

Thème 1 : Internet

I. Internet et protocole IP

A. Le réseau des réseaux :

Que se passe-t-il lorsque vous tapez dans la barre d'adresse de votre navigateur « http://www.google.fr » ?

Votre ordinateur va chercher à entrer en communication avec un autre ordinateur se trouvant probablement à des milliers de kilomètres de chez vous. Pour pouvoir établir cette communication, il faut bien sûr que les 2 ordinateurs soient « reliés ». On dira que nos deux ordinateurs sont **en réseau**.

Il existe énormément de réseaux (la plupart des ordinateurs du lycée sont en « réseau »), certains réseaux sont reliés à d'autres réseaux qui sont eux-mêmes reliés à d'autres réseaux... ce qui forme « des réseaux de réseaux de réseaux... ».

Savez-vous comment on appelle cet assemblage multiple de réseaux ?

Internet, c'est le réseau des réseaux.

Dès les années 1950, les ordinateurs ont été mis en réseau pour échanger des informations, mais de façon très liée aux constructeurs d'ordinateurs ou aux opérateurs téléphoniques. Les réseaux généraux, indépendants des constructeurs, sont nés aux États-Unis avec ARPAnet (1969) et en France avec Cyclades (1971). Cet effort a culminé avec Internet, né en 1983.

Afin de pouvoir s'identifier, tout ordinateur possède une adresse sur un réseau : son **adresse IP**.

B. Le protocole Internet :

Le protocole IP (Internet Protocol) est l'ensemble des normes qui permettent d'identifier et de nommer de façon uniforme tous les ordinateurs ou objets qui sont connectés à Internet.

Le protocole permet l'acheminement des données en utilisant un système d'adressage : les adresses IP.

Une adresse IP est par exemple "74.125.133.94" (cette adresse IP publique correspond au serveur de google "google.fr"). C'est la **norme IPv4**.

Les adresses IP sont de la forme **décimale pointée** : "a.b.c.d", avec a, b, c et d des nombres entiers compris entre 0 et 255. Nous aurons l'occasion de revenir là-dessus un peu plus tard dans ce thème.

N.B. Une autre norme est en train d'être déployée, la **norme IPv6**. Pourquoi cette nouvelle norme ? Parce qu'avec le système IPv4, on va, dans les prochaines années et à cause des objets connectés (thème n°6), manquer d'adresses IPv4 disponibles sur Internet.

Le protocole IP assure l'acheminement des données sans garantie de résultat. Il s'agit donc d'un protocole non fiable. Il agit comme un transporteur de données, sans connaître le contenu de ce qu'il transporte et il ne prend pas en compte les aspects suivants

- La validité des données.
- Leur ordre d'arrivée au destinataire.

- La perte de données.
- Le réacheminement de données.

II. Protocoles TCP/IP

Le protocole IP est accompagné de protocoles de transmission pour transférer l'information par paquets, le principal étant TCP (Transmission Control Protocol).

Le réseau Internet s'appuie sur une grande variété de réseaux physiques où le protocole IP est implémenté (C'est-à-dire installé). Ce dernier uniformise l'accès à tous les ordinateurs, les téléphones et les objets connectés. Mais son rôle n'est pas de contrôler si les données circulent correctement.

A. Le rôle de TCP :

Ce protocole va agir de différentes manières.

- Premièrement, il assure que le message est bien arrivé à son destinataire via un **accusé de réception** (comme à la Poste pour un recommandé avec AR !). Il garantit ainsi la **fiabilité** de transmission.



On identifie les trois étapes : 1. Ouverture de la communication (handshake), 2. Requête et envoi du message, 3. Accusé de réception et clôture de la communication.

- Deuxièmement, si le message est trop gros, le protocole TCP va le découper en **paquets**. Ces paquets sont numérotés pour pouvoir les remettre dans l'ordre. En effet, ils peuvent emprunter des voies d'accès différentes et donc, arriver dans le désordre (voir cours routage).

Ce découpage est obligatoire car la communication doit s'adapter aux capacités de transmission des réseaux physiques (MTU : Maximum Transmit Unit).

- Troisièmement, en cas de perte de paquets (panne, erreur de routage, etc.), le protocole TCP les renvoie, et reconstitue le message grâce à la numérotation des paquets.

