

Introduction au fonctionnement d'Internet : infrastructure, protocoles, applications

Table des matières

Objectifs	3
Introduction	4
I - Principe de la communication sur Internet	5
II - Infrastructure réseau d'Internet	7
III - Ordinateurs clients	8
IV - Ordinateurs serveurs	9
V - Ordinateurs routeurs	11
VI - Protocoles réseaux TCP/IP	12
VII - Adresses IP et noms de domaine	15
VIII - Protocoles applicatifs d'Internet : SSH, mail, Web...	16
IX - Quelques logiciels qui font fonctionner Internet	19
X - URL : Uniform Resource Locator	20
XI - Les clients et les serveurs web parlent HTTP	22
XII - Exercice	24
XIII - Cours-émission	28
Solutions des exercices	29
Abréviations	30
Crédits des ressources	31

Objectifs



- Savoir définir ce qu'est un client, un serveur, un routeur, un protocole
- Connaître les bases du protocole TCP/IP
- Savoir ce qu'est une adresse IP et un nom de domaine
- Savoir associer les principales applications d'Internet (SSH, mail et Web a minima), leurs principaux protocoles (SSH, IMAP, SMTP et HTTP) et leurs principaux outils (terminal, client mail, navigateur web)

Introduction



Ce cours est une introduction technique à la notion de communication sur Internet : quels protocoles et outils permettent de communiquer via Internet ?

Principe de la communication sur Internet



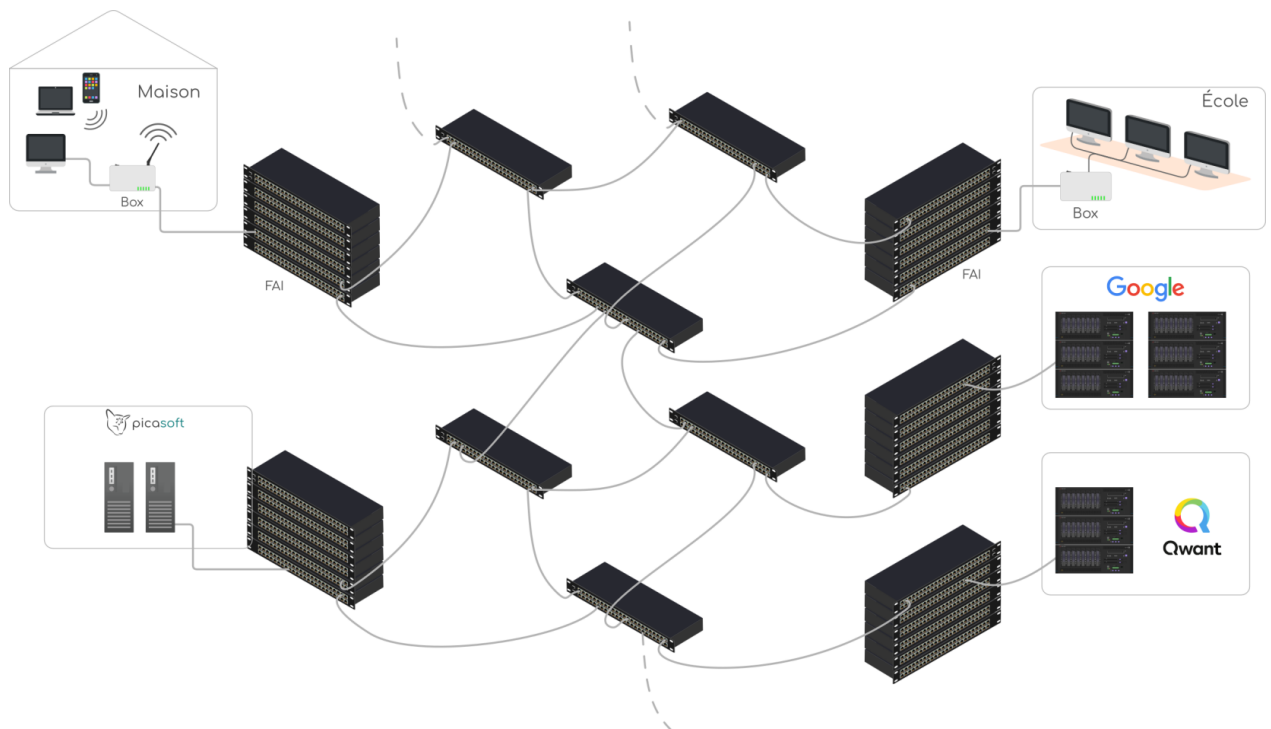
Une communication via Internet implique :

- les ordinateurs des personnes qui communiquent, on les appelle des **clients** ;
- les **routeurs**, ce sont des ordinateurs dont le rôle est de faire passer l'information d'un ordinateur à un autre sur Internet ;
- en général des **serveurs**, ce sont des ordinateurs en charge de gérer la communication pour un ensemble de clients ;
- des moyens physiques pour **relier** ces ordinateurs (câbles et ondes) ;
- des **protocoles** qui permettent à des logiciels installés sur ces machines de communiquer les uns avec les autres.

Internet est un vaste réseau d'ordinateurs qui communiquent entre eux



Pour que les **clients**, les **routeurs** et les **serveurs** puissent communiquer entre eux, ils faut qu'ils soient **reliés** physiquement et qu'ils hébergent des logiciels qui respectent des **protocoles** communs.



