

Le World Wide Web

*« Je n'ai fait que prendre le principe d'-
hypertexte et le relier au principe du
TCP et du DNS et alors – boum ! – ce fut
le World Wide Web ! »
Tim Berners-Lee*

Objectifs

- Se familiariser avec l'environnement réseau du web
- Acquérir les capacités d'auto-formation nécessaires pour suivre les évolutions à venir

Définitions

- Un *réseau* est un ensemble d'équipements informatiques interconnectés
- Un *réseau* s'appuie sur deux notions :
 - L'*interconnexion* : permet de transmettre les données d'un nœud à un autre
 - La *communication* : permet d'échanger des données entre processus (un programme en cours d'exécution)
- Un *réseau* désigne un ensemble d'équipements matériels et logiciels mis en oeuvre pour permettre la communication entre applications, quelles que soient les distances qui les séparent.

Éléments d'un réseau

- Les ordinateurs équipées d'une carte de communication
- Les logiciels
 - navigateur, client de messagerie, serveur web, ...
- Les supports
 - de LAN : câbles paires cuivre torsadées, prises RJ45, WIFI, CPL, ...
 - de WAN : ligne téléphonique, ADSL, fibre optique, ...
- Les équipements d'interconnexion
 - de LAN : répéteur (*transceiver*), concentrateur (*hub*), commutateur (*switch*)
 - de WAN : routeur



Caractéristiques

- La **topologie** définit l'architecture d'un réseau : on distinguera la topologie physique (qui définit la manière dont les équipements sont reliés entre eux, de la topologie logique (qui précise la manière dont les équipements communiquent entre eux) :
 - par exemple, une topologie logique en bus (Ethernet 10BASET) pourra se câbler avec une topologie physique en étoile (*hub*).
- Le **débit** mesure une quantité de données numériques (bits) transmises par seconde (bit/s ou bps).
- La **distance maximale** (ou portée), qui différencie essentiellement les LAN et WAN, dépend de la technologie mise en oeuvre :
 - WIFI 802.11g (54 Mbps – environ 50 m), Ethernet paires torsadées 100BASET (100 Mbps – 100 m) et fibre optique 100BASEFX (100 Mbps – 2 km)

Topologie

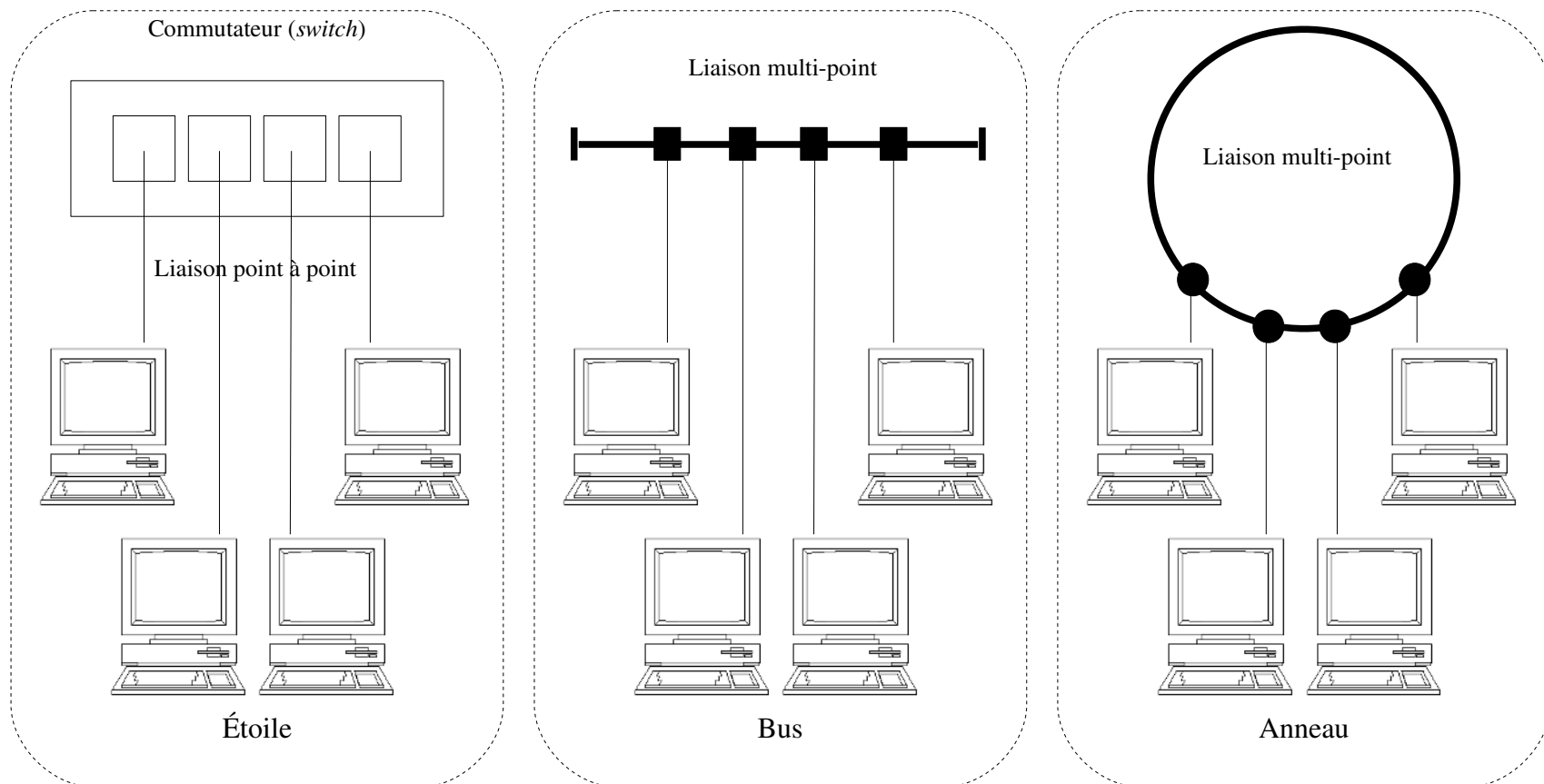
- Il existe 2 modes de diffusion classant les différentes topologies :
 - **MODE MULTI-POINT**
 - (exemple : topologie en bus ou en anneau)

Ce mode de fonctionnement consiste à n'utiliser qu'un seul support de transmission. Le principe est que le message est envoyé sur le réseau, toute unité réseau est capable de voir le message et d'analyser selon l'adresse du destinataire si le message lui est destiné ou non.
 - **MODE POINT A POINT**
 - (exemple : topologie en étoile, arbre ou maillée)

Dans ce mode, le support physique ne relie qu'une paire d'unités seulement. Pour que deux unités réseaux communiquent, elles passent obligatoirement par un équipement d'interconnexion (un routeur ou un commutateur).

