destinataire arrive au bon processus sur la machine (à bon port !)
Il ne s'agirait pas que le fichier HTML demandé par le navigateur web soit reçu par le client de messagerie alors que l'email soit reçu par le navigateur web !

Exemple : Imaginons que, sur une machine d'adresse IP (privée, sur le LAN) `192.168.20.35`, un processus communique avec le port 50842 avec un serveur web d'adresse IP distante `213.45.78.111` dont le processus de service web utilise le port 443 (port standard du HTTPS).
Ceci peut s'écrire `192.168.20.35:50842 -> 213.45.78.111:443`

Le port est un nombre de 16 bits (donc de valeur maximale 65535), écrit en décimal.

IV 1. Quelques ports standards

Certains numéros de port sont réservés et parmi ceux-ci certains sont standards pour des services très communs et très connus/

- 80 : port standard d'un serveur HTTP (web)
- 443 : port standard d'un serveur HTTPS (web crypté)
- 53 : port standard d'un serveur DNS (résolution de noms)
- 67 : port standard d'un serveur DHCP (attribution dynamique d'adresses IP)

- 110 : port standard d'un serveur POP (email réception)
- ...

IV 2. TCP

Deux protocoles différents peuvent prendre place dans la couche transport : TCP ou UDP.

Le premier d'entre eux, TCP est dit "en mode connecté".

Ce protocole permet un acheminement des données très fiable. De nombreux mécanismes sont prévus pour s'assurer que les données arrivent à destination sans perte.
Ces nombreux contrôles et dialogues avec le destinataire rendent la communication plutôt lente mais sécurisée.

De nombreuses utilisations sont faites de TCP, la plus fréquente étant le protocole HTTP/HTTPS du web.
De façon générale les utilisations du réseau ne pouvant accepter que les données soient corrompues à l'arrivée (protocole des emails, des téléchargements de fichiers)

IV 3. UDP

Le protocole UDP est dit "en mode non connecté".

Ce protocole permet un acheminement des données mais ne propose quasiment aucun mécanisme de contrôle et de fiabilité.
L'avantage est d'être plus rapide que TCP.

Les applications utilisant UDP :
- soit n'ont pas d'intérêt dans l'assurance de l'intégrité des données à la réception distante
- soit assurent eux-mêmes ces contrôles de fiabilité et de correction éventuelle.

Les utilisations classiques d'UDP sont le DNS, le DHCP, le jeu en ligne, la visioconférence, beaucoup de streaming (bien que Youtube et Netflix utilisent TCP)
Pour ces applications, la perte partielle de données n'entraîne que des défaut sans incidence pour la globalité du service (au pire une baisse de qualité acceptable)

V. DNS

Puisque les machines sur Internet sont référencées avec leur adresse IP, ce n'est pas très facile pour un être humain de les identifier, les différencier, les reconnaître...

C'est pourquoi on donne des noms aux machines, en plus de leurs numéros "adresses IP".

Exemple : Une machine possède le nom `www.google.fr`. C'est facile à retenir n'est-ce pas ?
En tout cas bien plus que `66.102.15.33`...

La machine, elle a besoin de l'adresse IP pour utiliser le réseau... Donc, à partir du nom de machine donnée par l'humain qui l'utilise, elle doit retrouver l'adresse IP.

Un service est conçu pour cela. Les *serveurs DNS* donnent l'adresse IP d'une machine dont on lui a indiqué le nom !