Des machines et des logiciels (confusing !)



Les termes de serveur et de client sont utilisés pour des machines, mais également pour des logiciels.

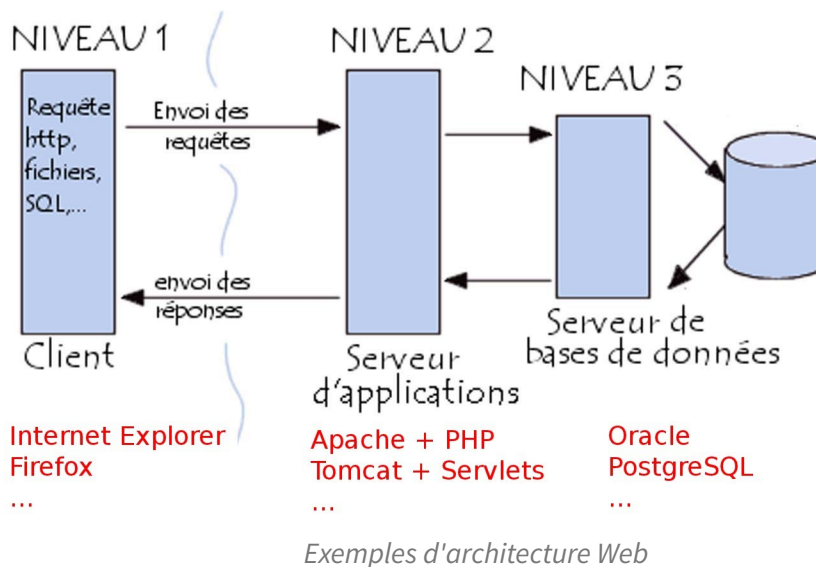
Ainsi Apache est un serveur web, c'est à dire un logiciel destiné à tourner sur une machine serveur et donc à répondre à des requêtes HTTP, tandis que Firefox est un client web, c'est à dire un logiciel destiné à tourner sur une machine cliente et donc à envoyer des requêtes HTTP.

Des clients qui sont aussi des serveurs...



Il est possible qu'une machine soit à la fois un client et un serveur :

- Si vous avez un ordinateur chez vous, vous pouvez tout à fait héberger un serveur web (qui est accessible quand la machine est allumée et connectée à Internet) et vous pouvez aussi vous en servir comme ordinateur personnel (donc par exemple pour consulter des sites web).
- Dans les architectures à plusieurs couches (architectures *n-tiers*) un ordinateur peut à la fois répondre à des requêtes et en formuler : il est ainsi courant qu'un serveur web ait besoin de demander des informations à un serveur de base de données, dans ce cas la machine est serveur du point de vue du navigateur qui lui demande une page web et client du point de vue de la base de données à laquelle il demande les informations (qui vont lui permettre de créer la page web à renvoyer au client).



Comment fonctionne Internet ?

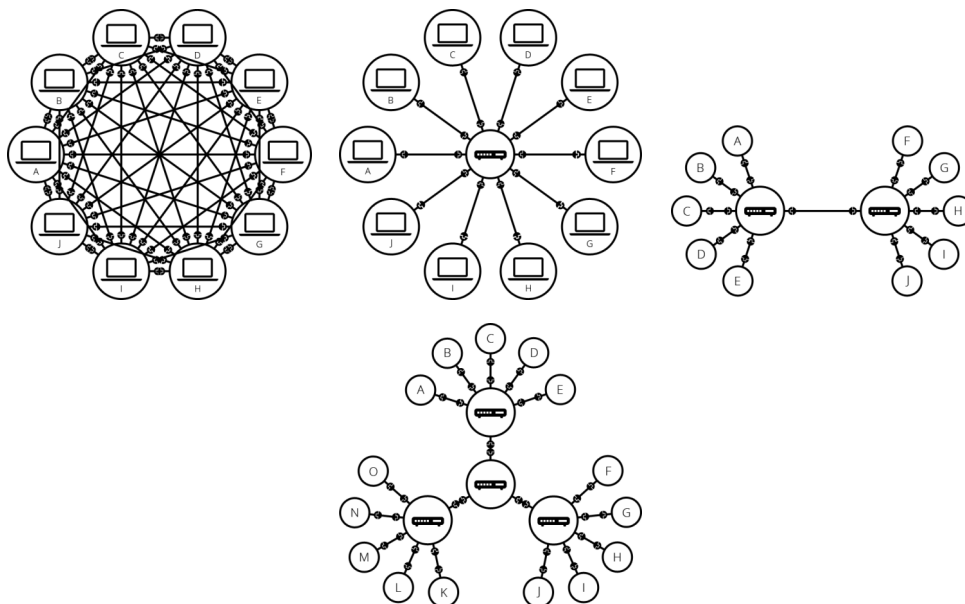
Les machines qui font fonctionner Internet

Le fonctionnement de l'Internet



https://developer.mozilla.org/fr/docs/Learn/Common_questions/How_does_the_Internet_work