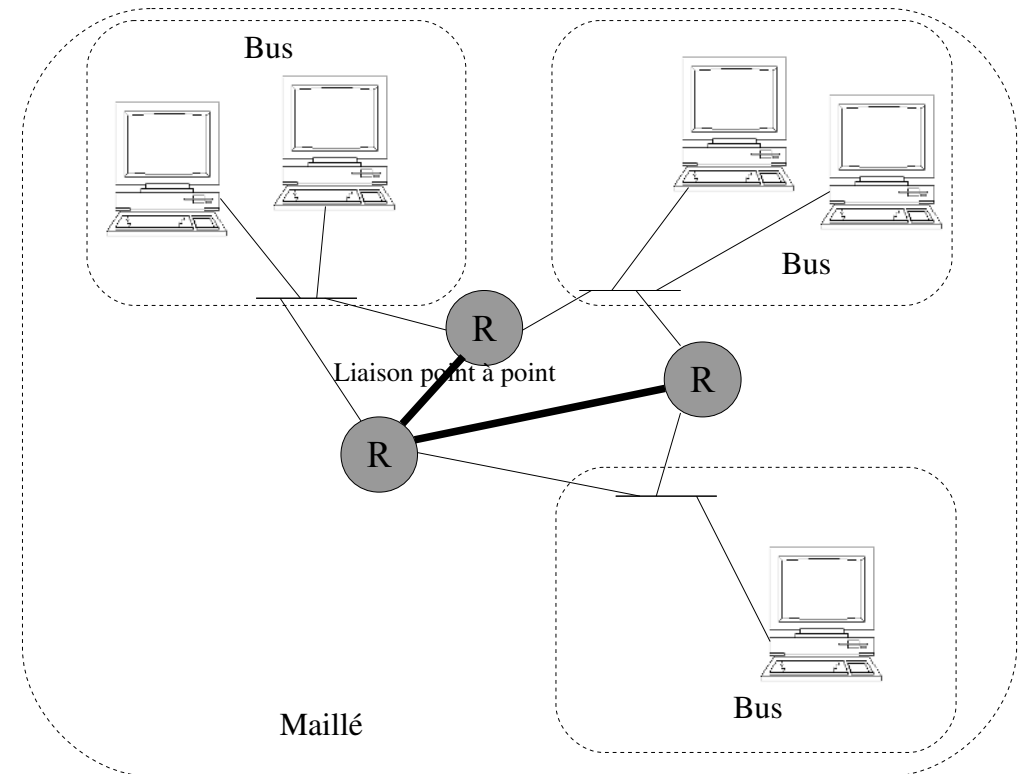
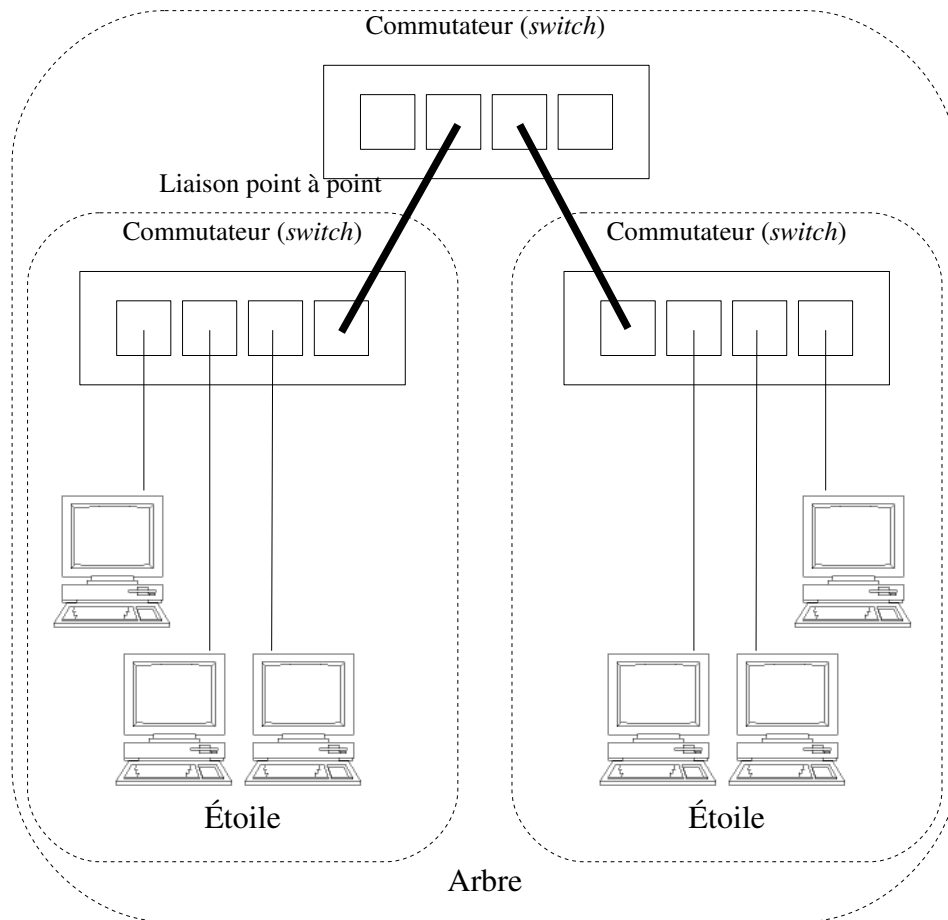
Topologies (I)

- Certaines topologies sont plus adaptées aux LAN (bus, anneau, étoile), d'autres aux WAN (maillé).



Topologies (II)

- Certaines topologies (arbre, maillé) sont plus adaptées pour interconnecter des LAN entre eux.



La communication en réseau

- Les échanges de données sont basés sur une communication *logique*.
- Les communications dans un réseau obéissent à des règles :
 - l'*adressage* qui permet d'identifier de manière unique les deux unités en communication
 - l'*architecture* qui définit les rôles endossés par les deux unités
 - les *protocoles* qui assurent l'échange des données

Les protocoles

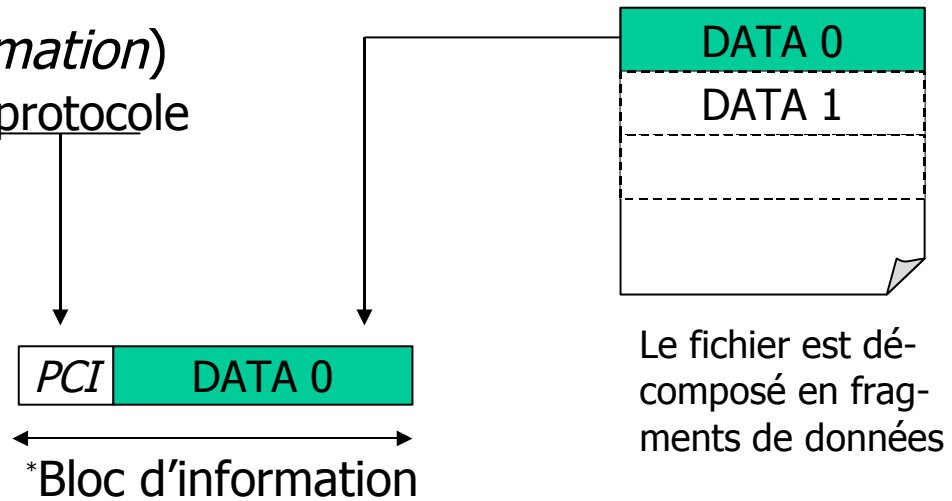
- Rendent possible le dialogue entre des machines différentes
- Un protocole de communication définit l'ensemble des procédures (ou règles) pour réaliser une communication :
 - Le dictionnaire : les primitives (demande connexion, acquittement, ...)
 - Le scénario du dialogue : enchaînement des primitives (diagramme de l'échange)
 - Les modalités : taille et représentation des informations, temps d'attente, etc ...
 - Les messages échangés : les différents champs (taille et contenu)

Protocole de communication

PCI (*Protocol Control Information*)

En-tête (*header*) ajouté par le protocole réseau

Il sera décodé par le récepteur, qui doit donc posséder le même protocole réseau.



Un protocole réseau définit :

- le contenu détaillé du PCI
- la taille du bloc d'information
- la manière de les échanger

*Termes couramment utilisés pour désigner un bloc d'information :

trame, paquet, datagramme, segment, message, fragment, ...