Les paquets sont normalisés sous la forme suivant :

En-tête IP avec n°	@IP source	@IP destinataire	Fraction du message/données
---------------------------	-------------------	-------------------------	------------------------------------

Le protocole TCP (Transmission Control Protocol) est l'ensemble des normes qui permettent d'assurer la bonne réception de données via le réseau Internet. Il est couplé au protocole IP.

Le protocole TCP assure le transport des paquets sans perte de données en

- Gérant le début et la fin de la communication.
- Numérotant les paquets.
- Classant les paquets.
- Assurant l'émission d'accusés de réception.
- Le renvoi de paquets si nécessaire.

B. Pour ceux qui veulent approfondir : (facultatif)

Si l'expéditeur ne reçoit pas d'accusé de réception au bout d'un certain temps, il renvoie le même message.

Dans le protocole TCP/IP, la durée au-delà de laquelle on considère que le paquet est perdu est appelée **RTT** (Round-Trip Time) et est comprise entre 0 et 90 ms. C'est très court à échelle humaine mais tout va très vite dans un câble, donc 90ms est un temps assez long pour considérer un paquet perdu.

De plus, un paquet a une durée de vie **TTL** (Time To Live). Ce champ est initialisé par l'émetteur puis diminué par chaque routeur traversé (voir cours routage). Quand le TTL arrive à 0 (TTL de départ = 255 ou 127), le paquet est supprimé par le routeur qui avertit l'expéditeur que ce paquet a été détruit. Cela évite que des paquets errent à l'infini sur le réseau.

III. Le système DNS

Internet Protocole manipule deux types d'information : les contenus envoyés et les adresses du destinataire et de l'émetteur. Ces deux types d'information sont regroupés, grâce au protocole TCP, dans des paquets de taille fixe, de façon uniforme et indépendante du type de données transportées (texte, images, sons, vidéos, etc.).

Les adresses sont numériques et hiérarchiques mais l'utilisateur connaît surtout des **adresses symboliques normalisées**, comme wikipedia.fr. Admettez que c'est plus facile à dire que « j'ai trouvé l'info sur 78.109.84.114 ! ». Et savez-vous qui est à 46.105.57.169 ?

A. Des adresses symboliques aux adresses numériques :

Le système **DNS (Domain Name System)** transforme une **adresse symbolique en adresse numérique**. Ce travail est réalisé par un grand nombre d'ordinateurs répartis sur le réseau et constamment mis à jour. Ce sont les « serveurs DNS ».

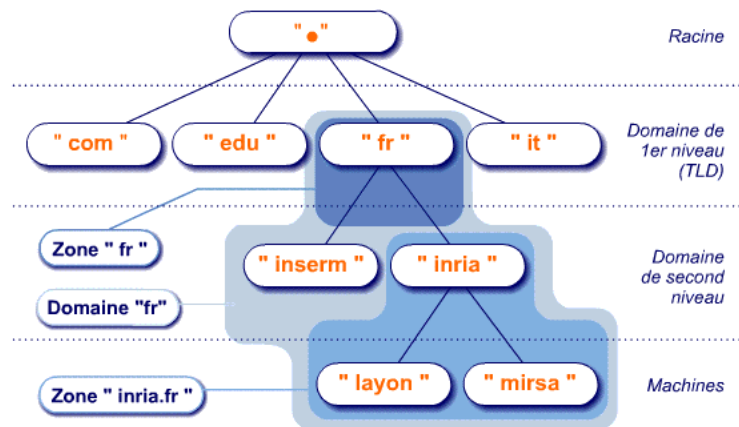
C'est Paul Mockapetris, alors qu'il travaillait à l'ISI (Information Sciences Institute) de l'Université de la Californie du Sud, qui proposa en 1983 une architecture de nommage des machines. Le DNS était né.



B. Un système hiérarchisé :

Le Domain Name System (DNS) est un **répertoire distribué** s'appuyant sur une structure de noms hiérarchisée.

La **résolution** (traduction de l'adresse symbolique en adresse IP) d'un nom de domaine est réalisée en partant du domaine racine et en suivant les délégations successives.



