Fiche

Internet est un réseau informatique mondial qui relie entre eux des appareils (ordinateur, smartphone, avion...) grâce à des routeurs (box Internet...) et des liens (câble, fibre optique...). Lorsqu'un individu souhaite aller de Lille à Toronto, il existe plusieurs étapes à effectuer dans un certain ordre : acheter un billet d'avion à l'aéroport, aller déposer ses bagages, se rendre à la porte d'embarquement, monter dans l'avion, attendre pendant le vol. Puis on effectue les étapes dans l'ordre contraire : descendre de l'avion.... Lorsqu'un appareil souhaite envoyer une donnée à un autre appareil via Internet, la donnée doit passer par plusieurs couches dans un certain ordre : application, transport, réseau, liaison et physique. Puis, la donnée passera par les couches dans l'ordre contraire.

I. La suite des protocoles Internet

• Cette suite (souvent appelée **TCP/IP**) n'est pas implémentée dans l'infrastructure, mais dans chacun des ordinateurs connectés. Il existe énormément de protocoles différents selon les couches.

Couche application

• Dans la **couche application**, des protocoles sont utilisés **directement par l'utilisateur**, notamment HTTP (Web), XMPP (WhatsApp), POP, SMTP, IMAP (courriel), BitTorrent (partage de fichiers), FTP ou encore Protocole skype (privé)...

• Lorsque l'on envoie un **message** (texte, image, son...) d'un appareil à un autre, le message est encodé en **binaire** (à l'aide de **bits** : 0 ou 1).

Couche transport

• Dans la **couche transport**, le principal protocole utilisé est **TCP (*Transmission Control Protocol*).** Il est responsable de l'envoi du message de l'application de l'émetteur à l'application du destinataire. TCP établit une connexion entre deux appareils (synchronisation), **garantit que tous les messages arrivent à destination**, en bon état (*checksum*) et remet les **messages dans l'ordre**. Il utilise pour cela des accusés de réception, demande éventuellement les messages manquants et diminue la congestion du réseau en gérant au mieux les demandes. TCP ajoute des informations (par exemple les numéros de port du destinataire et de l'émetteur) au message et crée un **segment**.

• **Numéro de port :** entier compris entre 0 et 65535 (codé en binaire avec 16 bits). Les entiers de 0 à 1023 sont reservés : 80 pour HTTP, 53 pour DNS... Il sert à indiquer quelle application d'Internet est concernée par le message.

Couche réseau

• Dans la **couche réseau** le protocole **IP** est le principal, **il détermine les meilleurs chemins pour traverser le réseau** et ainsi faire circuler l'information entre deux appareils. Il est responsable d'**assigner des adresses IP** aux appareils du réseau. Le protocole IP ajoute des informations (par exemple l'adresse IP du destinataire et de l'émetteur) au segment créé par la couche transport et crée un **paquet**.

Qu'est-ce qu'une adresse IP ?

• C'est une sorte de numéro d'identification de la machine sur un réseau. Une adresse IPv4 est un entier naturel composé de 32 bits. On la représente généralement en notation décimale sous la forme de quatre entiers naturels compris entre 0 et 255, séparés par des points.

• Par exemple, le serveur hébergeant le site web Météo France a pour adresse IP : 185.86.168.100.

• Il existe 2^{32} = 4 294 967 296 adresses IPv4 (plus de 4 milliards). C'est pourquoi la norme IPv6 va peu à peu remplacer celle d'IPv4, les adresses seront alors codées sur 64 bits.

• Les adresses IP sont gérées par L'ICANN (*Internet Corporation for Assigned Names and Numbers*) qui est la haute autorité d'Internet. Puis c'est l'IANA (*Internet Assigned Numbers Authority*) qui se charge de distribuer les adresses IP aux cinq différents RIR (*Regional Internet Registry*) qui allouent les adresses aux différents fournisseurs d'accès Internet (FAI). Un particulier obtient ensuite une connexion Internet en s'abonnant à un FAI. La France dépend du RIPE NCC, c'est le registre régional d'adresses IP qui dessert l'Europe et une partie de l'Asie, notamment le Moyen-Orient.

