

Décrire et manipuler un document numérique

Cours 4 : Internet et le Web

Loïc Grobol <lgrobol@parisnanterre.fr>

2022-02-07

Compléments

Dans ce cours, on va faire une très, très rapide introduction aux concepts réseaux et web

Par nécessité on ne fera que gratter la surface de la partie émergée de l'iceberg et il vous restera après beaucoup de choses à apprendre ou à préciser sur le fonctionnement d'Internet et du Web. N'hésitez surtout pas à vous documenter de votre côté.

Voici quelques points de départ :

- Vidéo « comment fonctionne Internet »
- Le très complet et bien sourcé « *History of the Internet* » de Wikipedia:en
- « Le fonctionnement du Web » et « Le fonctionnement d'Internet » sur MDN web docs.

Une très brève histoire d'Internet

- Années 1940—1950 : les ordinateurs sont **massifs** et peu nombreux. Pour des raisons de facilités on y accède donc depuis des terminaux distants
 - De plus en plus distants
 - On développe petit à petit des technologies pour augmenter la vitesse et la fiabilité des communications numériques à longue distance
- Années 1960 : l'idée d'interconnecter des ordinateurs entre eux pour le partage d'informations apparaît graduellement

-
- 1969 : premier réseau effectif entre machines, ARPANET, financé par l'Agence pour les projets de recherche avancée de défense (DARPA) du département de la Défense des États-Unis
 - Rapidement suivi par d'autres initiatives ailleurs dans le monde
 - 1970–2000 : émergence de nouveaux réseaux locaux et régionaux, progressivement fusionnés pour donner un réseau global, Internet.

Réseaux de machines

Définitions

Réseau : « Ensemble de systèmes informatiques reliés entre eux pour communiquer et partager des données » :

- La connexion peut se faire physiquement (câble, fibre optique) ou par ondes radio (wifi).
- Relation client/serveur :
 - Les serveurs fournissent des ressources.
 - Les clients accèdent aux ressources.
- Relation d'égal à égal (*peer to peer*) :
 - Tous les systèmes informatiques connectés sont serveurs et clients.

Types de réseaux

Mettre deux machines en réseau, c'est (conceptuellement) simple, on appelle ça une connexion point-à-point :