Exemple : voici l'adresse symbolique (on parle d'URL, on en reparlera dans le thème Web) du programme officiel de SNT :

https://cache.media.education.gouv.fr/file/SP1-MEN-22-1-2019/08/5/spe641_annexe_1063085.pdf

https : nom du protocole http sécurisé (voir thème Web).

Cache.media.education.gouv.fr : ce site se trouve à l'adresse IP 152.199.20.7. Le TLD (Top Level Domain) est en « .fr ». Puis dans le domaine « gouv », puis dans le sous-domaine « education », puis dans le sous-sous-domaine « media » et enfin dans le sous-sous-sous-domaine « cache ». Les adresses symboliques se lisent de droite à gauche.

file/SP1-MEN-22-1-2019/08/5/spe641_annexe_1063085.pdf : c'est le chemin d'accès. Cela signifie que le document au format .pdf se trouve dans le répertoire « file » qui contient le sous-répertoire « SP1-MEN-22-1-2019 » qui contient le sous-sous-répertoire « 08 » qui contient le sous-sous-sous-répertoire « 5 » qui contient le document. Les arborescences se lisent de gauche à droite.

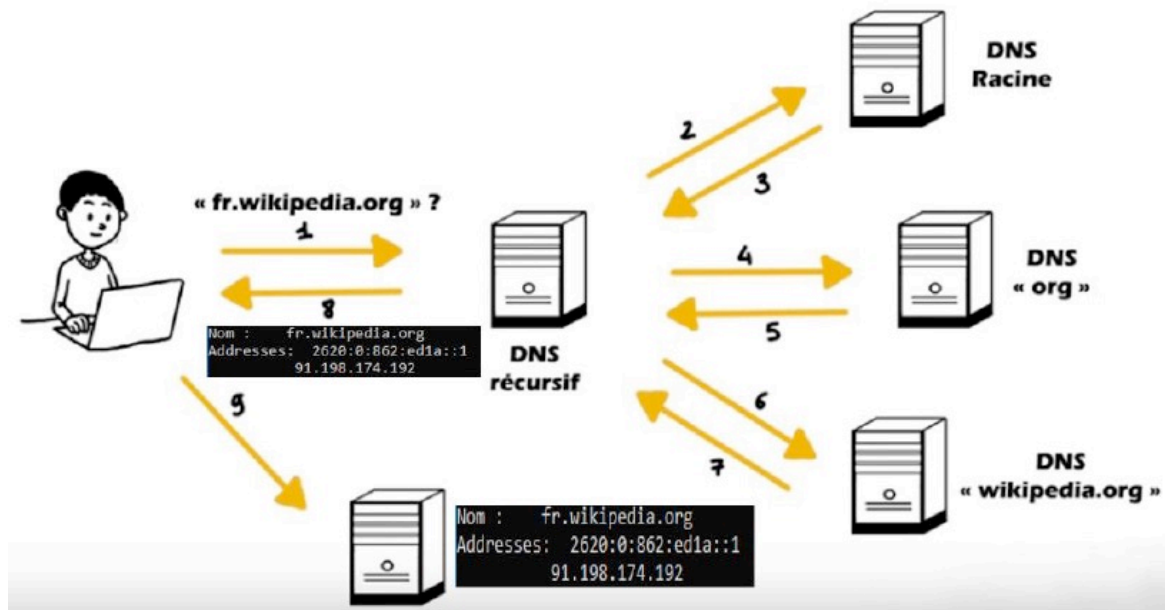
Les registres DNS sont gérés par IANA (International Assigned Numbers Authority), qui est un département de l'ICANN (Internet Corporation for Assigned Names and Numbers), une organisation à but non lucratif qui exécute la gestion de la zone racine dans le système de noms de domaine.

En France l'AFNIC est l'office d'enregistrement désigné par l'État pour la gestion des noms de domaine sous l'extension « .fr ».

C. Principe de fonctionnement :

Les répertoires DNS, ou registres, sont distribués sur des serveurs partout sur la planète.

Pour résoudre une adresse symbolique, un va-et-vient se met en place entre le client et les différents serveurs, depuis celui de la racine, à ceux des sous-domaines.