• Il existe des adresses IPv4 publiques et privées. **Les adresses IPv4 publiques sont uniques** à un instant donné sur Internet.

• Les adresses privées sont les adresses utilisées par les utilisateurs sur les réseaux domestiques et professionnels. Une adresse IPv4 provient d'un bloc d'adresses privées si elle :

- commence par 10, c'est-à-dire 10.0.0.0 à 10.255.255.255 ;

- commence par 172.16. jusqu'à 172.31 ;
- commence par 192.168.

• Les adresses IP sont en général délivrées de manière automatique grâce au protocole DHCP (installé dans une box Internet, par exemple).

• Ainsi, deux ordinateurs connectés à la même box Internet possèdent une adresse IP privée différente, mais ils possèdent la même adresse IP publique.

Déterminer une adresse IP

• Pour déterminer l'adresse IP privée de son ordinateur, on peut utiliser le logiciel cmd (invite de commandes) sur Windows ou le Terminal sur macOS. Il suffit alors de taper **ipconfig** (ou ifconfig pour Mac) et de valider à l'aide de la touche « Entrée » pour voir apparaître de nombreux renseignements sur sa machine, dont l'adresse IP privée, mais aussi l'adresse MAC, par exemple.

• Pour savoir si un appareil possédant une adresse IP est accessible, on peut utiliser la commande **ping**.
Par exemple ping -c 5 007.com va envoyer cinq petits paquets au serveur du site web 007.com et calculer le temps de propagation en boucle. Le Terminal affiche alors :

```
ping -c 5 007.com
PING 007.com (104.28.12.104): 56 data bytes
64 bytes from 104.28.12.104: icmp_seq=0 ttl=54 time=26.637 ms
64 bytes from 104.28.12.104: icmp_seq=1 ttl=54 time=28.197 ms
64 bytes from 104.28.12.104: icmp_seq=2 ttl=54 time=24.784 ms
64 bytes from 104.28.12.104: icmp_seq=3 ttl=54 time=23.091 ms
64 bytes from 104.28.12.104: icmp_seq=4 ttl=54 time=20.850 ms
--- 007.com ping statistics ---
5 packets transmitted, 5 packets received, 0.0% packet loss
round-trip min/avg/max/stddev = 20.850/24.712/28.197/2.585 ms
```

Routage

• Un appareil ne peut communiquer qu'avec un appareil du même **sous-réseau**. Les appareils d'un même sous-réseau sont reliés par des commutateurs (**switch**). Un appareil chargé de transmettre des paquets entre des réseaux différents est appelé **routeur** : il permet une communication entre réseaux.

• Lorsqu'un appareil veut envoyer un paquet à un autre, il va d'abord regarder si l'appareil est sur son propre sous-réseau. Si ce n'est pas le cas, il enverra son paquet à l'adresse « passerelle » d'un routeur qui effectuera ce que l'on appelle un **routage** : trouver une suite de routeurs pour acheminer le paquet vers sa destination.

• Un routeur contient une **table** (*forwarding table*) qui contient une liste de blocs d'adresses IP avec une liste de liens associés. Chaque paquet transite ainsi par un certain nombre de routeurs, chacun l'envoyant à un autre routeur selon sa carte locale et la destination prévue.

• La *forwarding table* d'un routeur peut être obtenue de plusieurs manières. L'utilisation de l'algorithme de Dijsktra est fréquente : les routeurs d'un même FAI élaborent un graphe du réseau avec les temps de parcours entre chaque routeur. Ils s'ajustent en permanence quand un routeur est ajouté au réseau ou quand un routeur voisin disparaît. Il n'y a pas besoin de carte globale, ce qui permet le routage à grande échelle. Dans le cas d'un paquet qui doit être envoyé d'un système autonome à un autre, certains routeurs utilisent le protocole eBGP (*external Border Gateway Protocol*) pour signaler aux autres routeurs les systèmes autonomes auxquels ils peuvent délivrer un paquet.