Protocole HTTP

- Le Protocole HTTP (*HyperText Transfert Protocol*) sert notamment au dialogue entre un client web (navigateur par exemple) et un serveur (apache par exemple).
- Comme la plupart des protocoles de la couche Application, c'est un **protocole orienté texte (ASCII)**, basé sur TCP. Il existe deux spécifications la 1.0 et la 1.1 (RFC 1945).

Requête HTTP

```
GET /index.html HTTP/1.1\r\nHost: www.btsiris.net\r\n\r\n
```

→ Ligne vide = fin de l'en-tête HTTP

En-tête
Corps

Le corps est vide

Réponse HTTP

```
HTTP/1.1 200 OK → Ligne de statut  
Date: Wed, 10 Mar 2010 09:58:08 GMT  
Server: Apache/2.2.11 (Mandriva  
Linux/PREFORK-10.7mdv2009.1)  
Content-Length: 215  
Connection: close  
Content-Type: text/html
```

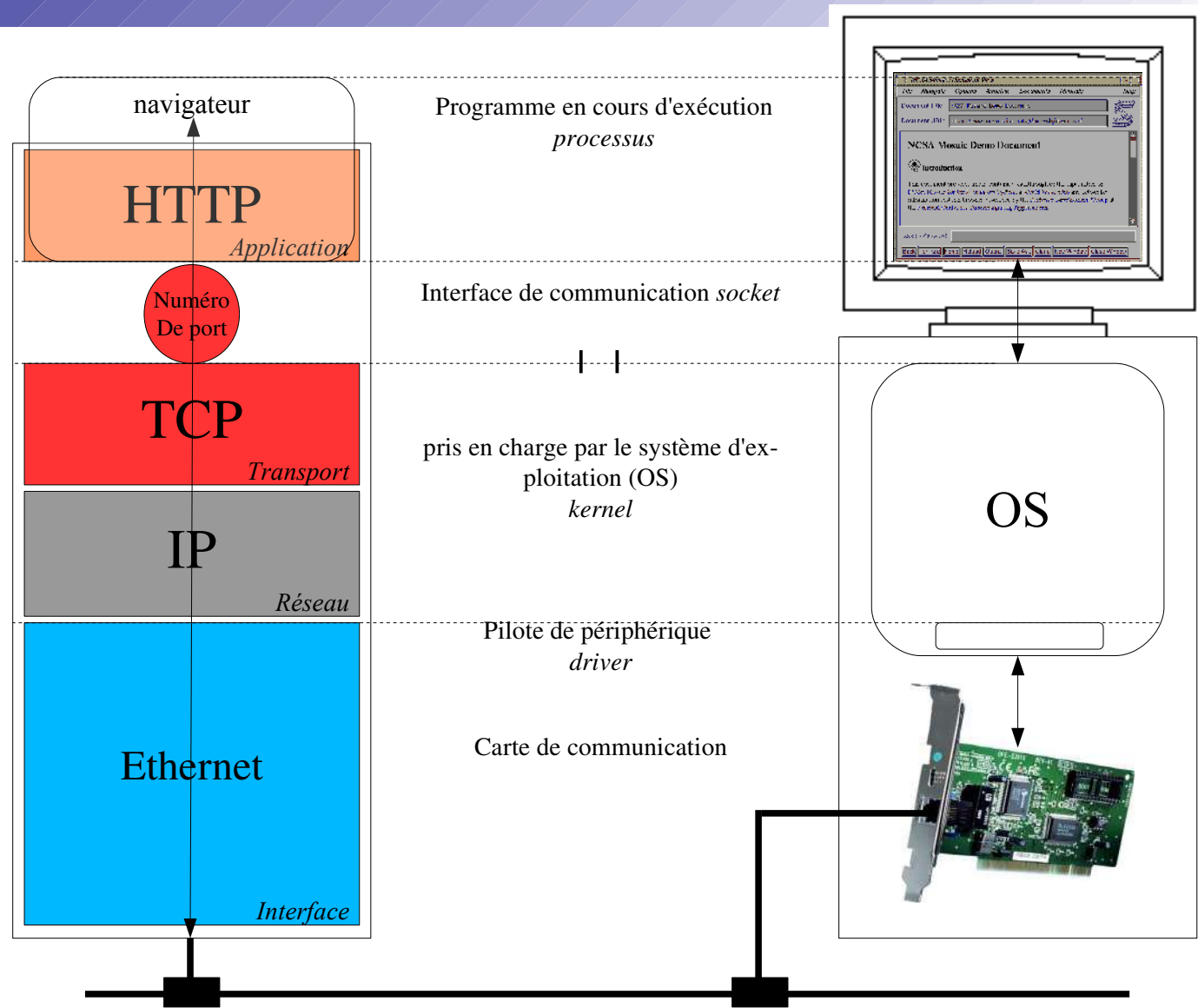
En-tête
Corps

→ Ligne vide = fin de l'en-tête HTTP

```
<html>  
<body> Le corps de la réponse contient le  
<h1>It works!</h1> contenu du fichier index.html deman-  
</body> dé dans la requête  
</html>
```

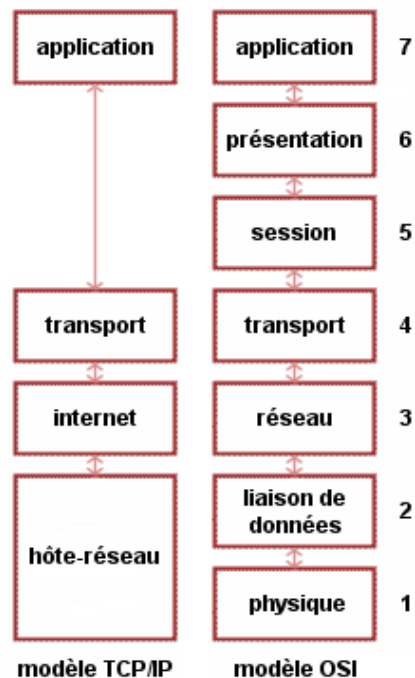
Modèle de référence (I)

- Un modèle de référence est utilisé pour décrire la structure et le fonctionnement des communications réseaux
- Le modèle DoD (*Department of Defense*) ou « TCP/IP » est composé de 4 couches
- En raison de son apparence, la structure est très souvent appelé **pile** ou **pile de protocoles**.



Modèle de référence (II)