Un utilisateur saisit l'adresse symbolique *fr.wikipedia.org*. Son ordinateur va interroger le serveur DNS racine qui l'envoie vers le serveur DNS de l'extension « org », qui le renvoie sur le DNS du domaine *wikipedia*, qui lui renvoie l'adresse IP. L'ordinateur de l'utilisateur peut donc envoyer sa requête à l'adresse IP correspondante à l'adresse symbolique. Tout cela est totalement invisible pour l'utilisateur !

A retenir :

- ✓ Le système DNS (Domain Name System) transforme une adresse symbolique en adresse numérique.
- ✓ Il est réalisé par un grand nombre d'ordinateurs répartis dans le monde et est constamment mis à jour.

Une vidéo bilan sur IP, TCP et DNS

https://www.youtube.com/watch?v=5xSNH6Rf_CO

IV. Le principe du routage des paquets

Un autre algorithme très important d'Internet est l'algorithme de routage des paquets de leurs émetteurs vers leurs destinataires.

A. Le principe du routage :

Le routage est effectué par des machines appelées **routeurs**, qui échangent en permanence avec leurs voisins pour établir une carte locale de ce qu'ils voient du réseau.

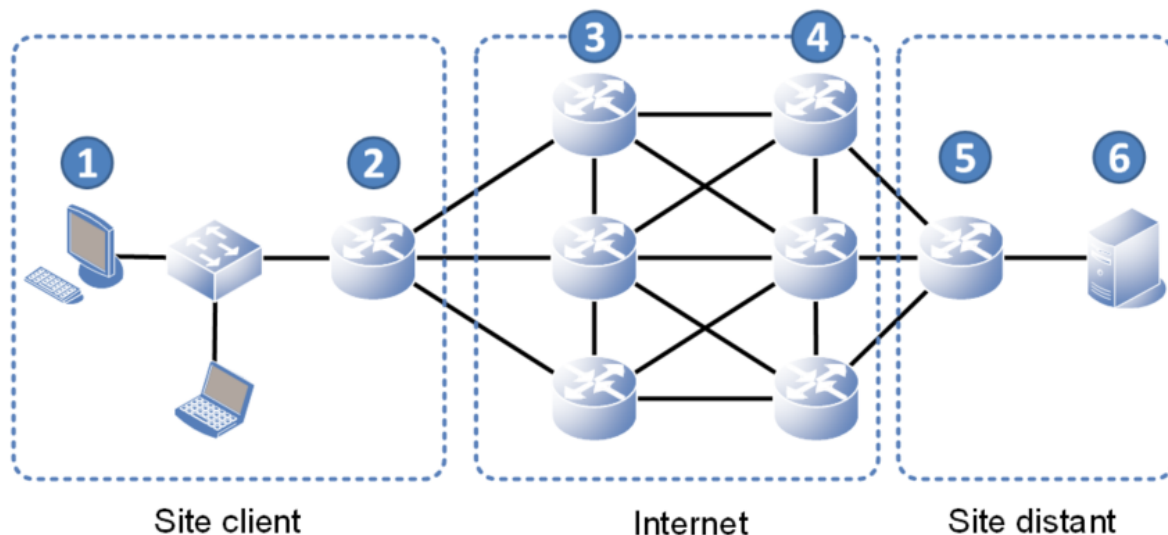
Chaque paquet transite par une série de routeurs, chacun l'envoyant à un autre routeur selon sa carte locale et la destination prévue. Les routeurs s'ajustent en permanence et de proche en proche quand on les ajoute au réseau ou quand un routeur voisin disparaît. **Il n'y a plus besoin de carte globale**, ce qui permet le routage à grande échelle.

Lors du routage, un paquet peut ne pas arriver pour deux raisons : une panne matérielle d'une ligne ou d'un routeur, ou sa destruction. Chaque paquet contient l'information d'un nombre maximal de routeurs à traverser : pour ne pas encombrer le réseau, il est détruit si ce nombre est atteint (c'est sa TTL : time to live).

C'est le protocole TCP qui fiabilise la communication en redemandant les paquets manquants. Il garantit que tout paquet finira par arriver, sauf panne matérielle incontournable.