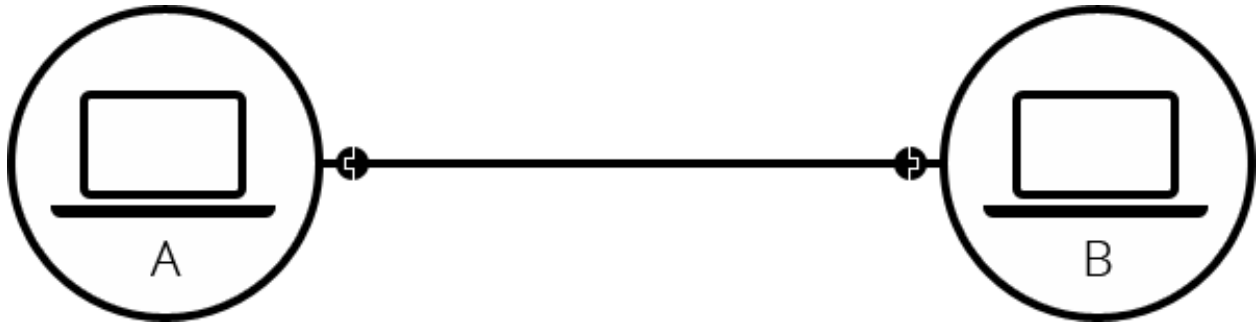
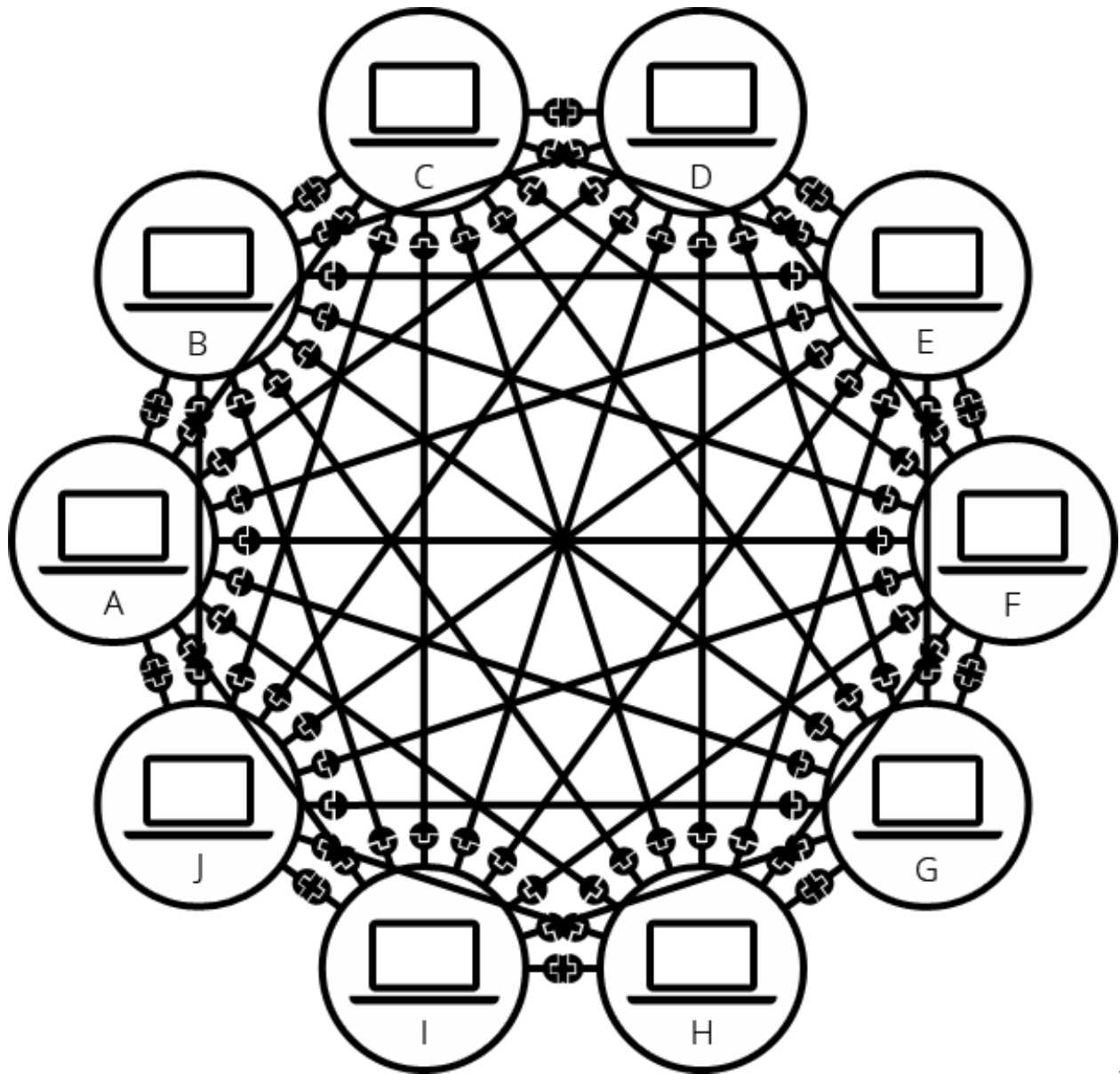


FIGURE 1 – Deux ordinateurs, nommés A et B reliés par un câble

Et théoriquement, c'est possible de le faire pour autant de machines qu'on veut :



Mais ça fait beaucoup de câbles et de prises. Combien d'ailleurs ?

Pour éviter ça, on peut plutôt avoir une machine qui sert de centre. Le plus souvent une machine spécialisée, qu'on appelle *routeur*.

On parle alors de réseau *en étoile*. Ça permet de simplifier considérablement le réseau. Dans ce cas, quand A passe un message à B, il le passe d'abord au hub (en lui disant que le destinataire est B) et le routeur le passe à B.