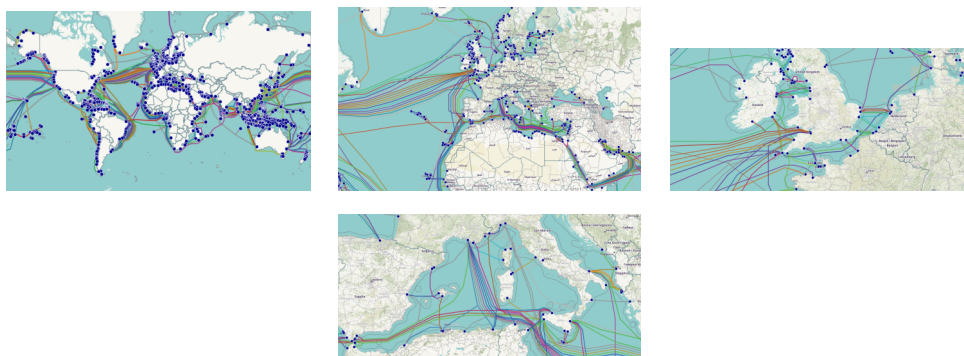
Infrastructure réseau d'Internet



Internet est un réseau de réseaux

Fondamental

Les composants fondamentaux du réseau Internet sont les câbles très haut débit par lesquels transitent les données.



Carte des câbles sous-marins du réseau Internet

Complément

Les acteurs du réseau : FAI, FCA, hébergeur, transitaires

Ordinateurs clients



Client



Définition

Un ordinateur client est un ordinateur qui demande un service à un autre ordinateur, par exemple : envoie ce mail à telle adresse de ma part.

Tout ordinateur qui utilise des services sur Internet est un client.

Envoyer un mail



Exemple

Si j'envoie un mail à *contact@picasoft.net* depuis mon mail *stephane.crozat@utc.fr* :

- Mon ordinateur est le client, via le logiciel de mail Thunderbird il demande au serveur mail de l'UTC d'envoyer mon mail au serveur mail de Picasoft (il utilise pour cela le protocole SMTPS).
- Ensuite quand quelqu'un chez Picasoft consultera les mails du compte *contact*, il demandera à son serveur quels nouveaux mails il a reçu et il y aura le mien parmi ceux-là (il utilise pour cela le protocole IMAPS).

Demander une page web



Exemple

Si je consulte la page <https://fr.wikipedia.org/wiki/Internet> :

- Mon ordinateur est le client, via le navigateur web Firefox il demande au serveur web de Wikipédia de m'envoyer une page HTML (il utilise pour cela la requête GET du protocole HTTPS).
- Le serveur web de Wikipédia renvoie la page HTML demandée (il utilise pour cela une réponse HTTPS).

Ordinateurs serveurs



Serveur



Un serveur est un ordinateur dont le rôle est de proposer des services, par exemple : stocker des pages web ou acheminer des mails.



- Les serveurs sont souvent des ordinateurs sans écran et sans clavier situés dans des datacenters, par exemple Picacoft dispose de deux serveurs dans un datacenter de Toulouse géré par l'association Tetaneutral. Aucun membre de Picasoft n'a jamais rencontré physiquement ces machines.
- Chacun peut installer un serveur chez lui ou transformer son ordinateur personnel en serveur, par exemple <https://pi.crzt.fr> est un serveur web que j'héberge sur un Raspberry Pi chez moi.



<https://tecnologiahabitual.blogspot.com/2014/07/mantenimiento-en-datacenters.html>

Exemple de serveur



L'UTC propose un serveur pour envoyer des mails en utilisant le protocole SMTPS, il s'appelle smtps.utc.fr.

```
1 stph@mypc:~$ ping smtps.utc.fr
2 PING smtps.utc.fr (195.83.155.8) 56(84) bytes of data.
3 64 bytes from smtps.utc.fr (195.83.155.8): icmp_seq=1 ttl=62 time=0.454 ms
4 64 bytes from smtps.utc.fr (195.83.155.8): icmp_seq=2 ttl=62 time=0.752 ms
5 64 bytes from smtps.utc.fr (195.83.155.8): icmp_seq=3 ttl=62 time=0.739 ms
6 64 bytes from smtps.utc.fr (195.83.155.8): icmp_seq=4 ttl=62 time=0.726 ms
```

La commande *ping* permet de tester l'accessibilité d'une machine sur Internet.

Ordinateurs routeurs



Routeur



Définition

Un routeur est un ordinateur dont le rôle est d'acheminer des paquets d'information sur Internet, on peut voir cela comme un coursier.



Remarque

Pour qu'un message se déplace d'un ordinateur à un autre sur Internet, il devra passer par plusieurs routeurs, de plus en plus proches de sa destination.

traceroute from pic.crzt.fr to www.utc.fr

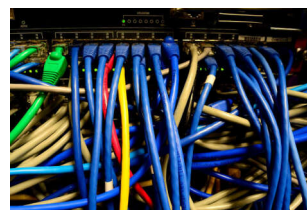
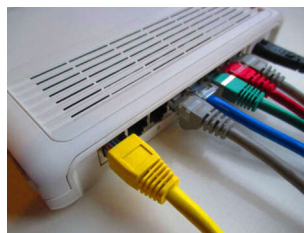


Exemple

```
1 1 h7.tetaneutral.net (91.224.148.1)
2 2 renater.par.franceix.net (37.49.236.19)
3 3 te0-0-0-1-ren-nr-compiegne-rtr-091.noc.renater.fr (193.51.177.53)
4 4 rrtt-vl221-te0-0-0-0-compiegne-rtr-011.noc.renater.fr (193.51.181.189)
5 5 * * *
6 6 * * *
7 7 gw-utc.utc.fr (195.83.154.242)
```



Exemple



Routeurs

Protocoles réseaux TCP/IP



Afin de pouvoir communiquer entre elles toutes les machines d'Internet respectent des règles communes. Ces règles sont standardisées au sein de *protocoles*.