• On peut utiliser la commande **tracert** (ou **traceroute** pour macOS et Linux) pour visualiser le chemin emprunté pour joindre un appareil connecté à Internet.

Fiabilité de transmission mais absence de garantie temporelle

• Les paquets sont envoyés jusqu'à atteindre la destination, mais le temps d'acheminement n'est pas connu à l'avance. Un paquet pourrait ne pas arriver en raison d'une **panne matérielle** d'un lien ou d'un routeur, ou de sa **destruction**. Chaque paquet contient l'information d'un nombre maximal de routeurs à traverser (durée de vie). Pour ne pas encombrer le réseau, il est détruit si ce nombre est atteint. Enfin, il existe évidemment des possibilités d'attaque par saturation, lorsque l'on envoie un très grand nombre de messages à un destinataire, pour provoquer un déni de service.

Couche liaison

• Dans la **couche liaison**, les protocoles sont **r**esponsables de **trouver la route entre deux nœuds du réseau** (ordinateur à borne wifi ou routeur à routeur), de permettre une connexion à plusieurs appareils vers un routeur et de vérifier que les données ne sont pas corrompues (à l'aide du code CRC). **Ethernet et IEEE 802.11 (wifi) sont les deux principaux protocoles**. Est utilisé aussi le

protocole ARP qui ajoute des informations (**adresses MAC**) au paquet créé par la couche réseau et crée une **trame**.

Couche physique

• Dans la **couche physique** , on envoie l'intégralité des bits contenus dans la trame à travers le lien. (conversion des bits en signaux électriques).

Conclusion

On peut assimiler l'envoi d'une donnée sur Internet au déplacement d'un énorme meuble d'une maison A à une maison B. L'équipe de « déménageur TCP » va se charger de démonter le meuble et de numéroter les « pièces segments ». L'équipe de « transporteur IP » va se charger, à l'aide de « voitures paquets », de transporter les « pièces segments » de l'adresse IP de la maison A à l'adresse IP de la maison B. À chaque fois qu'une voiture arrive à la maison B, elle repart vers la maison A avec un accusé de réception, puis repart vers la B avec une nouvelle pièce... Un chemin de A vers B est composé de « routes liens » reliées entre elles par des « intersections routeurs ». À chaque intersection, un « panneau table » précise la route la plus rapide pour se rendre à une intersection proche. Évidemment les voitures n'emprunteront pas toujours le même chemin, car une route pourrait être en travaux, fermée ou détruite.

II. Adresses symboliques et serveurs DNS

• Les adresses sont numériques et hiérarchiques, mais l'utilisateur connaît surtout des **adresses symboliques** normalisées, comme lilletourism.com. Le système DNS (*Domain Name System*) transforme une adresse symbolique en adresse numérique IP. Il est réalisé par un grand nombre d'ordinateurs répartis sur le réseau et constamment mis à jour. C'est un système essentiel afin de ne pas avoir à saisir des adresses IP en permanence et cela permet également de remplacer un serveur par un autre sans que l'utilisateur n'ait à changer l'adresse.

• **Exemple :** 194.199.8.109 est l'adresse IP de l'adresse symbolique bnf.fr. 195.254.146.9 est l'adresse IP de l'adresse symbolique musee-orsay.fr.

• Il existe trois types de serveurs DNS : racines, TLD (*Top Level Domain*) et autorité.

• Il y a treize serveurs racines dans le monde ; ils connaissent les adresses IP des serveurs TLD qui sont responsables d'un nom de domaine de premier niveau (com, fr, org, be...).