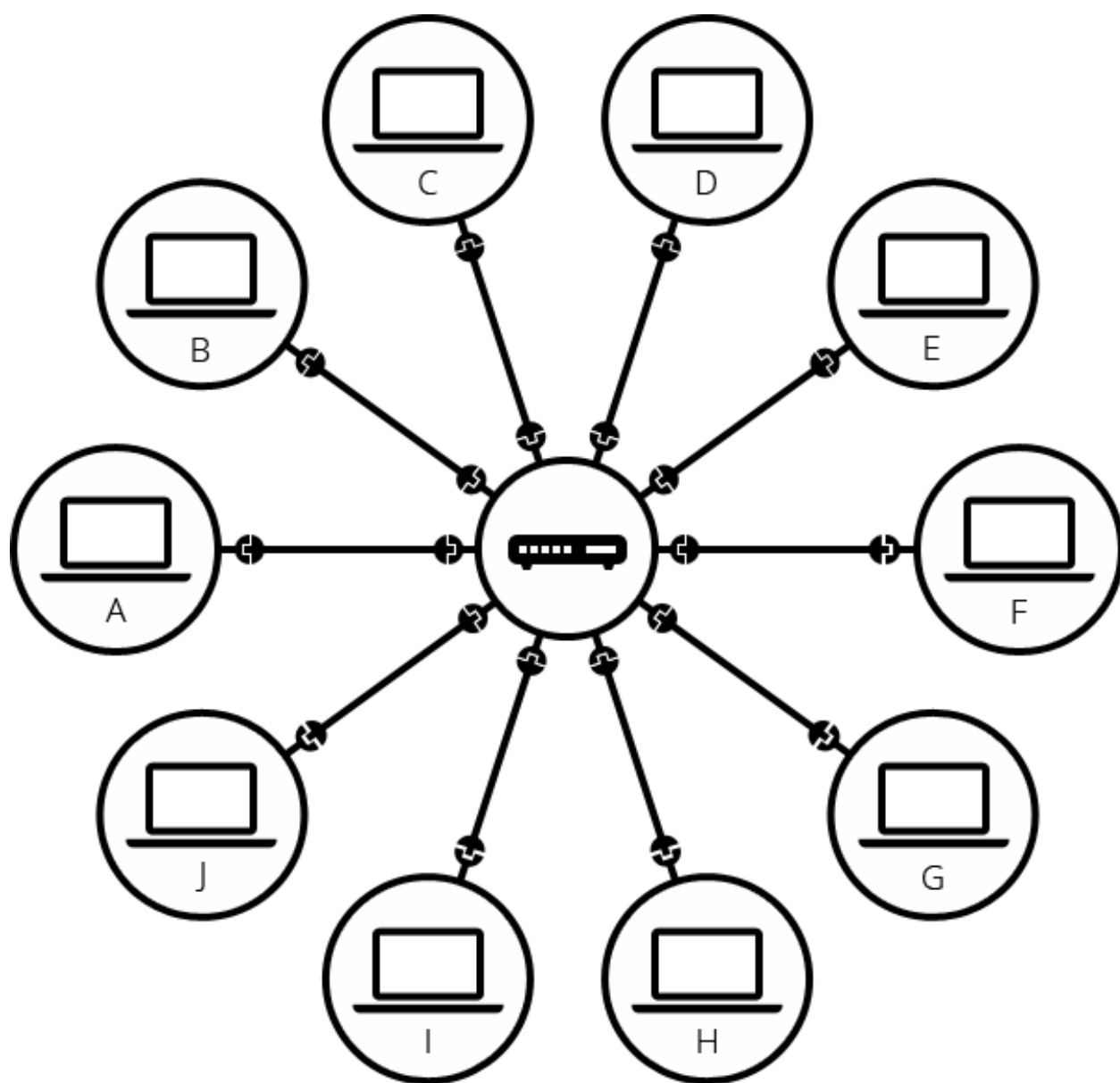