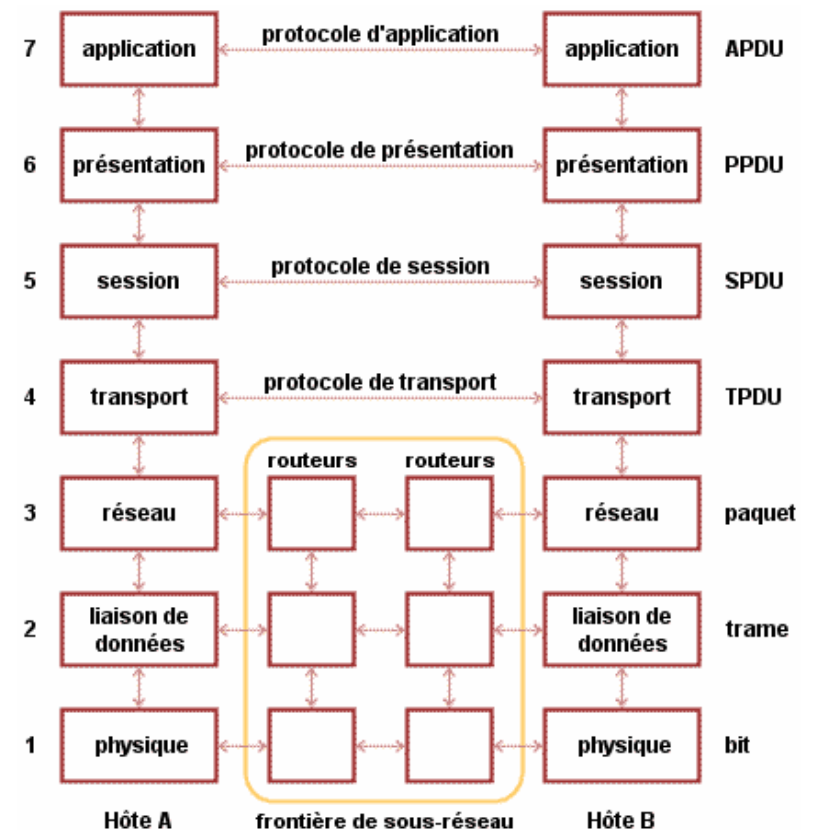
- Un modèle de représentation développé par l'**ISO** (*International Standards Organization*) est souvent utilisé pour décrire la structure et le fonctionnement des communications réseaux : le modèle **OSI** (*Open Systems Interconnect Reference Model*).
- Le modèle OSI contient **7 couches ou niveaux** qui définissent les **fonctions** des protocoles de communication qui vont de l'interface physique à l'interface avec les applicatifs utilisant le réseau.



à l'interface avec les applicatifs utilisant le réseau.

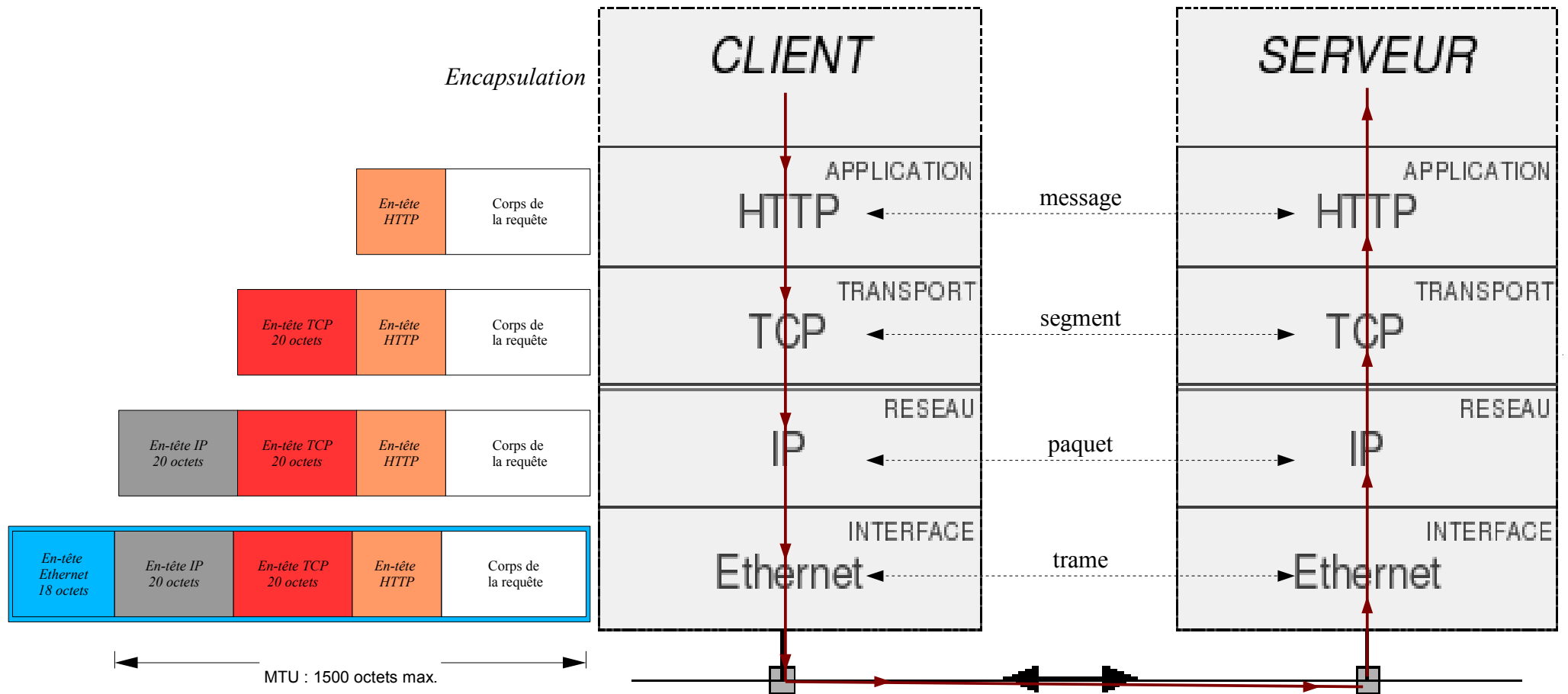
Critiques du modèle OSI :

- Ce n'était pas le bon moment : trop tôt ou trop tard ?
- Ce n'était pas la bonne technologie : trop complet et trop complexe
- Ce n'était pas la bonne implémentation : trop lourd et trop lent
- Ce n'était pas la bonne politique : trop normalisé et trop bureaucratique



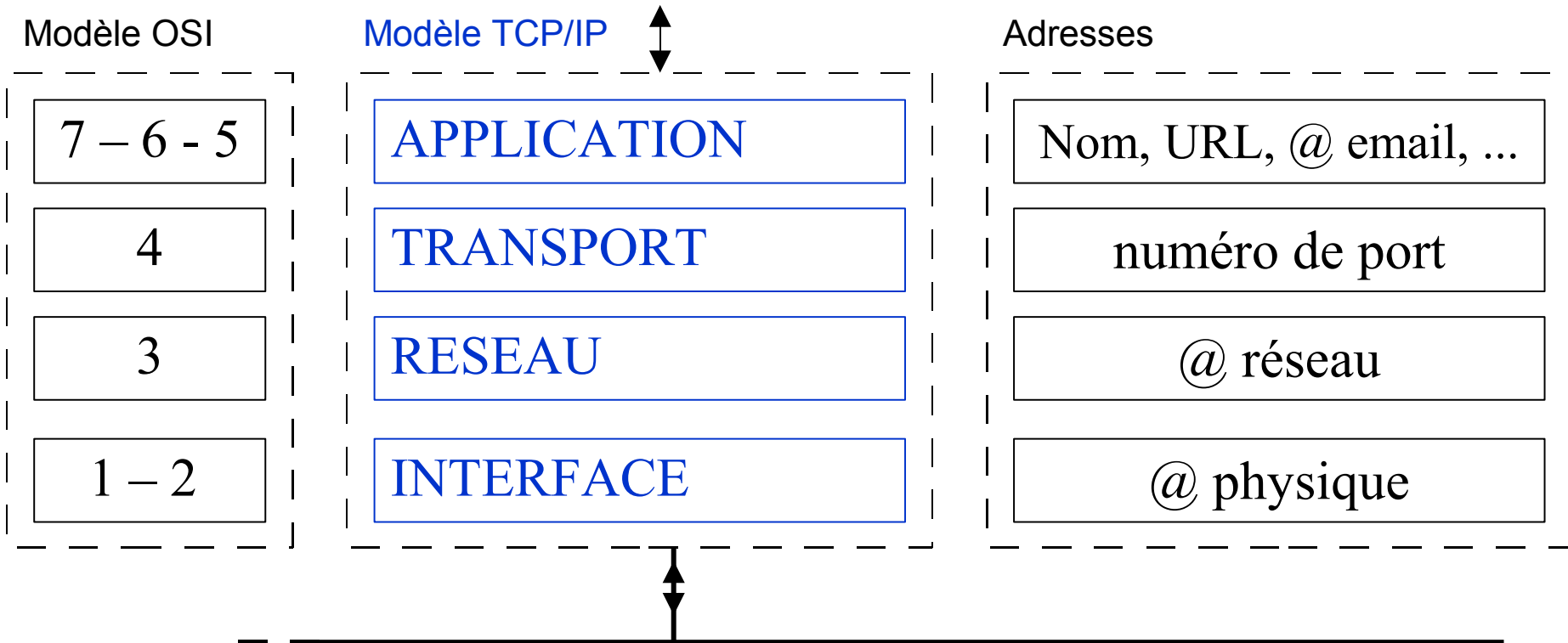
L'encapsulation

- Dans une communication entre deux équipements, de nombreux protocoles sont mis en oeuvre : ils sont encapsulés les uns dans les autres