• Les serveurs TLD connaissent les adresses IP des serveurs d'autorité.

• La commande **nslookup** permet de déterminer l'adresse IP d'un serveur hébergeant un site web et inversement. La commande **whois** permet d'obtenir plus d'informations sur l'appareil en question.

III. Réseaux pair-à-pair

• Contrairement à la configuration classique client-serveur (les données sont détenues par un serveur), chaque ordinateur sert à la fois d'émetteur et de récepteur. Les hôtes communiquent directement entre eux sans passer par un serveur, ils partagent un fichier constitué de plusieurs fragments répartis sur les pairs.

• Un client pair-à-pair commence par se connecter à un **tracker**, un serveur qui contient une liste des clients ayant au moins un fragment du fichier. Le protocole utilisé (BitTorrent par exemple) récupère alors les adresses IP de ces clients. Il contacte les clients en même temps afin de télécharger le fragment. Il reconstitue à la fin le fichier. Il diffuse également les fragments récupérés pendant ce temps.

• Il est aussi possible de créer des systèmes de **calcul distribué** comme dans le cas du projet BOINC (*Compute for Science*, par exemple), du projet WCG (*Help Fight Childhood Cancer*, par exemple) ou du projet français Décrypthon, qui propose de répartir des calculs extrêmement complexes sur un grand nombre d'ordinateurs personnels, chaque ordinateur ayant un morceau du calcul à effectuer, et tout cela dans le but de faire avancer la recherche. On peut cumuler la puissance de beaucoup de machines afin d'exécuter davantage de tâches, tout en diminuant les coûts des recherches. En effet, on économise les ressources électriques pour refroidir un centre de données (*datacenter*), et comme les pairs sont souvent proches on réalise une économie d'électricité sur les communications.

• Des **services de paiement** récents (PayPal, WhatsApp...) et certaines **cryptomonnaies** (comme le Bitcoin basé sur le principe de la *blockchain* : réplique de preuves chiffrées et vérifiables d'un ensemble d'informations enregistrées) utilisent également les réseaux pair-à-pair. Quelques services de VoIP utilisent en partie des réseaux pair-à-pair.

• Il est possible d'utiliser un réseau pair-à-pair de manière illicite et illégale en partageant des fichiers sans autorisation (documents dont on ne possède pas de droits de diffusion).

IV. Indépendance d'Internet par rapport au réseau physique

• En 2018, le trafic était de 50 000 Go par seconde avec 4,2 milliards d'utilisateurs. Le trafic prévu pour 2021 est de 3 300 milliards de milliards d'octets ($3{,}3 \times 10^{21}$ octets).