Le fonctionnement d'Internet repose fondamentalement sur les protocoles réseaux IP et TCP.

Protocole IP



Le protocole (ou couche) IP gère les communications entre plusieurs machines d'un réseau.

Elle est basée sur l'**adressage**, le **routage** et la **fragmentation**.

- Pour s'adresser à une machine à laquelle on n'est pas directement connecté, il faut un moyen de l'identifier : c'est son **adresse**.
- Pour arriver jusqu'à cette machine, il faut trouver le chemin qui nous sépare d'elle : c'est le **routage**.
- Si le message à transmettre est trop gros pour les couches inférieures, il faut le découper : c'est la **fragmentation**.

Paquet IP



Les entités transmises par le protocole IP sont des **paquets** (ou datagramme).

Un message sur Internet est généralement découpé en plusieurs paquets qui sont transmis indépendants par le protocole IP.

Le protocole TCP est en charge de s'assurer que tous les paquets sont transmis et remis dans le bon ordre.





	1	2	3	4		
1	Version	IHL	Type of service	Total length		4
5	Identification		Flag	Fragment offset		8
9	Time to live	Protocole	Header checksum			12
13	Source IP adresse				16	
17	Destination IP adresse				20	
21	Options			Padding		24
25	DATA				28	
29					32	
33					36	
37					40	
41					44	
45					48	
49					52	
53					56	
57					60	
61					64	
65					68	
69					72	
73					76	
77					80	
81					84	
85					88	
89					92	
93	96					
97	100					
101	104					
105	108					
109	112					
113	116					
117	120					
121	124					
125	128					

Codage d'un paquet IP

Protocole TCP



Le protocole (ou couche) TCP gère les communications **de bout en bout**.

Elle est basée sur la **correction des erreurs**, le **ré-ordonnancement** et le **contrôle de flux**.

- Plutôt que de raisonner en terme de paquets, elle raisonne en terme de messages complets.
- Elle est capable de **corriger des erreurs** dans la transmission des messages (congestion du réseau...) et de **ré-ordonner** des paquets qui auraient pris une route différente et seraient arrivés dans le mauvais ordre.
- Ses mécanismes de **contrôle de flux** permettent d'adapter la vitesse d'envoi des données à la capacité de réception du destinataire.

Message



Les entités transmises par la couche TCP sont des **messages**.



Sur la base du protocole de communication TCP/IP on peut bâtir des applications, tels que HTTP pour consulter des fichiers sur le Web ou FTP pour transférer des fichiers.



Les couches réseaux du modèle OSI

Adresses IP et noms de domaine



Adresses IP



Une adresse IP est un **numéro d'identification d'une machine**.

Il est attribué de façon à chaque appareil relié à Internet.

Les adresses ont un format différent selon la version d'IP utilisée :

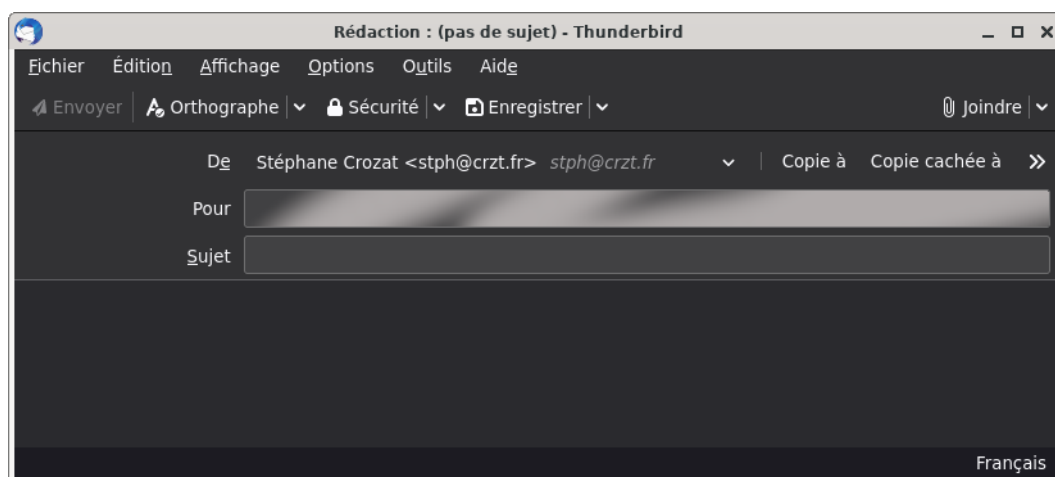
- pour IPv4, les adresses sont sur 4 octets (exemple : 192 . 168 . 3 . 200),
- pour IPv6, les adresses sont sur 16 octets (exemple : 2001:0db8:0000:85a3:0000:0000:ac1f:8001).