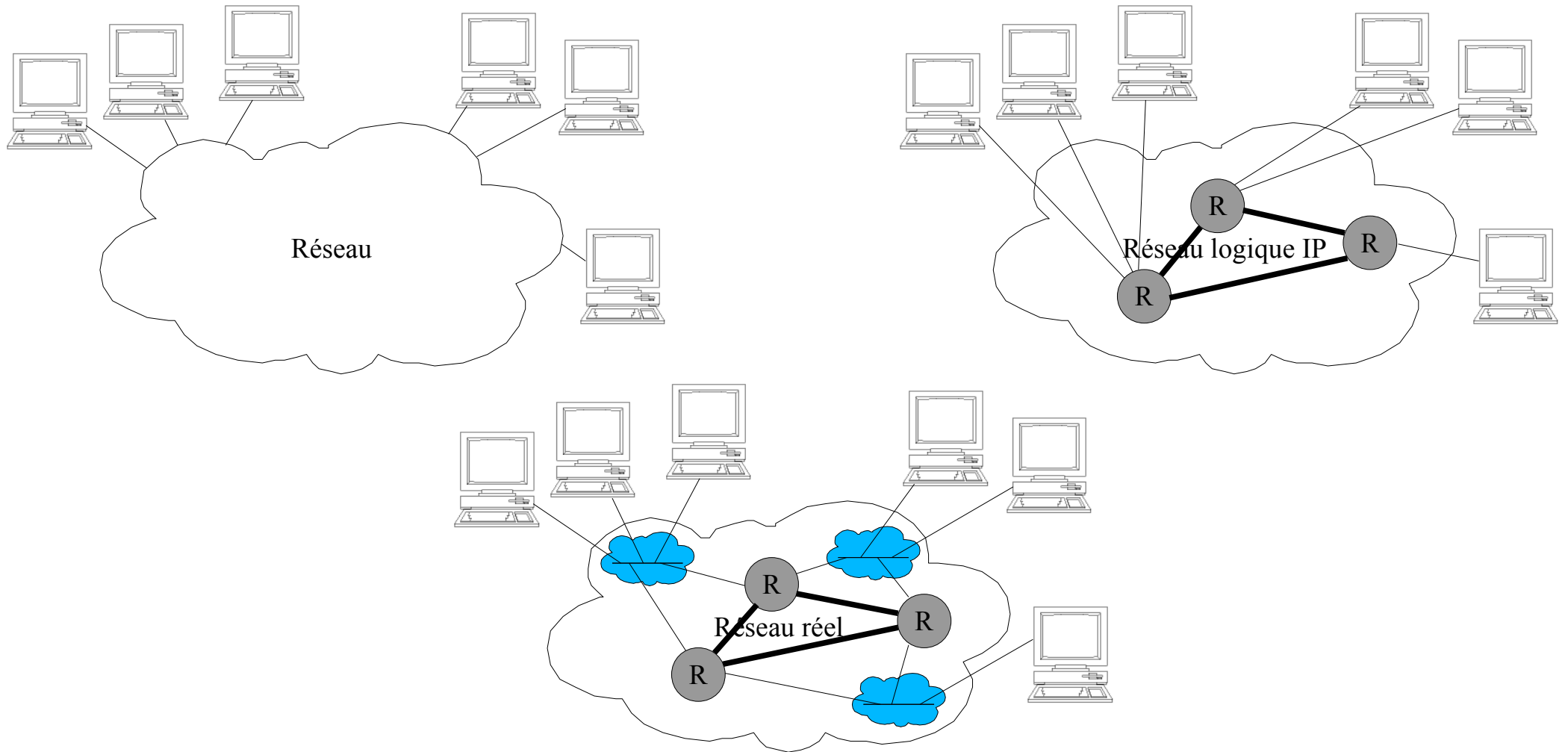


L'adressage (I)

- Identifier de manière unique une interface, un poste dans un réseau, une application (un processus), une ressource, un fichier, un document, un utilisateur, ... sur un réseau
- On distinguera donc plusieurs types d'adresse :



L'adressage (II)



L'adressage IP (II)

- Une adresse IP est décomposée en deux parties : une partie identifie le réseau (*net-id*) auquel appartient l'hôte et une partie identifie le numéro de l'hôte (*host-id*) dans ce réseau.



- Le masque de sous-réseau permet de savoir quelle partie d'une adresse IP correspond à la partie numéro de réseau et laquelle correspond à la partie numéro de l'hôte. On utilise une opération de ET bit à bit entre l'adresse IP et le masque de sous-réseau pour extraire la partie réseau de l'adresse.
- L'adressage CIDR (*Classless Inter-Domain Routing*) a été mis au point afin (principalement) d'insuffler une plus grande durée de vie aux adresses IPv4 dans l'attente d'un passage à IPv6. La notation CIDR abandonne l'adressage par classe et indique une adresse réseau suivi d'un '/' et d'un nombre indiquant les bits à 1 constituant le masque de sous-réseau (en partant de la gauche). Exemple : 10.0.0.0/8 correspond un masque 255.0.0.0.

L'adressage web : URI/URL

- Un URI (*Uniform Resource Identifier*) soit littéralement « identifiant uniforme de ressource », est une courte chaîne de caractères identifiant une ressource sur un réseau (par exemple une ressource Web) physique ou abstraite, et dont la syntaxe respecte une norme d'Internet mise en place pour le World Wide Web (voir RFC 3986).
- Un URL (*Uniform Resource Locator*) littéralement « localisateur uniforme de ressource », est une chaîne de caractères utilisée pour adresser les ressources du World Wide Web : document HTML, image, son, forum Usenet, boîte aux lettres électronique, etc. Elle est aussi appelée une adresse web.
- Le format d'une adresse web ou URL :

`protocole://[<login>:<mot_de_passe>]<serveur>[:<numero_port>]/[<chemin>/]<ressource>`