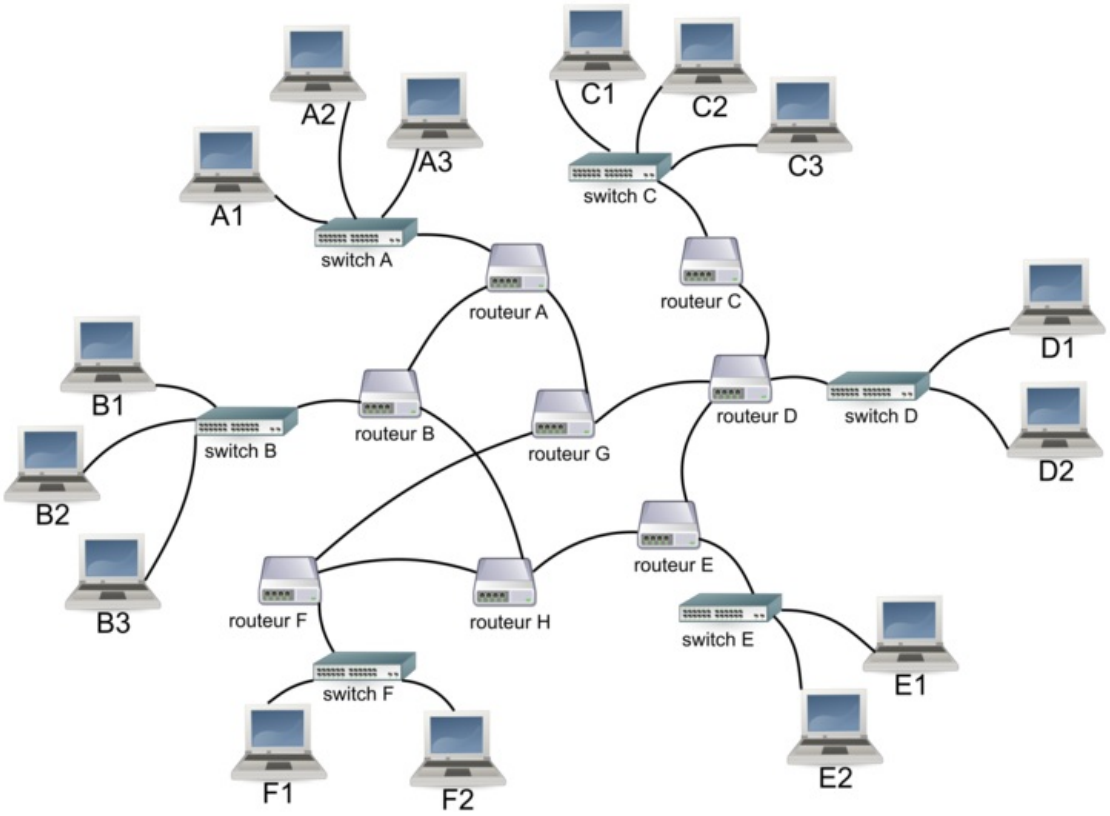
• Les protocoles d'Internet étant des logiciels, ils peuvent s'appuyer sur n'importe quel réseau physique qui les implémente : 4G,

Ethernet, ADSL, wifi, Bluetooth... La majorité du trafic passe par des câbles sous-marins (submarinecablemap.com). À l'intérieur d'un pays, on trouve de plus en plus de fibres optiques (pour le très haut débit) et des antennes 4G (bientôt 5G).

• La neutralité du Net, présente dès l'origine du réseau, signifie que les routeurs doivent transmettre les paquets indépendamment du type de leur contenu (texte, vidéo...) à la même vitesse et sans filtrage. Mais elle est constamment remise en cause par certains *lobbies* industriels.

V. Simulation d'un réseau

• À l'aide d'un logiciel libre et gratuit comme Filius, on peut simuler plusieurs réseaux locaux reliés entre eux par des routeurs. Il s'agit d'une sorte de mini-Internet. On peut également utiliser les commandes ping, traceroute, ipconfig depuis des ordinateurs du réseau afin de visualiser le trajet des paquets (les câbles sollicités deviennent verts). Il est en outre possible de simuler un serveur DNS.