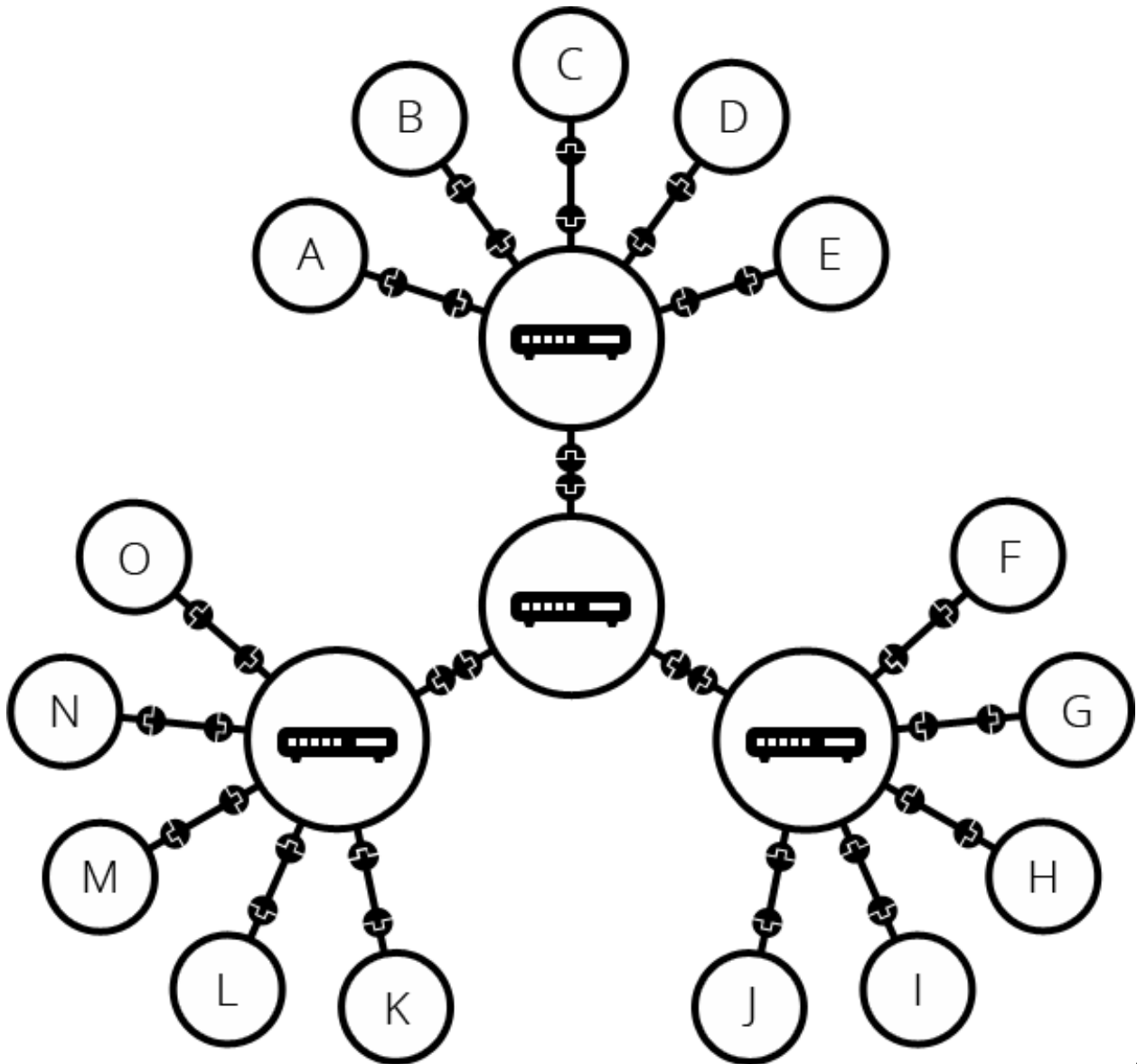