Nom de domaine



Un nom de domaine est une **adresse symbolique** permettant d'éviter de devoir retenir des adresses IP.

Ainsi, à un nom de domaine tel que wikipedia.org on peut associer une adresse telle que 91.198.174.192.



DNS



Un **serveur DNS** (pour *Domain Name System*) est un serveur qui a connaissance de la **correspondance entre noms de domaines et adresses IP**. Il est capable de répondre à des requêtes qui lui demandent de **résoudre** un nom de domaine en une adresse IP.



Adresses IP

Noms de domaine

Protocoles applicatifs d'Internet : SSH, mail, Web...



Protocole



Un protocole est un langage standard que tous les ordinateurs qui veulent communiquer entre eux respectent.

Les protocoles sont la condition de fonctionnement d'Internet, ils permettent à des ordinateurs différents qui utilisent des logiciels différents de communiquer entre eux.

Qui établit les protocoles ?



Internet est géré de façon décentralisée, il n'y a pas une structure qui gère l'ensemble, les protocoles sont souvent le résultat de discussions ouvertes à toutes et tous.

En pratique il y a quelques organismes de standardisation, comme l'*IETF* ^{p.30} qui publie les RFC ou le W3C qui standardise les langages du Web, comme HTML ou CSS.

Internet ≠ Web (métonymie)



Le protocole principal du Web est HTTP.

Il existe d'autres protocoles sur Internet : FTP pour le transfert de fichier, SSH pour l'accès à des ordinateurs distants, SMTP et IMAP pour le mail, etc.

Il y a donc plusieurs applications sur Internet qui ne sont pas du Web.

Notons néanmoins que certaines applications non web à l'origine ont aussi une couche web aujourd'hui :

- si vous utilisez Thunderbird, vous faites du mail sur Internet (vous utilisez les protocoles STMP et IMAP), mais pas du Web ;
- si vous faites du webmail, vous faites du Web et du mail (HTTP, IMAP, SMTP) ;
- si vous communiquez via la messagerie de Twitter, vous faites du Web.

Connexion distante (1969)



- Naissance : 1969
- Protocole : Telnet (1969, RFC 15), SSH (1995, RFC 4251)


```

Terminal - stph@pi-stole: ~
Fichier  Édition  Affichage  Terminal  Onglets  Aide
stph@hal9017:~$ ssh stph@192.168.1.97
Enter passphrase for key '/home/stph/.ssh/id_rsa':
Linux pi-stole 5.10.103-v7l+ #1529 SMP Tue Mar 8 12:24:00 GMT 2022 armv7l

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Tue Sep 20 11:41:11 2022 from 192.168.1.53

Wi-Fi is currently blocked by rfkill.
Use raspi-config to set the country before use.

stph@pi-stole:~ $ whoami
stph
stph@pi-stole:~ $
    
```

Mail (1971)

? Exemple

- Naissance : 1971
- Protocoles : SMTP (1982, RFC 821¹) — POP (1985, RFC 937²) — IMAP (2003, RFC 3561³)

Forum (1979)

? Exemple

- Naissance : *usenet*, 1979
- Application web : Discourse (décentralisé, libre)
- → Médias sociaux

Messagerie instantanée (1988)

? Exemple

- Naissance : commande Unix *talk* (années 1970)
- Protocole : IRC, 1988
- Applications web de messagerie instantanée :
 - Slack (centralisé, propriétaire)
 - Mattermost (décentralisé, libre)
 - Element/Matrix (décentralisé, libre, fédéré, chiffré E2E)

Web (1990)

? Exemple

- Naissance : 1990
- Protocole : HTTP (1990, W3C, RFC 1945⁴) (et le langage de formatage de documents HTML)

1. <https://tools.ietf.org/html/rfc821>

2. <https://tools.ietf.org/html/rfc937>

3. <https://tools.ietf.org/html/rfc3501>

4. <https://www.w3.org/Protocols/rfc1945/rfc1945.txt>

RFC



<https://www.ietf.org/standards/rfcs>⁵

<https://www.bortzmeyer.org>⁶



Internet et ses protocoles

FTP : File Transfer Protocol

HTTP : HyperText Transfert Protocol

SSH - Secure Shell (introduction)

⁵. <https://www.ietf.org/standards/rfcs/>

⁶. <https://www.bortzmeyer.org/>

Quelques logiciels qui font fonctionner Internet



Logiciels côté clients

- Navigateur web : Mozilla Firefox ou Google Chrome (clients HTTP)
- Client Mail : Mozilla Thunderbird (client IMAP ou POP)

```
Terminal - stc@pclf001215: ~
Fichier  Édition  Affichage  Terminal  Onglets  Aide
stc@pclf001215:~$ ssh stph@myserver.utc.fr
```

Client SSH sous Linux

Logiciels côté serveur

- Serveur web : Apache, Nginx
- Serveur mail : Postfix (SMTP), Dovecot (IMAP)

Typologie de logiciels



- propriétaire + centralisé (Zoom, Discord, Youtube)
- libre + centralisé (Signal)
- libre + décentralisé (Mattermost, BigBlueButton, Etherpad)
- libre + décentralisé + fédéré (Mastodon, Element, Mobilizon, Peertube)

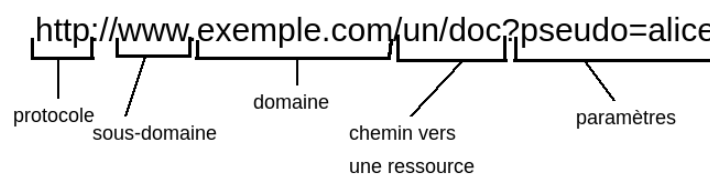


URL : Uniform Resource Locator

Les URL



Les URL sont nées en même temps que le World Wide Web et permettent d'identifier une ressource sur un serveur web.