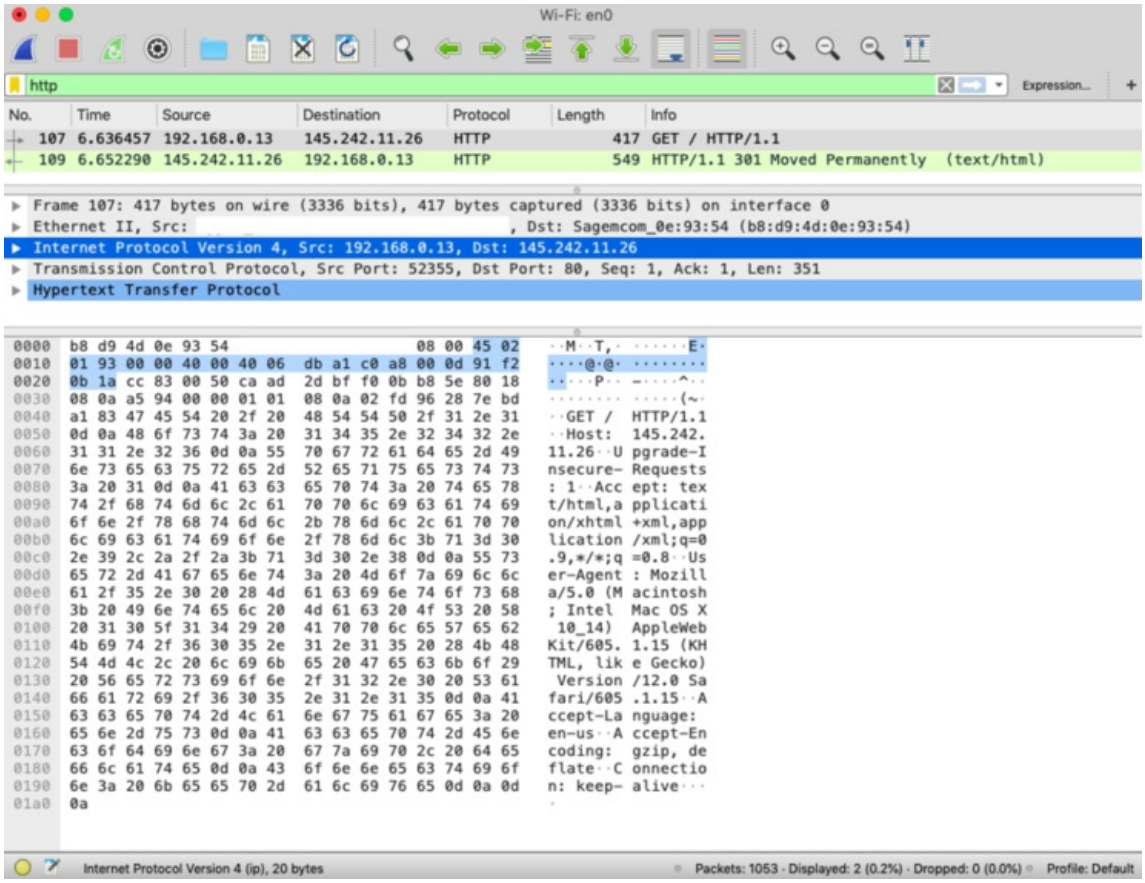


VI. Visualisation des paquets

• Pour visualiser les échanges de paquets entre deux appareils, on peut utiliser un logiciel libre et gratuit comme Wireshark (disponible ici wireshark.org/#download).

• Lancer le logiciel, dans l'onglet « Capture » sélectionner « Démarrer », se connecter à un site web de son choix, puis dans l'onglet « Capture » sélectionner « Arrêter ». Cliquer sur la première ligne dont le protocole est HTTP. Dans la deuxième partie de la fenêtre, sélectionner le menu déroulant « HyperText Transfer Protocol », vérifier que la première commande est bien du type GET. Puis sélectionner le menu déroulant « Internet Protocol » et repérer l'en-tête que le protocole IP a ajouté au segment pour obtenir un paquet (on dit aussi un datagramme). On repère évidemment l'adresse IP privée dans le sous-réseau (192.168.0.13) et l'adresse IP du site web auquel on veut accéder (145.242.11.26).

• On visualise la trame finale, puis en survolant les différentes couches on repère les informations qui ont été ajoutées. L'affichage est en hexadécimal (pour faciliter la lecture), mais les informations sont bien encodées en binaire avant la transmission par électricité ou par les ondes.



• L'adresse MAC de l'émetteur a été remplacée sur ce document par un rectangle blanc. On peut cependant toujours lire l'adresse MAC du destinataire : b8 :d9 :4d :0e :93 :54.

Exercice n°1

Exercice n°2

Exercice n°3

Exercice n°4

Exercice n°5