FIGURE 2 – Dix ordinateurs, nommés A, ..., J, disposés en cercle et tous reliés à une machine centrale

Il existe d'autres types de réseaux (on parle de *topologie*), le plus important étant les réseaux en anneau, qui ne nécessitent pas de centre, mais sont moins efficaces.

On peut aussi connecter entre eux des réseaux de ce type, en formant des arbres :



Cette idée de réseaux locaux interconnectés en arborescence pour former un réseau global est le principe de base d'Internet.

En pratique, c'est un peu plus complexe que ça, avec notamment des liens redondants qui le rendent plus modulable en fonction du trafic et plus robuste en cas d'interruption.

Les réseaux sont organisés hiérarchiquement :

- Les LAN (*Local Area Network*) : Réseau local
 - À l'échelle d'un bâtiment ou d'un complexe
- Les WAN (*Wide Area Network*) : Réseau étendu
 - Couvrent des zones géographiques étendues : pays, voire continent
- Les GAN (*Global Area Network*) : Réseau global → Internet

On peut aussi construire des sous-réseaux en s'appuyant sur l'architecture globale : par exemple avec un VPN (*Virtual Private Network*).

Adresse IP et noms de domaine

Chaque machine est identifiée sur le réseau Internet de manière unique par son adresse IP.

Vous pouvez trouver la vôtre en utilisant par exemple <https://whatismyipaddress.com/>.

Deux versions coexistent :

- IPv4 : Combinaison de 4 chiffres allant de 0 à 255 (ex : 128.11.25.2) codé sur 32 bits
- IPv6 : Créé pour éviter la saturation de l'IPv4. Combinaison de 8 groupes de 4 chiffres en hexadécimal sur 128 bits. (5800:10C3:E3C3:F1AA:48E3:D923:D494:AAFF)

L'adresse est attribuée par le FAI (Fournisseur d'Accès Internet) : Orange, SFR, Renater...

Il n'est évidemment pas pratique de travailler directement avec les adresses IP pour accéder à des machines publiques.

- Complicées à retenir.
- Ne permet pas d'héberger plusieurs services sur une même machine.
- Ni un même service sur plusieurs machines.

On fonctionne donc avec un système de *noms de domaines* qui ajoute une couche d'indirection : pour accéder à une machine on demande à un serveur de noms de domaines (DNS) l'adresse qui correspond à une chaîne de caractères : son nom de domaine. Comme :

- <http://parisnanterre.fr>
- <http://plurital.org>
- <http://w3c.org>

Services et protocoles

Internet est donc un **réseau**, utilisable pour des activités supposant des communications entre machines :

- Web
- Courrier électronique
- Messagerie instantanée

- Jeux vidéos en ligne
 - ...
-

Ces communications sont encodées en utilisant une **pile de protocole** :

- La couche *applicative* : communications entre logiciels, applications, services... (HTTP, TLS, IMAP...)
 - La couche *de transport* : communication entre machines connectées (TCP, UDP...)
 - La couche *internet* : gestions des connexions entre réseaux (IP)
 - La couche *de lien* : implémentation matérielle/physique des réseaux (Ethernet, wifi...)
-

Chaque protocole permet de gérer les communications à un certain niveau de granularité. En principe, un protocole :

- Traite les couches inférieures comme une boîte noire.
 - Ignore l'existence des couches supérieures.
-

Le Web

URL

Le **World Wide Web** ou web, est un système d'information reposant sur Internet.

Il est constitué de documents et de ressources, chacune possédant des *Universal Resource Locators* (URL) : des chaînes de caractères qui permettent d'y accéder de façon stable depuis n'importe quel point d'accès.

Une URL est habituellement de la forme :

```
http://example.com/path/page.html
```

- **http** désigne le **protocole applicatif** de mise à disposition
 - **example.com** est le **nom de domaine associé** à une machine qui met la ressource à disposition
 - **path/page.html** est un **chemin interne** à cette machine
 - Peut correspondre ou non à un chemin dans son système de fichiers.
-

Une URL peut également comporter un sous-domaine : par exemple dans le cas de `https://loicgrobol.github.io/document-numerique`, **loicgrobol** désigne un sous-domaine du nom de domaine **github.io**.

Le web est construit sur le concept d'**hypertextualité** :

- Les documents sont prévus pour être des documents textuels.
- Liés par des *liens hypertexte* : des URL de documents ou de ressources liées.

Ainsi, ce qu'on appelle un *site web* ou *site Internet* est conceptuellement simplement un ensemble cohérent de documents textuels (ou *pages web*) liées par des liens hypertextes.