TCP réordonne aussi les paquets arrivés dans le désordre, et diminue la congestion du réseau en gérant au mieux les redemandes. Mais ni IP, ni TCP ne possèdent de garantie temporelle d'arrivée des paquets, ce qui nuit à la qualité du streaming du son ou des vidéos et de la téléconférence. En effet, dans une vidéo on peut perdre une image isolée, mais pas le fil du temps.

D'autres protocoles s'appuient sur ceux d'internet, par exemple les protocoles du Web (HTTP et HTTPS) et le protocole NTP (Network Time Protocol) qui permet de synchroniser finement les heures des ordinateurs et objets connectés.



Un client (1) a demandé une page Web. Le serveur du site distant (6) l'envoie à son routeur (5). Celui-ci utilise le protocole TCP pour le découper en paquets. Le protocole de routage envoie les paquets sur des routeurs voisins (4), qui les transmettent à leurs voisins (3). Le routeur du client (2) réceptionne tous les paquets et le protocole TCP les remet dans le bon ordre (les paquets ayant pu emprunter des routes différentes, ils sont sûrement arrivés dans le désordre. C'est l'absence de garantie temporelle).

B. Les machines :

Réseau mondial, internet fonctionne à l'aide de routeurs, de lignes de transmissions à très hauts débits (fibres optiques) entre routeurs, de réseaux de téléphonie mobile, et de réseaux locaux. Ses protocoles étant logiciels, il peut s'appuyer sur n'importe quel réseau physique qui les implémente : 4G, Ethernet, ADSL, Wi-Fi, Bluetooth, Câbles, Satellites, etc.

TCP/IP n'est pas implémenté dans l'infrastructure, mais dans chacun des ordinateurs connectés, et un serveur DNS est aussi un ordinateur connecté. Des mécanismes complexes assurent la continuité de la connexion, par exemple pour passer sans interruption de téléphonie 4G au Wi-Fi, ou son ubiquité, par exemple pour passer de façon invisible d'antenne à antenne avec un téléphone portable quand on voyage.

A retenir :

- ✓ Les protocoles IP et TCP garantissent une transmission fiable (pas de perte de données).
- ✓ Les protocoles IP et TCP ne garantissent pas la temporalité.
- ✓ Le protocole de routage envoie les paquets qui ont une durée de vie limitée et via éventuellement différentes routes de routeurs.
- ✓ Les routeurs sont des machines qui se connaissent de proche en proche (mais pas globalement).

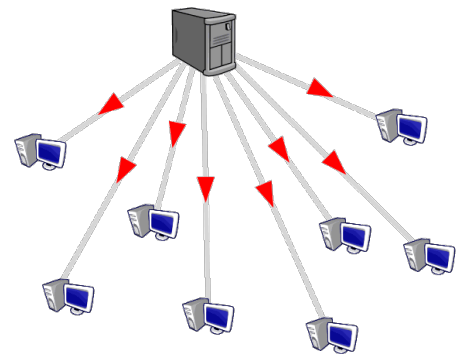
V. Réseau pair-à-pair

Dans les réseaux pair-à-pair, **chaque ordinateur sert à la fois de serveur (émetteur) et de client** (récepteur).

A. Le mode client/serveur : à quoi ça sert ?

Imaginez : Un éditeur veut distribuer une version de démonstration d'un jeu très attendu. Le fichier est très volumineux et est donc long à télécharger.

Habituellement, il va placer le fichier sur son site web, et tous les internautes vont venir le télécharger. Ce serveur web va donc être la seule source du fichier.



Le serveur de l'éditeur envoie le fichier à chaque client.

Problème : Comme le fichier est très populaire, il y a énormément d'internautes qui viennent le télécharger en même temps. Du coup, le serveur a de plus en plus de mal à répondre aux demandes, au point de ne plus pouvoir répondre du tout ! C'est le DoS (déni de service).

C'est ennuyeux, car le serveur veut distribuer ce fichier pour en faire profiter tout le monde et les internautes veulent obtenir ce fichier...