Décomposition d'une URL

Elles contiennent :

- Le **protocole** à utiliser pour accéder à la ressource, ici HTTP,
- L'**adresse du serveur** à rejoindre, ici le nom de domaine `www.exemple.com`,
- Le **chemin de la ressource** dans le serveur,
- Éventuellement des paramètres que l'on veut envoyer au serveur (utilisés par les requêtes HTTP GET).



Lorsque `https://www.wikipedia.org/wiki/URL` est rentrée dans un navigateur web, ce dernier fait une requête GET sur la ressource « `/wiki/URL` » du serveur pointé par `www.wikipedia.org`, en sécurisant la connexion grâce à HTTPS (que nous verrons par la suite).

Éléments de syntaxe d'une URL



Certains caractères sont réservés pour un usage spécifique. Par exemple :

- le caractère `/` est réservé pour indiquer le chemin des fichiers à consulter,
- le caractère `?` est réservé pour annoncer les paramètres à transmettre au serveur,
- le caractère `=` est réservé pour indiquer la valeur des paramètres,
- le caractère `#` est réservé pour indiquer un **fragment**, souvent une partie spécifique de la page web (par exemple, un paragraphe).

Lorsqu'un caractère réservé doit être utilisé pour autre chose dans l'URL, il est remplacé par un symbole pourcent suivi de son code ASCII au format hexadécimal (`'/'` devient `%2F` et `'?'` devient `%3F`).

À retenir

- Une URL permet d'identifier une ressource précise sur un serveur web donné.
- Les URL permettent de passer des paramètres aux serveurs web.

Les clients et les serveurs web parlent HTTP



Serveur web



Un serveur web est un logiciel qui **traite les requêtes HTTP** de clients.

Il est installé sur une machine hébergeant des documents (HTML, CSS, JavaScript, etc.) et est accessible depuis Internet.

Navigateur web



Un navigateur web est client HTTP qui a trois rôles.

1. Effectuer des requêtes HTTP (GET, POST, etc.).
2. Comprendre les réponses à ces requêtes.
3. Interpréter les fichiers web (HTML, CSS, JavaScript) retournés.

Méthodes HTTP



Les méthodes HTTP sont les commande que le client (navigateur) envoie au serveur.



Les plus courantes sont :

- GET (pour demander une ressource, une page HTML par exemple)
- POST (pour transmettre des données au serveur, par exemple quand on remplit un formulaire en ligne)

Code de retour HTTP



Un code HTTP est un code à 3 chiffres qui est présent dans chaque réponse HTTP du serveur.



Les plus courants sont :

- 200 OK: « Tout est bon »,
- 401 Unauthorized: « Vous n'avez pas le droit d'accéder à cette ressource, il faut une authentification »,
- 404 Not Found: « Ressource non trouvée »,
- 500 Internal Server Error: « Erreur du serveur ».

Serveurs web

Navigateurs web

URL : Uniform Resource Locator (cf. p.20)

Codes de retour HTTP

Exercice



Internet c'est pas du virtuel

Vous participez à la Fête De la Science et dans ce contexte vous devez expliquer le fonctionnement d'Internet à des collégiens.

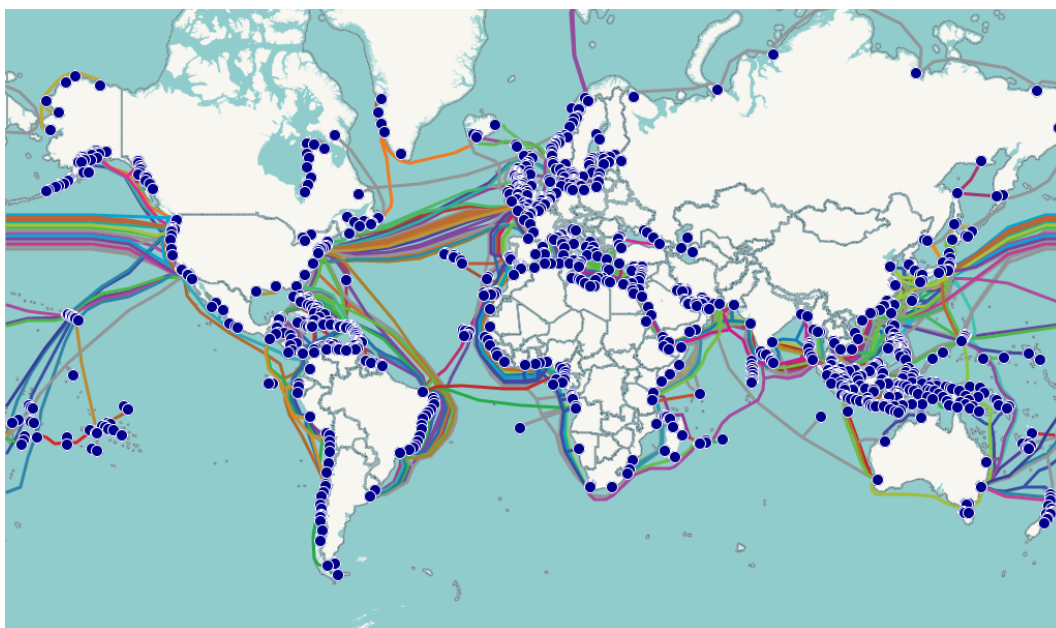
Préparez votre exposé en répondant aux questions suivantes en une ou deux phrases simples et précises.