Exemples : `http://www.example.com/tim/page.html`

`http://192.168.52.83/index.php`

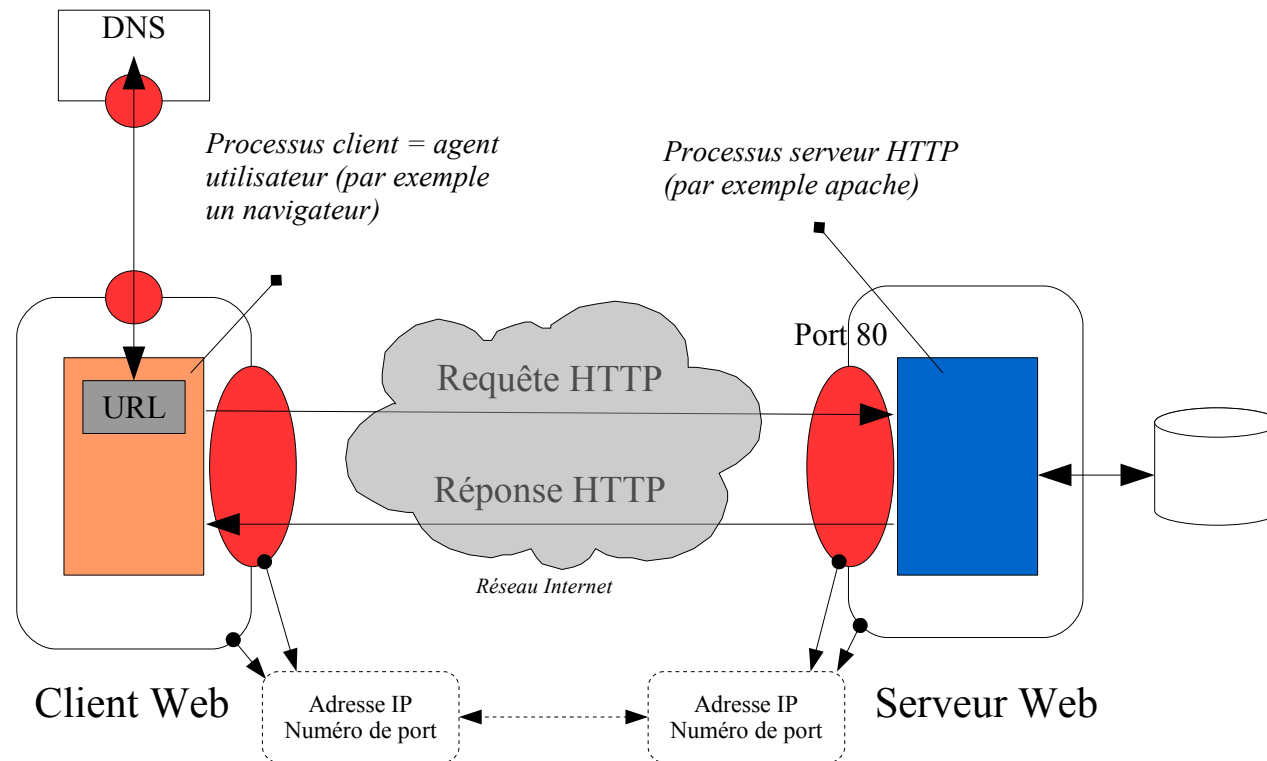
`ftp://ftp.is.co.za:21/rfc/rfc1808.txt`

DNS

- DNS (*Domain Name System* ou système de noms de domaine) est un système permettant d'établir une correspondance entre une adresse IP et un nom de domaine et, plus généralement, de trouver une information à partir d'un nom de domaine.
- Avant le DNS, la résolution devait se faire grâce à un fichier texte appelé HOSTS, local à chaque ordinateur. Sous UNIX/Linux, il se trouve dans le répertoire /etc. Sous Windows, il se trouve par défaut dans %SystemRoot%\system32\drivers\etc.
- Avec DNS, la résolution se fait par l'intermédiaire d'un serveur (port 53 sur UDP). Quand un utilisateur souhaite accéder à un serveur web, par exemple celui de fr.wikipedia.org, son ordinateur émet une requête vers un serveur DNS, demandant 'Quelle est l'adresse de fr.wikipedia.org ?'. Le serveur répond en retournant l'adresse IP du serveur, qui est dans ce cas-ci, 91.198.174.2.

Architecture Client/Serveur (I)

- Serveur : offre un service (en attente)
- Client : demandeur d'un service
- La communication s'initie TOUJOURS à la demande du client.



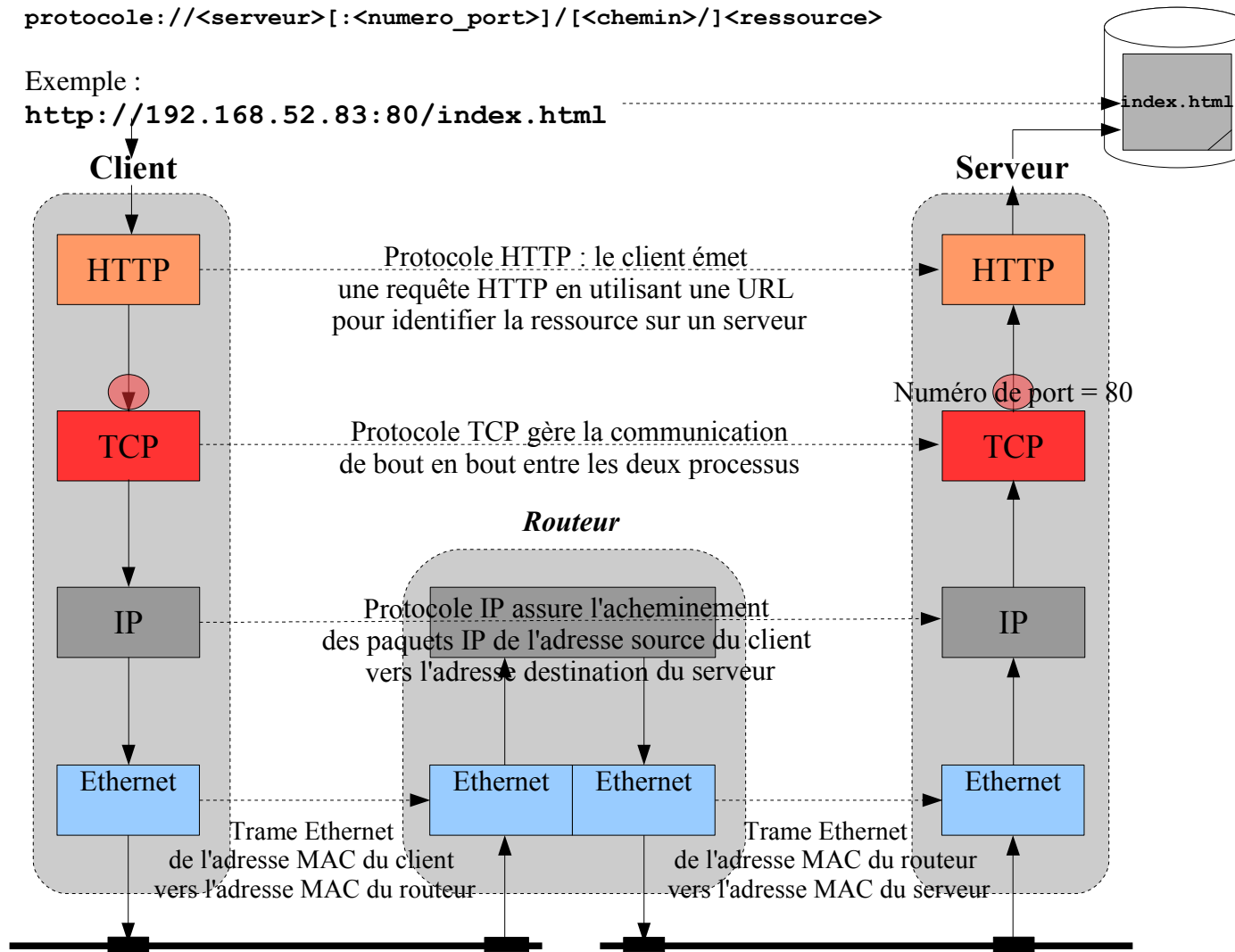
Architecture Client/Serveur (II)