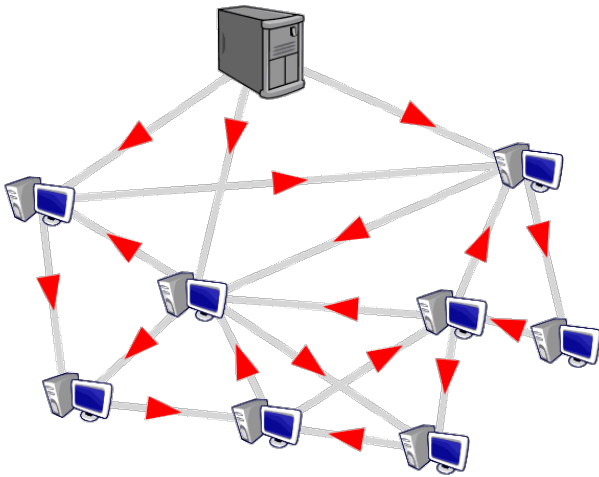
D'égale à égale :

Peer-to-peer (P2P) signifie "d'égal à égal", "de pair à pair".

Et de fait, sur internet, tous les ordinateurs sont égaux. Ils peuvent tous envoyer et recevoir des données !

L'idée :

Puisque chaque internaute est capable aussi d'envoyer des données, il pourrait fournir aux autres internautes les bouts du fichier qu'il possède déjà. Cela allègerait le serveur. C'est sur ce principe qu'est basé le Peer-to-Peer.



Chaque fois qu'un internaute télécharge un fichier, il partage en même temps les bouts du fichier qu'il a déjà reçu avec les autres internautes. **L'ordinateur de chaque internaute se comporte automatiquement en petit serveur, même s'il ne possède qu'une toute petite partie du fichier.**

Le serveur d'origine n'est plus la seule source du fichier : La charge est répartie sur tous ceux qui sont en train de télécharger.

Dès qu'une partie du fichier est reçue par un client, celui-ci devient serveur pour l'envoyer à d'autres clients.

A retenir :

- ✓ **L'utilisation de logiciels de peer-to-peer est légale.**
- ✓ **Ce qui est illégal, c'est d'utiliser des logiciels de Peer-to-peer pour télécharger ou distribuer des œuvres protégées par le droit d'auteur, ou d'autres choses (pédopornographie, terrorisme, etc.).**
- ✓ **Il ne faut pas confondre l'outil et les utilisations qui en sont faites. Toute technologie peut être utilisée à de mauvaises fins.**
- ✓ **Il y a également des tas de fichiers distribués légalement par les logiciels de Peer-to-peer (comme les distributions Linux ou des logiciels libres, des documents scientifiques, des œuvres libres de droit, etc.). Mais il faut admettre que la grande majorité des fichiers qu'on y trouve sont illégaux dans la plupart des pays.**

B. Impacts sur les pratiques humaines :

Internet a fait progressivement disparaître beaucoup des moyens de communication précédents : télégramme, télex, le courrier postal pour une bonne partie, et bientôt le téléphone fixe grâce à VoIP (voix sur IP). Son trafic prévu pour 2021 est de 3 300 milliards de milliards d'octets ($3,3 \times 10^{21}$ octets).

Internet a aussi ses problèmes : absence de garantie temporelle sur l'arrivée des paquets, et possibilité d'attaques par saturation en envoyant un très grand nombre de messages à un site donné, pour y provoquer un déni de service.

La neutralité du Net, présente dès l'origine du réseau, exprime l'idée que les routeurs doivent transmettre les paquets indépendamment du type de leur contenu (texte, vidéo, etc.) et indépendamment des requêtes (particulier, entreprise, état, etc.). Mais elle est constamment remise en cause par certains lobbies industriels.

Le réseau Internet est utilisé pour différents services : les mails (SMTP), le partage de fichier (FTP), le pilotage à distance (TELNET), etc. Mais surtout, le plus utilisé : le service WEB... Ce sera l'objet d'étude de notre thème n°2.

En conclusion :

Dans ce thème, on a vu que le réseau Internet fonctionne grâce à de nombreux algorithmes appelés protocoles : IP et TCP, routage.

Pour faciliter l'adressage aux humains, on utilise des adresses symboliques, traduites grâce au système DNS.

Enfin, il existe deux environnements de réseau : client/serveur ou pair-à-pair.