Question 1

[solution n°1 p. 29]

Expliquez en quoi Internet repose sur une infrastructure physique qui consomme des matériaux, de l'énergie et de l'espace.

Indice :

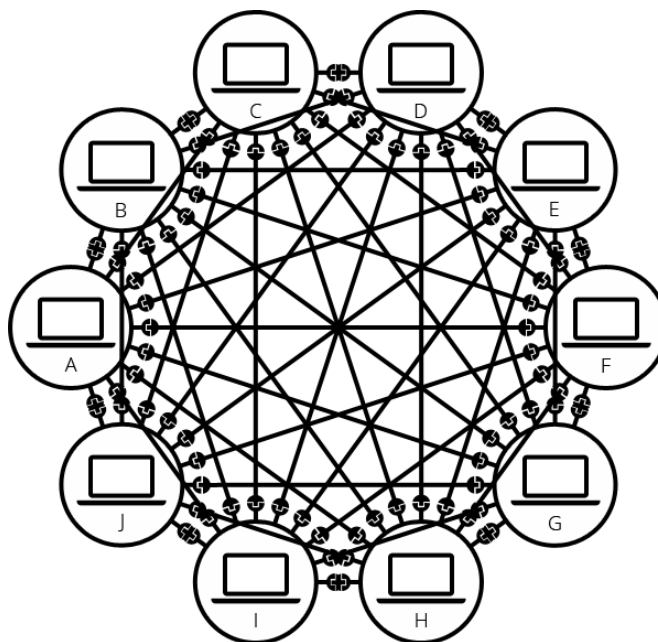


Carte des câbles sous-marins du réseau Internet (Monde)

Question 2

Quel est le rôle des routeurs au sein du réseau Internet et pourquoi ne peut-on pas s'en passer ?

Indice :

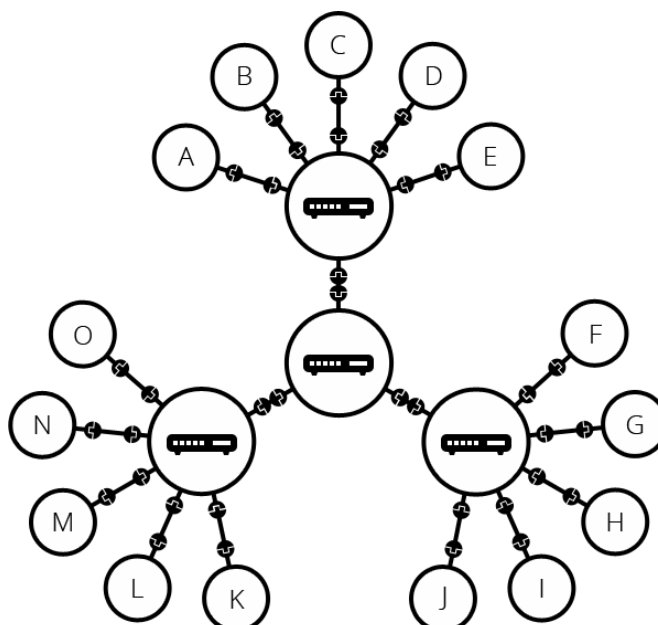


Pour relier 10 ordinateurs on a besoin de 45 câbles et de 9 prises sur chaque ordinateur

Question 3

Qu'est-ce qu'un protocole et pourquoi est-ce indispensable pour faire fonctionner Internet ?

Indice :



Internet est un réseau de réseaux qui relie des routeurs du monde entier

Question 4

[solution n°4 p. 29]

Expliquez le principe de fragmentation des messages en paquets sur lequel repose le protocole réseau TCP/IP.