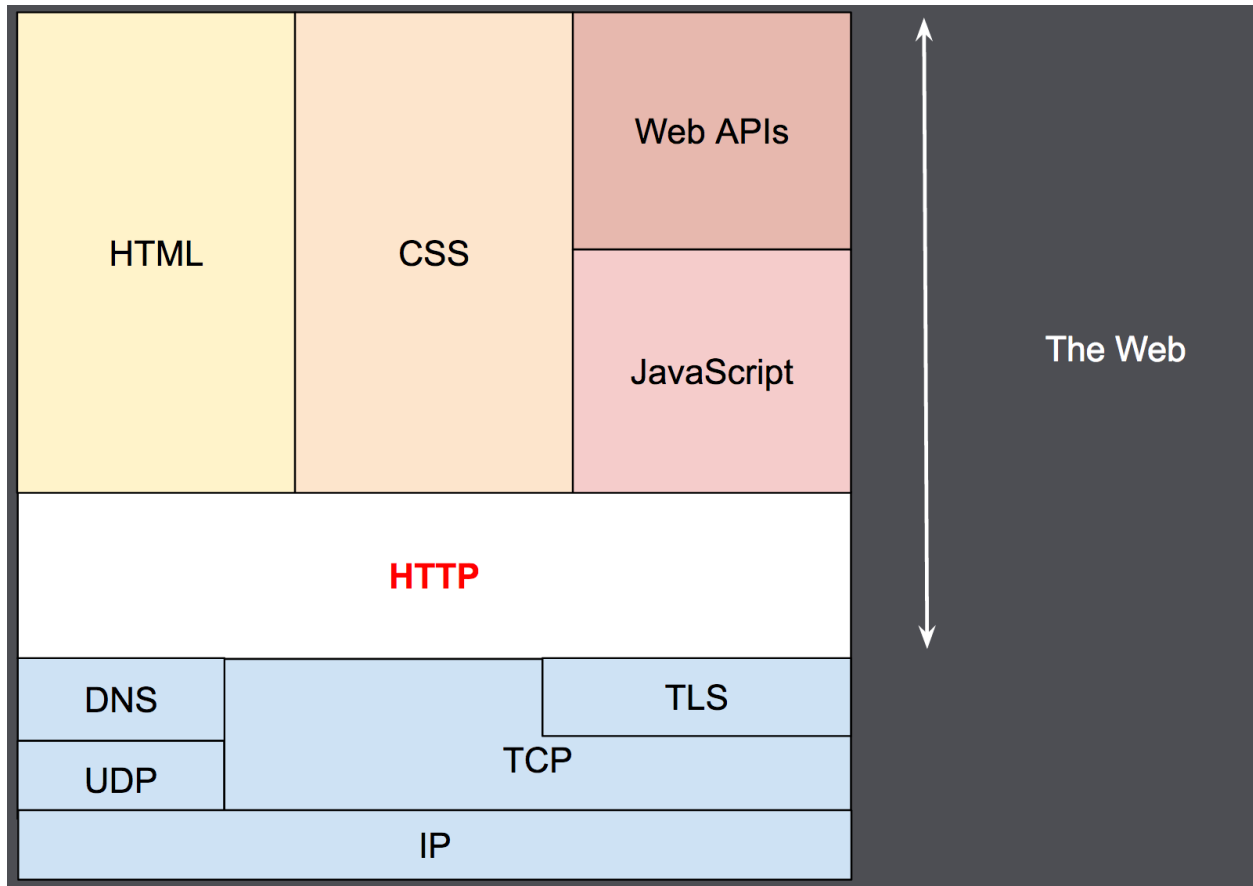


FIGURE 3 – Représentation de la pile de protocoles internet : IP supporte UDP et TCP, UDP supporte DNS, TCP supporte TLS qui supporte HTTP, qui supporte HTML, CSS et JavaScript, qui supporte les API web

Pages, sites, navigateurs

Pages et sites

Une **page web** est un document électronique accessible sur le web.

Elles sont souvent constituées d'une ressource centrale : un document au format HTML et d'autres ressources

- Multimédia : images, vidéos, sons...
- Style : un fichier au format CSS qui décrit l'apparence du document.
- Scripts : un programme dans le langage JavaScript qui s'exécute sur la machine cliente et rend la page interactive.

Un **site web** est un ensemble de pages web hiérarchisées.

Moteurs de recherche

Un **moteur de recherche** est un site web permettant de chercher des pages webs.

Quelques exemples :

- <https://google.com>
 - <https://duckduckgo.com>
 - <https://www.bing.com>
 - <https://www.startpage.com>
-

Navigateurs

Un **navigateur web** est une application permettant d'accéder à et d'afficher des pages web. Quelques exemples

- Mozilla Firefox
- Chromium
- Lynx