Format d'une URL :

protocole://<serveur>[:<numero_port>]/[<chemin>/]<ressource>

Exemple :

http://192.168.52.83:80/index.html



Serveur mandataire (*proxy*)

- Un serveur mandataire ou *proxy* est un serveur informatique qui a pour fonction de relayer des requêtes entre un poste client et un serveur.
- Les serveurs proxy sont notamment utilisés pour assurer les fonctions suivantes :
 - mémoire cache (amélioration des performances)
 - la journalisation des requêtes (« logging »)
 - la sécurité du réseau local
 - le filtrage et l'anonymat
 - l'authentification pour autoriser ou non l'accès au service

Historique : Internet

- 1958 : La BELL crée le premier Modem permettant de transmettre des données binaires sur une simple ligne téléphonique
- 1961 : Leonard Kleinrock du MIT publie une première théorie sur l'utilisation de la commutation de paquets pour transférer des données
- 1962 : Début de la recherche par ARPA, une agence du ministère de la Défense américain
- 1964 : Leonard Kleinrock du MIT publie un livre sur la communication par commutation de paquets pour réaliser un réseau
- 1969 : Connexion des premiers ordinateurs sur l'ARPANET
- 1979 : Création des NewsGroups (forums de discussion) par des étudiants américains
- 1982 : Définition du protocole TCP/IP et du mot « Internet »
- 1983 : Premier serveur de noms de sites (DNS)
- 1988 : Première connexion Internet en France
- 1991 : Annonce publique du *World Wide Web* qui est basé sur trois inventions, le protocole de communication client/serveur HTTP (*Hypertext Transfer Protocol*), les adresses web (URI/URL) et le langage HTML (*HyperText Markup Language*).
- 1994 : Premier moteur de recherche

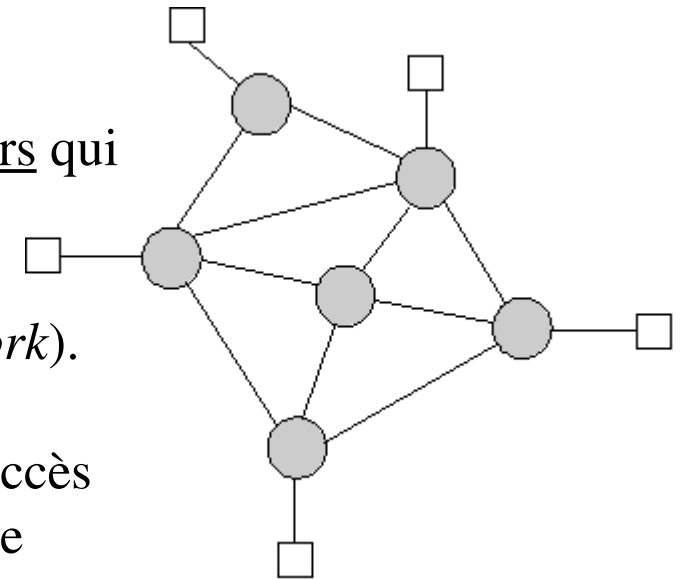
Historique : le web

- **1991 : Origine**, le *World Wide Web* est basé sur trois inventions, le protocole de communication client/serveur HTTP (*Hypertext Transfer Protocol*), les adresses web (URI/URL) et le langage HTML (*HyperText Markup Language*).
- **1993-1997 : Evolution du langage HTML par l'apports des navigateurs (NCSA Mosaic, Netscape Navigator, Microsoft Internet Explorer, Mozilla Firefox, ...)**
- **1994 : Naissance du W3C (*World Wide Web Consortium*)** qui est un organisme de normalisation chargé de promouvoir la compatibilité des technologies du *World Wide Web* telles que HTML, XHTML, XML, CSS, PNG, SVG et SOAP. Le W3C n'émet pas des normes au sens européen, mais des recommandations à valeur de standards industriels. Tim Berners-Lee préside aujourd'hui le W3C.
- **1996 : Standardisation HTTP/1.0 (RFC 1945)**, cette version supporte les serveurs HTTP virtuels, la gestion de cache et l'identification.
- **1997 : Standardisation HTTP/1.1 (RFC 2068 puis dans la RFC 2616 en juin 1999)**, cette version ajoute le support des connexions persistantes et la négociation de type de contenu (format de données, langue).
- **1999 : dernière spécification de HTML, la 4.01**

Internet (I)

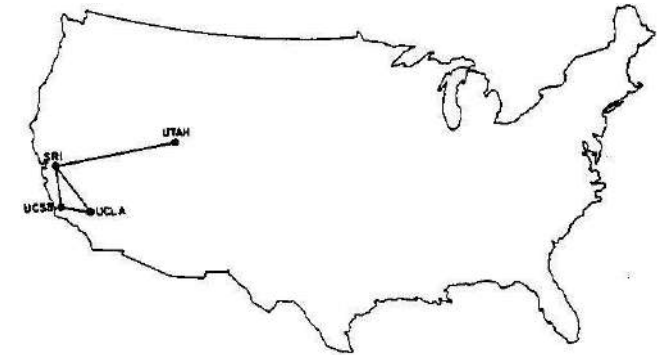
- Un réseau de réseaux

- Internet est un réseau public mondial qui relie des milliers de réseaux plus petits ou des ordinateurs isolés.
- L'interconnexion des réseaux est réalisé par des routeurs qui donnent une topologie de type maillé.
- Internet est un réseau de type WAN (*Wide Area Network*).
- Le plus souvent, on passe par un FAI (Fournisseur d'Accès Internet) ou ISP (*Internet Service Provider*) pour se connecter (raccorder) au réseau Internet.

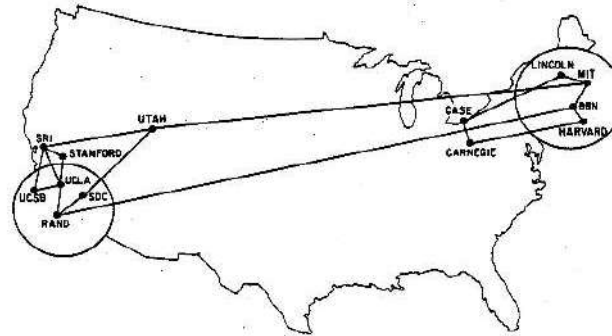


Internet : évolution (II)

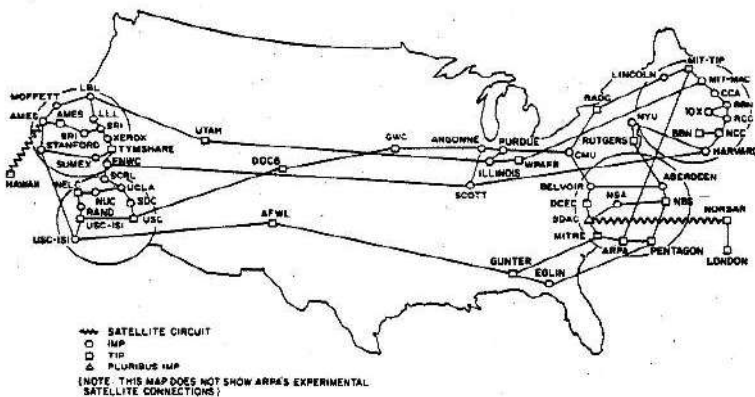
- Depuis 2006, il y a plus d'un milliard d'ordinateurs connectés à Internet ...



Il y a seulement 4 noeuds à la création du réseau ARPAnet fin 1969.



Un an plus tard, fin 1970, il y a 13 noeuds d'interconnectés, le réseau maillé se construit ...