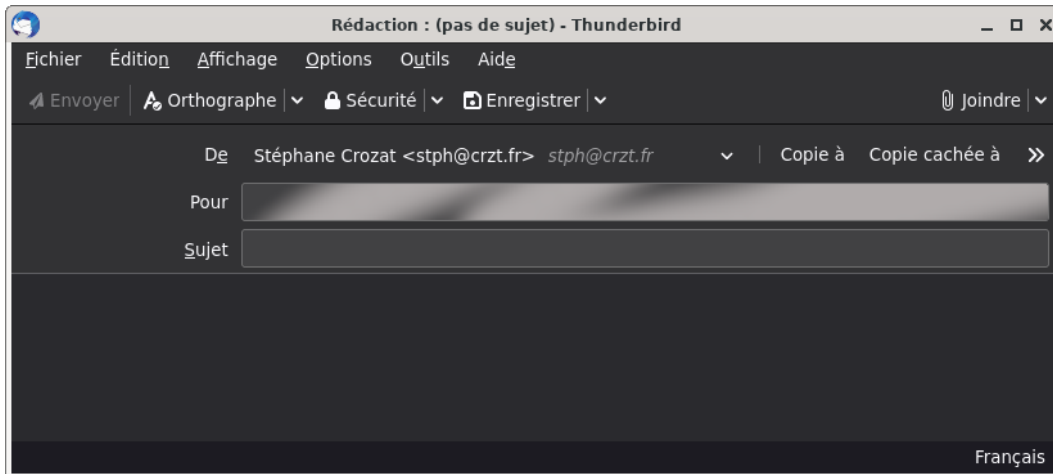
Indice :

Question 5

[solution n°5 p. 29]

À quoi servent les adresses IP et quelles sont les machines qui en disposent ? À quoi servent les noms de domaine ?

Indice :

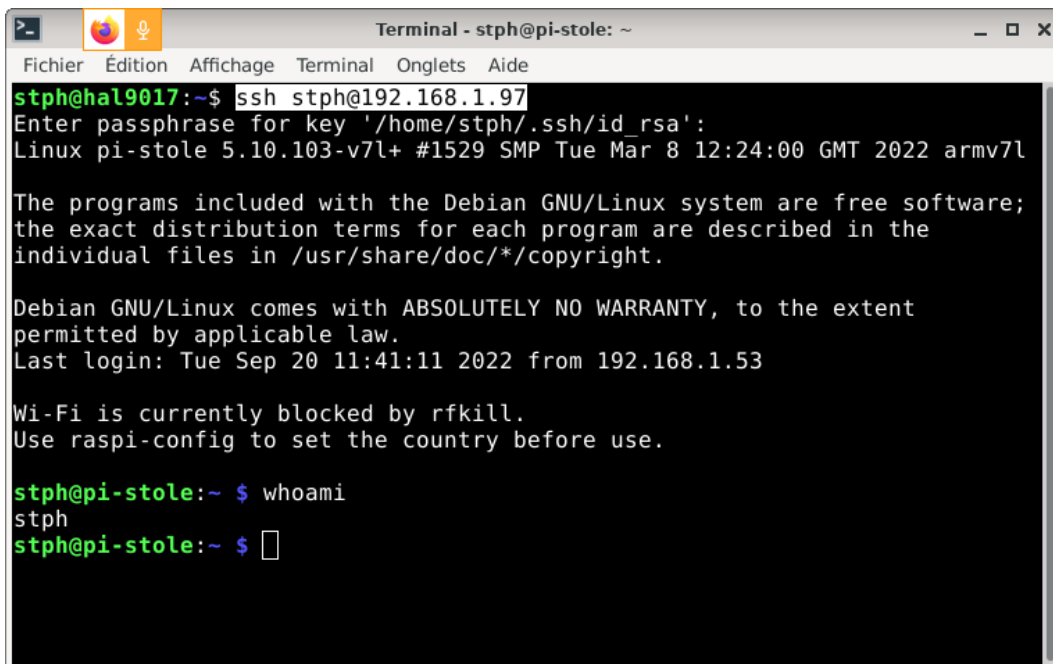


Question 6

[solution n°6 p. 29]

À quoi sert l'application Telnet, depuis remplacée par SSH, et pourquoi était-ce la première application sur Internet ?

Indice :



Question 7

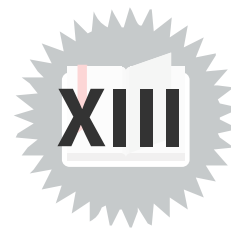
[solution n°7 p. 29]

Expliquez comment fonctionne un navigateur web à partir de la méthode GET, la plus courante du protocole HTTP.

Indice :



Cours-émission



Première partie (jusque 35:40)

Solutions des exercices



[exercice p. 24] **Solution n°1**

[exercice p. 25] **Solution n°2**

[exercice p. 25] **Solution n°3**

[exercice p. 26] **Solution n°4**

[exercice p. 26] **Solution n°5**

[exercice p. 26] **Solution n°6**

[exercice p. 27] **Solution n°7**

Abréviations



IETF : Internet Engineering Task Force

Crédits des ressources



p. 5

Universel - Transfert dans le Domaine Public - Picasoft

Exemples d'architecture Web p. 6

Licence de documentation libre GNU - Pillou, commentcamarche.net

<https://tecnologiahabitual.blogspot.com/2014/07/mantenimiento-en-datacenters.html> p. 9

Attribution - Partage dans les Mêmes Conditions - Mariano Simón Lázaro

Codage d'un paquet IP p. 13

Attribution - Partage dans les Mêmes Conditions - Stéphane Crozat

Décomposition d'une URL p. 20

Licence : Domaine Public

Carte des câbles sous-marins du réseau Internet (Monde) p. 24

<https://framacarte.org/m/72142>

Pour relier 10 ordinateurs on a besoin de 45 câbles et de 9 prises sur chaque ordinateur p. 25

https://developer.mozilla.org/fr/docs/Learn/Common_questions/How_does_the_Internet_work