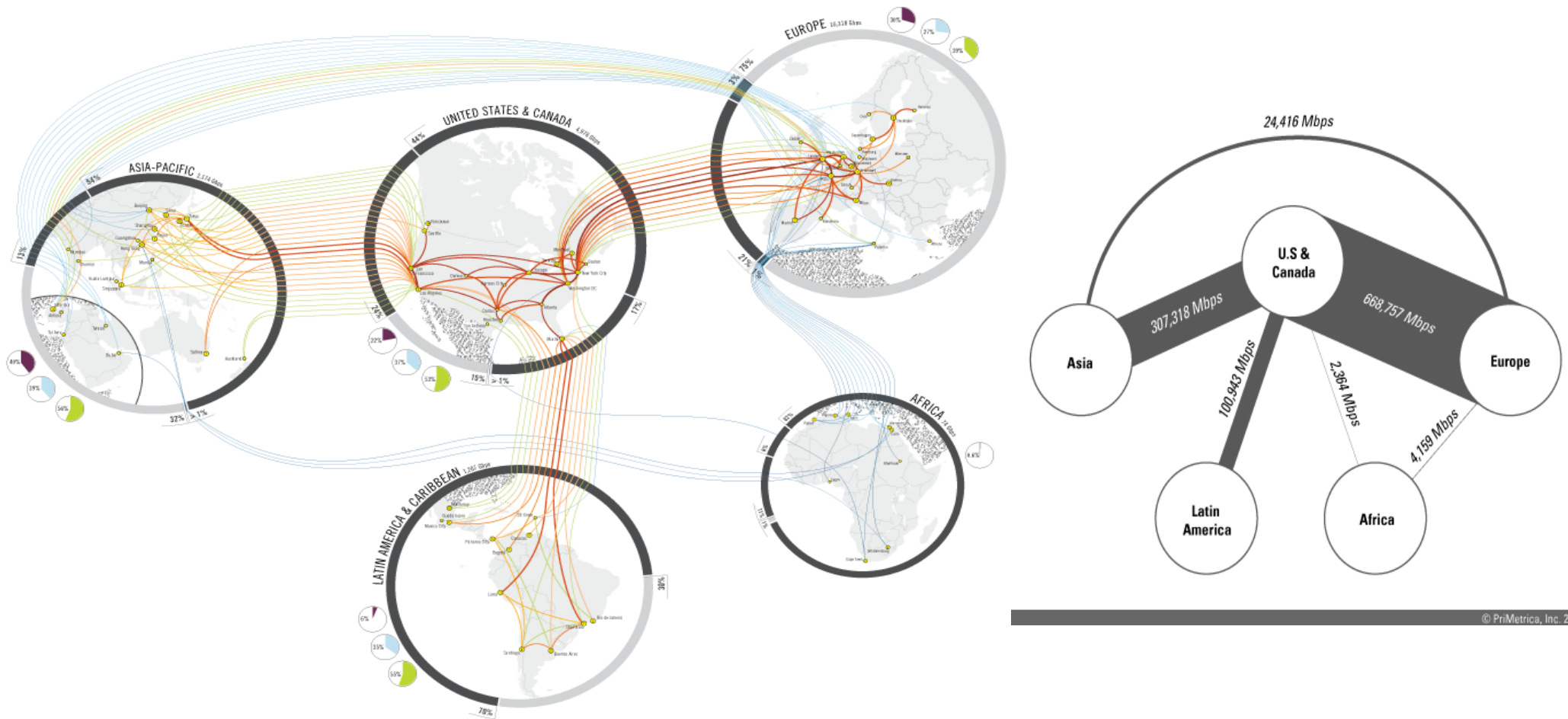
Il y aura plus de 100 000 noeuds à la fin des années 80



L'équipe de Christian HUITEMA à l'INRIA Sophia Antipolis réalise la première connexion Internet en France en juillet 1988.

Internet : réseau mondial (III)

- Source : http://www.telegeography.com/ee/free_resources/



Internet (IV)

- Un réseau de services

- On utilise le réseau Internet pour les services (ou application) qu'il propose : web, messagerie,
- Les services Internet sont fournis par des serveurs. Les demandeurs du service sont nommé les clients. L'architecture qui en découle se nomme client/serveur.
- Chque service, pris en charge par un processus (un programme en cours d'exécution), est identifié par un numéro (numéro de port).
- Chaque service est associé à un protocole (ensemble de règles pour communiquer) :
 - Web (www) : HTTP - Transfert de fichiers : FTP, BitTorrent, eDonkey, ...
 - Courrier électronique (mail) : SMTP, POP, IMAP, ...
 - Messagerie instantanée : AIM, ICQ, Jabber, XMPP, MSN Messenger, ...
 - Discussion (chat) : IRC - Système de fichiers : NFS, SMB, ...
 - Session distante (émulation de terminal) : Telnet, Rlogin, SSH, ...
 - Forum de discussion (news) : NNTP (Usenet), ... - Supervision : SNMP, ...
 - Résolution d'adresse : DNS, ... - Synchronisation horaire : NTP - Affichage distant : XDMCP
 - etc ...

RFC

- Les *Requests For Comment* (RFC), littéralement demande de commentaires, sont une série numérotée de documents électroniques documentant les aspects techniques d'Internet.
- Peu de RFC sont des standards, mais tous les standards d'Internet sont des RFC.
- Les RFC sont rédigées pas des experts techniques. En mai 2008, le nombre de RFC a atteint les 5 000.
- La première RFC (RFC 1), titrée "Logiciel hôte", a été publiée le 7 avril 1969 par Steve Crocker.

Terminologie Internet

- Une passerelle (*gateway*) est un dispositif permettant de relier deux réseaux informatiques, comme par exemple un réseau local et Internet. Cependant, le terme passerelle (sans autre précision) est couramment employé comme exact synonyme du terme routeur. Par exemple, on parle de passerelle par défaut (*default gateway*) ou **gateway IP** pour désigner un routeur qui interconnecte deux réseaux IP. Le routeur est un équipement réseau qui permet de relayer les paquets d'un réseau vers un autre.
- **Internet** est le réseau informatique mondial qui rend accessibles au public des services (comme le courrier électronique et le World Wide Web). Ses utilisateurs sont désignés par le néologisme « **internaute** ». Techniquement, Internet se définit comme le réseau public mondial utilisant le protocole de communication « TCP/IP » (au sens les protocoles de la famille TCP/IP).
- Lorsque les technologies Internet (TCP/IP, services, etc.) sont mises en oeuvre au sein de réseaux privés (entreprises, administrations, etc ...), on parle alors d'**intranet**.

Bibliographie

- "TCP/IP sous Linux" de JF Bouchaudy - Formation Tsoft © Ed. Eyrolles
- "TCP/IP Administration de réseau" de Craig Hunt © Ed. O'Reilly
- "Les protocoles TCP/IP et Internet" d'Eric Lapaille © NetLine 1999
- "Webmaster in a nutshell" © Ed. O'Reilly
- "Technique des réseaux locaux sous Unix" de L. Toutain © Ed. Hermes
- "Pratique des réseaux locaux d'entreprise" de JL Montagnier © Ed. Eyrolles
- "Transmission et Réseaux" de S. Lohier et D. Present © ED. DUNOD
- Les sites www.frameip.com, fr.wikipedia.org, www.w3.org, etc ...

© Copyright 2010 tv <thierry.vaira@orange.fr>

Permission is granted to copy, distribute and/or modify this document under the terms of the **GNU Free Documentation License**, Version 1.1 or any later version published by the Free Software Foundation; with no Invariant Sections, with no Front-Cover Texts, and with no Back-Cover.

You can obtain a copy of the GNU General Public License :

write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA