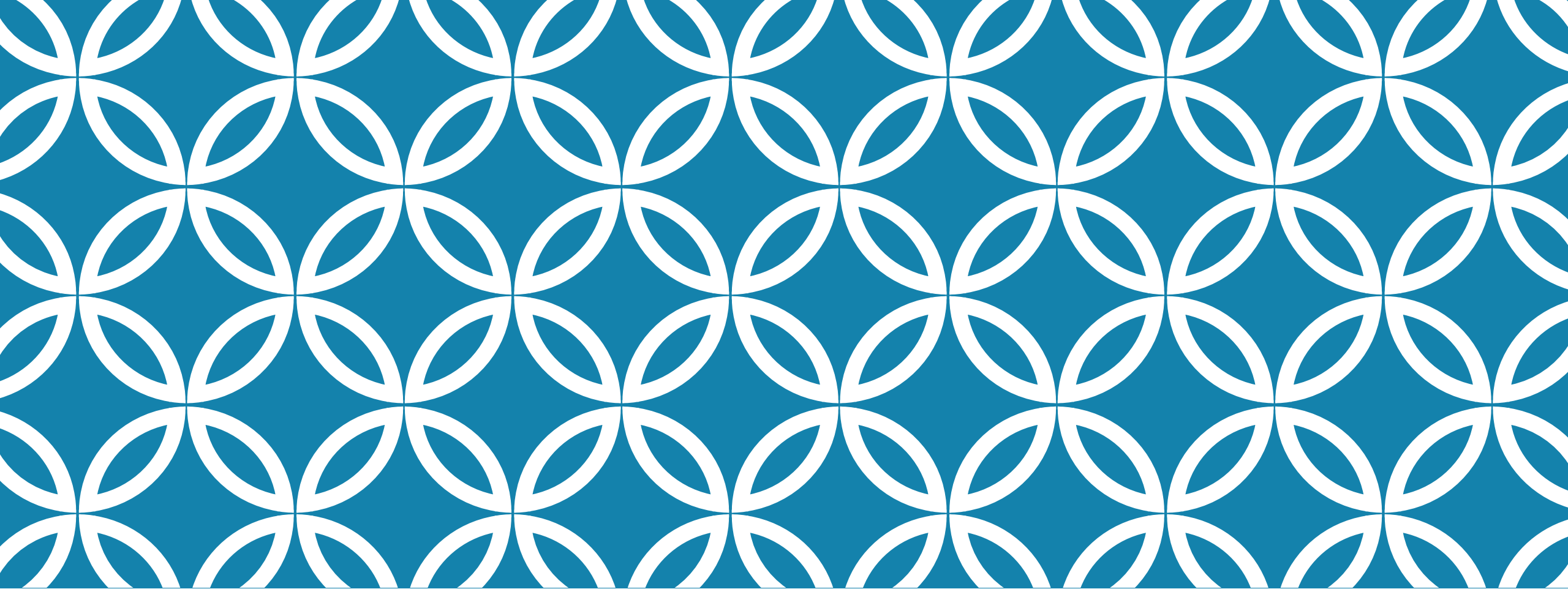




INTRODUCTION AU WEB

ARCHITECTURE DU WEB

Par Jean-Pierre Lozi
Basé sur les cours d'Andrea
Tettamanzi et Philippe Renevier

PRÉSENTATION/PROGRAMME DE L'UE

PRÉSENTATION DE L'UE

- 8 séances de cours (1h30), 10 séances de TP en salle machine
- 50% de la note : contrôle continu (présence + certains TP notés aléatoirement)
- 50% de la note : contrôle final
- Site web de l'UE : http://i3s.unice.fr/~jplozi/intro_web_2016/
- Contact : jplozi@unice.fr

PROGRAMME DE L'UE

- Ce que vous apprendrez (à titre indicatif !)
 - Qu'est-ce que le web ? Comment cela fonctionne-t-il ?
 - Créer des pages web statiques en HTML
 - Utiliser CSS (feuilles de style) pour mettre en forme vos pages
 - Programmation dynamique côté client en Javascript
 - Programmation dynamique côté serveur en PHP

QUESTIONS

- Avez-vous des notions...
 - De création de pages web ?
 - Des notions de réseaux ?
 - De programmation ?
- Posez des questions !

INTRODUCTION

DIFFERENCE ENTRE WEB ET INTERNET ?

À votre avis ?

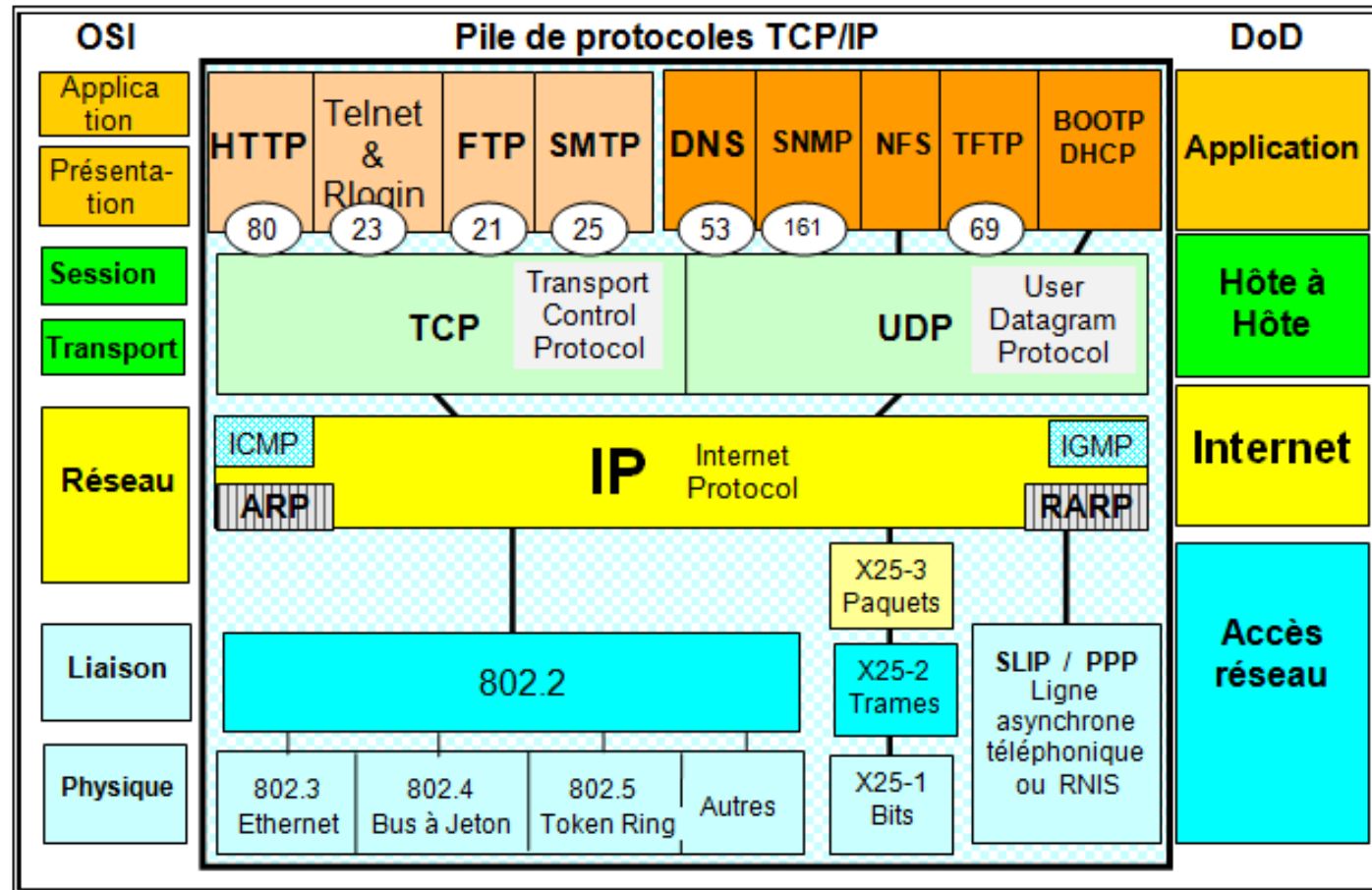
INTERNET

- Réseau global mondial, réseau de réseaux
- Définition : tout ce qui est basé sur le protocole de bas niveau TCP/IP !
- Tout ordinateur a une adresse IP (par exemple, 134.59.130.2)
- Souvent, ordinateurs identifiables par un nom de domaine / sous-domaine etc.
 - Plus facile à retenir pour les humains !
 - Par exemple, www.unice.fr → 134.59.204.9
 - eden.i3s.unice.fr → 134.59.133.6
- Traductions entre adresses IP et domaines par DNS (Domain Name System)

INTERNET

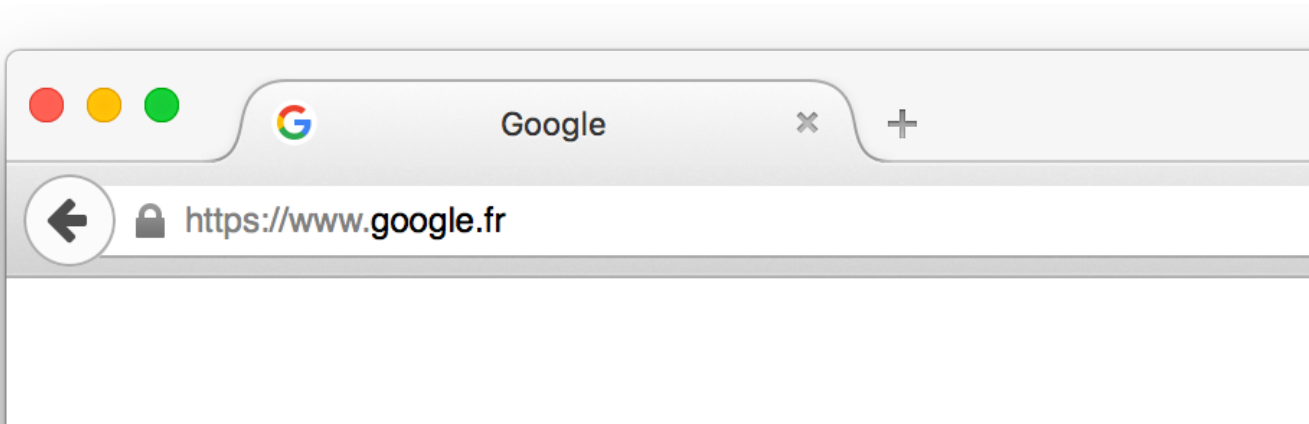
- Au-dessus de TCP/IP, de très nombreux protocoles
- Protocole HTTP : navigation de pages web dans votre navigateur
- Protocoles POP/IMAP/SMTP : envoi et réception d'e-mails
- Protocole BitTorrent : échange de fichiers pair à pair
- Protocole FTP : échange de fichiers
- Protocole IRC : messagerie instantanée
- ...

INTERNET



LE WEB

- Seulement le protocole HTTP ! (le plus utilisé depuis les années 90)
 - Ou HTTPS, S comme sécurisé...
- C'est-à-dire, les pages web que vous consultez dans votre navigateur...
- Connectées par des *liens hypertextes*... HTTP = HyperText Transfer Protocol



- <http://www.google.fr> ?
- Qu'est-ce que ça veut dire ?
- Différence entre http:// et www ?

LE WEB

- URL : Universal Resource Locator / URI : Universal Resource Indicator

- Différence entre les deux pas très bien définie

- Format dans un cas simple :

[protocole://sous-domaine.domaine/chemin/vers/fichier/fichier.html](#)

- Exemple :

http://www.lemonde.fr/idees/article/2015/09/11/climat-regarder-la-realite-en-face_4753004_3232.html

- Format plus général :

[protocole://sous-sous-domaine.sous-domaine.domaine:port/chemin/vers/fichier/fichier#ancree?param=valeur¶m=valeur](#)

- Ancre = position dans la page, paramètres = nom d'utilisateur par exemple
- Paramètres utiles pour les pages générées dynamiquement, en PHP par exemple

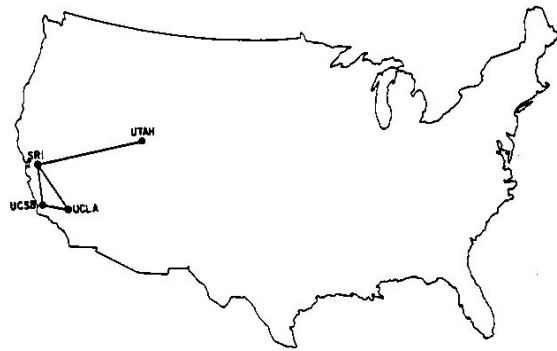
HISTORIQUE

UN PEU D'HISTORIQUE : INTERNET

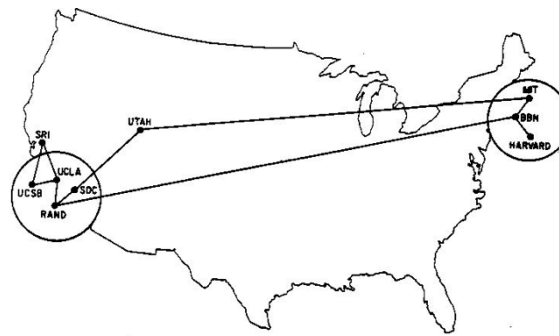
- Dans les années 60, création de réseaux utilisant la commutation de paquets
 - Le premier d'entre eux créé par Donald Davies au Royaume-Uni (National Physical Laboratory)
 - Division des données en paquets, avec en-tête pour le routage
- Création d'ARPANET en 1977, par le US Department of Defense
 - « Advanced Research Projects Agency Network »
 - Premier réseau à utiliser TCP/IP
 - TCP/IP créé par Robert Kahn et Vint Cerf
 - Incorporé des idées du réseau français CYCLADES (Louis Pouzin)
- Petit à petit, de plus en plus d'universités connectées
 - Emergence des premiers fournisseurs d'accès à la fin des années 80
 - Tout ceci avant le web !



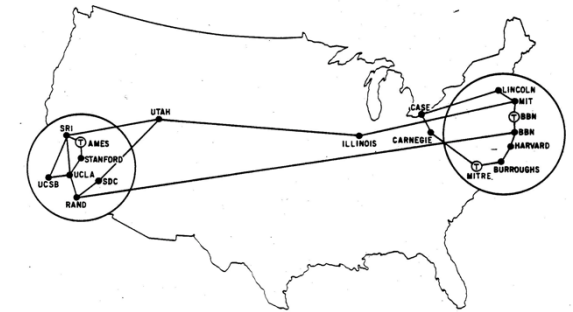
UN PEU D'HISTORIQUE : INTERNET



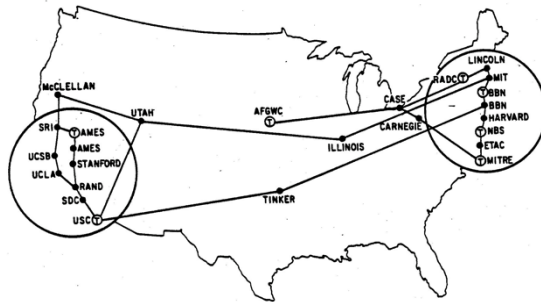
1969



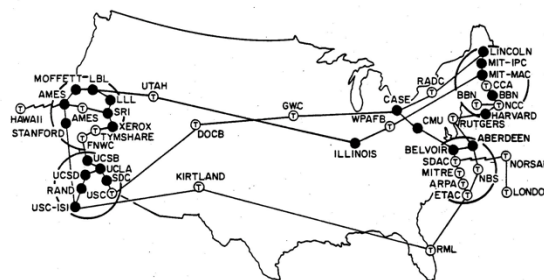
1970



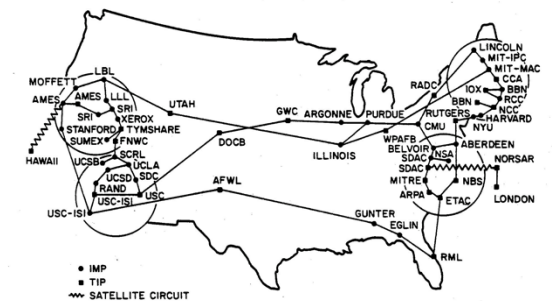
1971



1972



1973

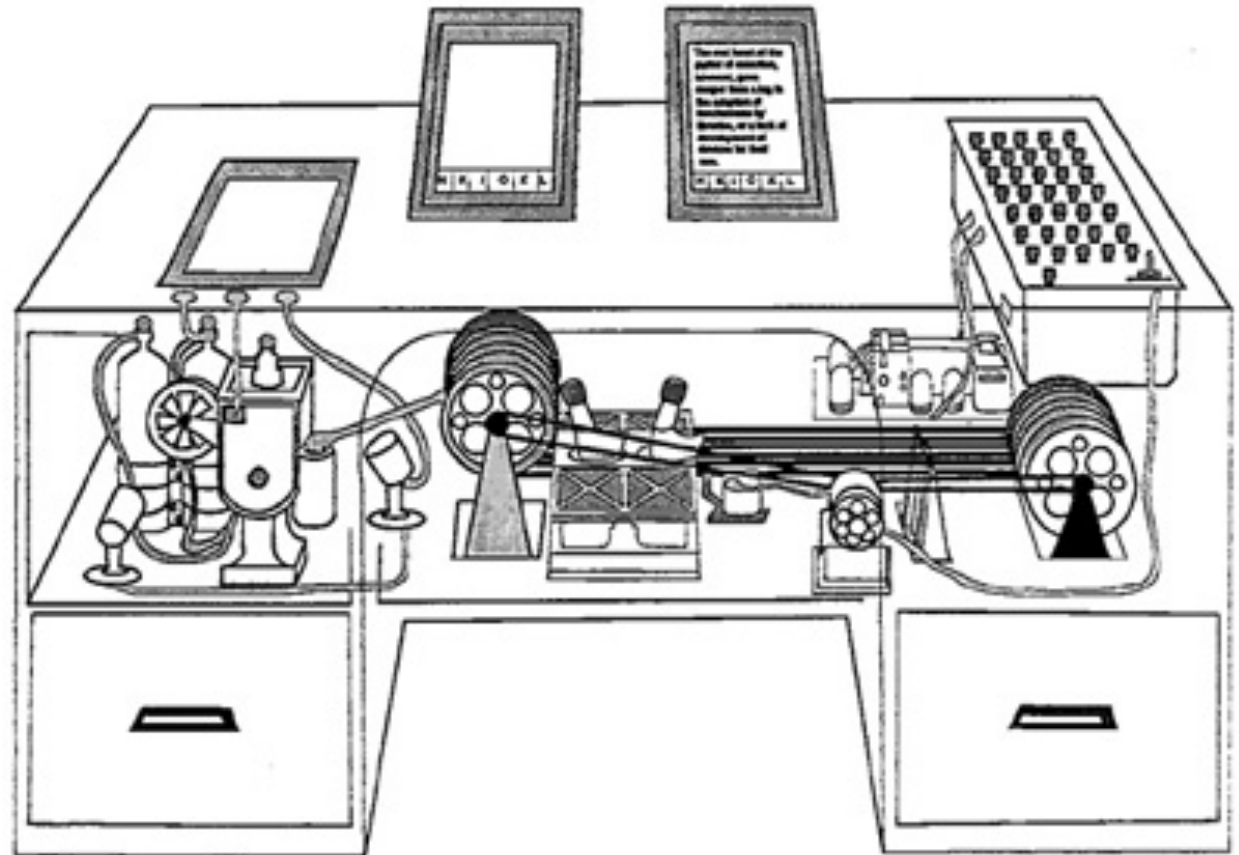
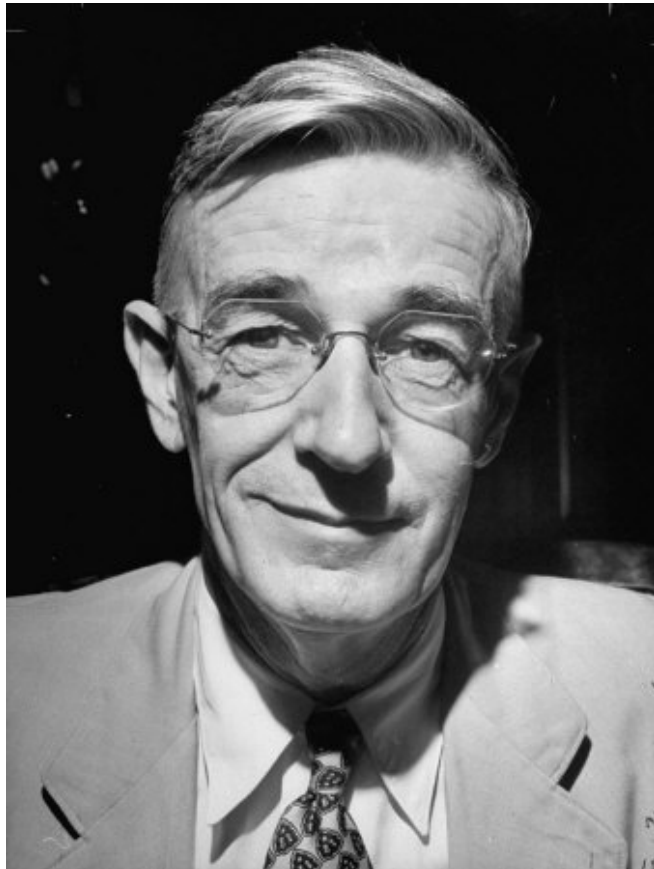


1974

UN PEU D'HISTORIQUE : LE WEB

- Idée principale du web : système hypertexte basé sur internet
 - Nombreuses pages interconnectées, passage de l'une à l'autre par des liens
- En 1945, Vannevar Bush (ingénieur) publie l'article « As we may think »
 - Dans Atlantic Monthly (pas une revue scientifique)
 - Il décrit le Memex, un ordinateur analogique, électro-mécanique, avec microfilms
 - Rélié à une bibliothèque permettant d'afficher des livres et projeter des films
 - Liens entre paires d'images de microfilms : première idée basique de lien

UN PEU D'HISTORIQUE : LE WEB



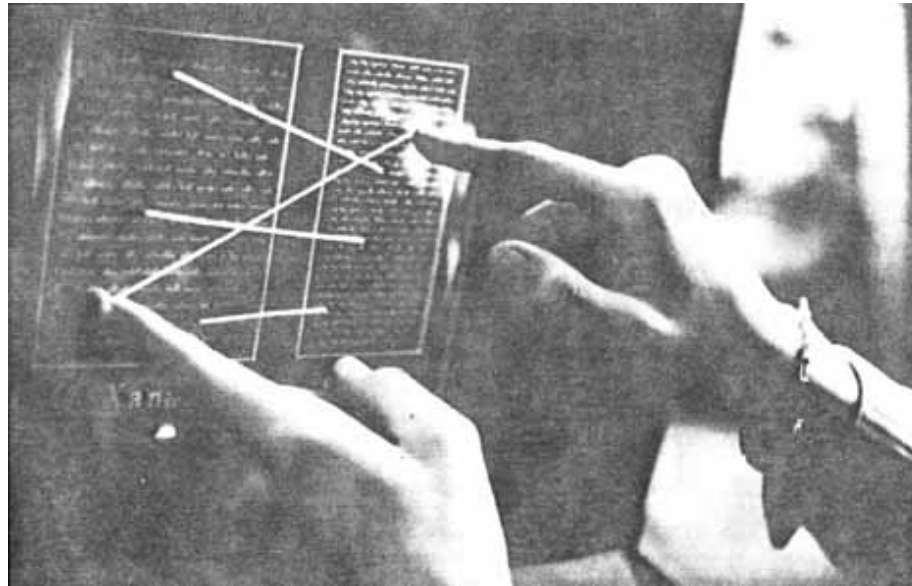
UN PEU D'HISTORIQUE : LE WEB

- Stanford, années 60 : Doug Engelbart crée le prototype NLS (oN-Line System)
 - Révolutionnaire : utilise liens, mais aussi un écran, une souris, fenêtre, programmes...



UN PEU D'HISTORIQUE : LE WEB

- En 1960, à Harvard, Ted Nelson crée le Projet Xanadu, premier système hypertexte
 - Inventeur du terme « hypertexte » !
 - En avance sur son temps, très complexe... Pas un succès mais influent.



UN PEU D'HISTORIQUE : LE WEB

- Dans l'idée de Ted Nelson : système hypertexte = ensemble de nœuds liés entre eux par des hyperliens permettant de passer automatiquement de l'un à l'autre.
- Nœud = unité minimale d'information, par exemple un paragraphe
- Les liens entre parties du texte gérés par ordinateur
 - Accès à l'information de manière associative
 - Navigation non-linéaire (pas évident à l'époque !), personnalisée
- Les nœuds peuvent aussi être audiovisuels : on parle de système hypermédia
- **Ted Nelson** : « Let me introduce the word 'hypertext' to mean a body of written or pictorial material interconnected in such a complex way that it could not conveniently be presented or represented on paper » (Proceedings of the 20th ACM National Conference, 1965).

UN PEU D'HISTORIQUE : LE WEB

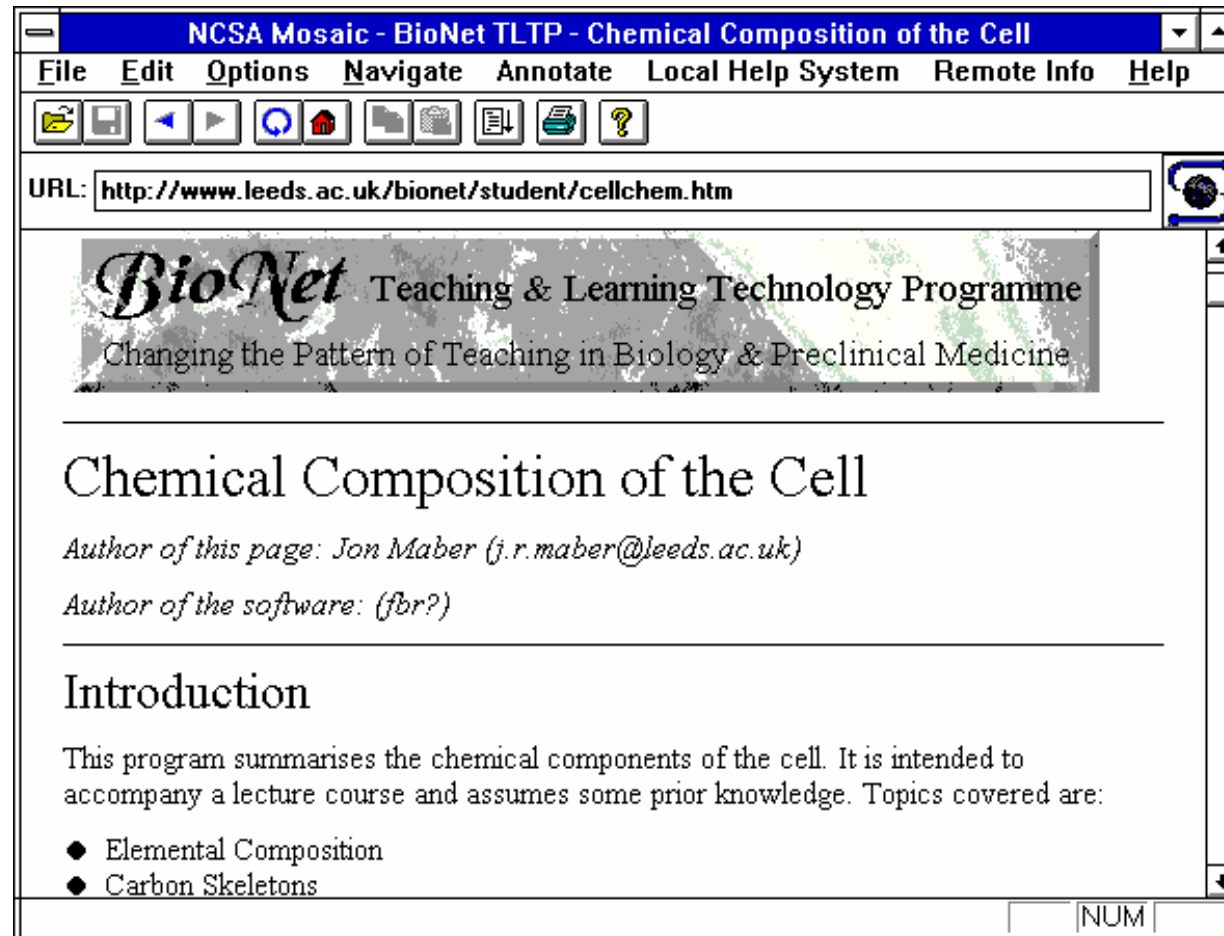
- En 1980, Tim Berners-Lee crée ENQUIRE, une base de données de personnes et logiciels
 - Premier logiciel à utiliser les liens hypertexte
 - **Au CERN, en Suisse**
- Peu après, connexion de machines du CERN au réseau TCP/IP
 - Tim Berners-Lee écrit « A large hypertext database with typed links »
 - Acquisition d'une station de travail NeXT par Tim Berners-Lee...
 - ...sur laquelle il a commencé à implémenter le World Wide Web
- En 1990, Tim Berners-Lee avait créé tous les outils nécessaires : le protocole HTTP, le langage HTML, le premier navigateur web, le premier serveur HTTP...
- 30 avril 1993 : le CERN verse toutes les technologies autour du WWW dans le domaine public



PREMIÈRE IMAGE PARTAGÉE SUR LE WEB



UN DES TOUS PREMIERS NAVIGATEURS: NCSA MOSAIC



QUELQUES BASES DE RÉSEAUX

RÉSEAUX : UNE ARCHITECTURE EN COUCHES

OSI (Open Source Interconnection) 7 Layer Model

Layer	Application/Example	Central Device/ Protocols		DOD4 Model
Application (7) Serves as the window for users and application processes to access the network services.	End User layer Program that opens what was sent or creates what is to be sent Resource sharing • Remote file access • Remote printer access • Directory services • Network management	User Applications SMTP	G A T E W A Y Can be used on all layers	Process
Presentation (6) Formats the data to be presented to the Application layer. It can be viewed as the "Translator" for the network.	Syntax layer encrypt & decrypt (if needed) Character code translation • Data conversion • Data compression • Data encryption • Character Set Translation	JPEG/ASCII EBDIC/TIFF/GIF PICT		
Session (5) Allows session establishment between processes running on different stations.	Synch & send to ports (logical ports) Session establishment, maintenance and termination • Session support - perform security, name recognition, logging, etc.	Logical Ports RPC/SQL/NFS NetBIOS names		
Transport (4) Ensures that messages are delivered error-free, in sequence, and with no losses or duplications.	TCP Host to Host, Flow Control Message segmentation • Message acknowledgement • Message traffic control • Session multiplexing	F I L T E R I N G P A C K E T	TCP/SPX/UDP	Host to Host
Network (3) Controls the operations of the subnet, deciding which physical path the data takes.	Packets ("letter", contains IP address) Routing • Subnet traffic control • Frame fragmentation • Logical-physical address mapping • Subnet usage accounting			
Data Link (2) Provides error-free transfer of data frames from one node to another over the Physical layer.	Frames ("envelopes", contains MAC address) [NIC card — Switch — NIC card] (end to end) Establishes & terminates the logical link between nodes • Frame traffic control • Frame sequencing • Frame acknowledgment • Frame delimiting • Frame error checking • Media access control	Switch Bridge WAP PPP/SLIP	Land Based Layers	Internet
Physical (1) Concerned with the transmission and reception of the unstructured raw bit stream over the physical medium.	Physical structure Cables, hubs, etc. Data Encoding • Physical medium attachment • Transmission technique - Baseband or Broadband • Physical medium transmission Bits & Volts	Hub		
				Network

NIVEAU PHYSIQUE

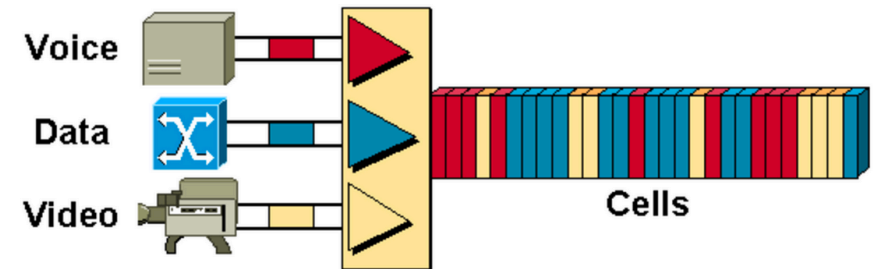
- 100/1000 Base-T filaire : technologie d'interconnection de machines
 - Standard le plus courant à l'heure actuelle
- Autre exemple, sans fil : le WiFi (IEEE 802.11)



NIVEAU DATA LINK : ETHERNET

- Ethernet : type de réseau le plus courant pour la plupart d'entre nous
 - Avant Ethernet, d'autres types de réseau : Token Ring (topologie en anneau), AppleTalk...
- Autres technologie, exemple : ATM (« Asynchronous Transfer Mode »)
 - Garanties spécifiques : pas de pertes de données
 - Réseaux nationaux de FAI

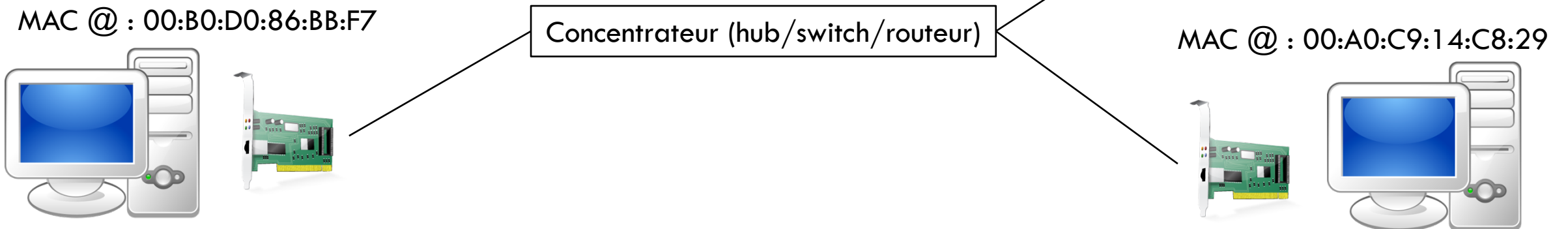
Introduction to ATM



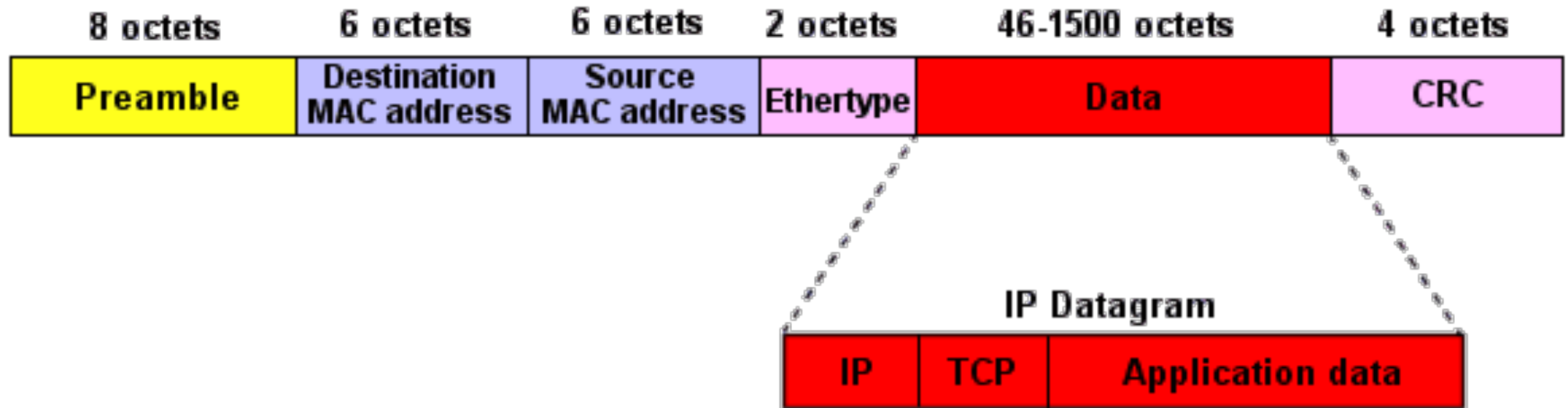
Asynchronous Transfer Mode - ATM, attempts to combine the best of both worlds -- the guaranteed delivery of circuit-switched networks and the robustness and efficiency of packet-switching networks.

NIVEAU DATA LINK : ETHERNET

- Une adresse MAC fixée pour chaque interface réseau (encodée en hard)
- Utilisé pour communiquer dans les réseaux locaux
 - Pas fait pour le routage à grande échelle
- En-dessous d'IP, avant même l'acquisition d'une adresse IP !



NIVEAU DATA LINK : ETHERNET



NIVEAU RÉSEAU : ADRESSES IP

- But : permettre la communication entre deux machines n'importe où dans le monde
- Deux protocoles d'adresse IP existent : version 4 et version 6
- IPv4 : par exemple, i3s.unice.fr a pour adresse 134.59.204.9
 - 4 294 967 296 adresses possibles en théorie, moins en pratique (pas allouées linéairement)
 - Problème : ces adresses s'épuisent, même si protocole encore le plus utilisé
- IPv6 : 2607:f0d0:1002:0051:0000:0000:0000:0004
 - Abrégé en 2607:f0d0:1002:51::4
 - 340 282 366 920 938 463 463 374 607 431 768 211 456 adresses !
 - 667 millions de milliards d'appareils connectés sur chaque millimètre carré du globe !

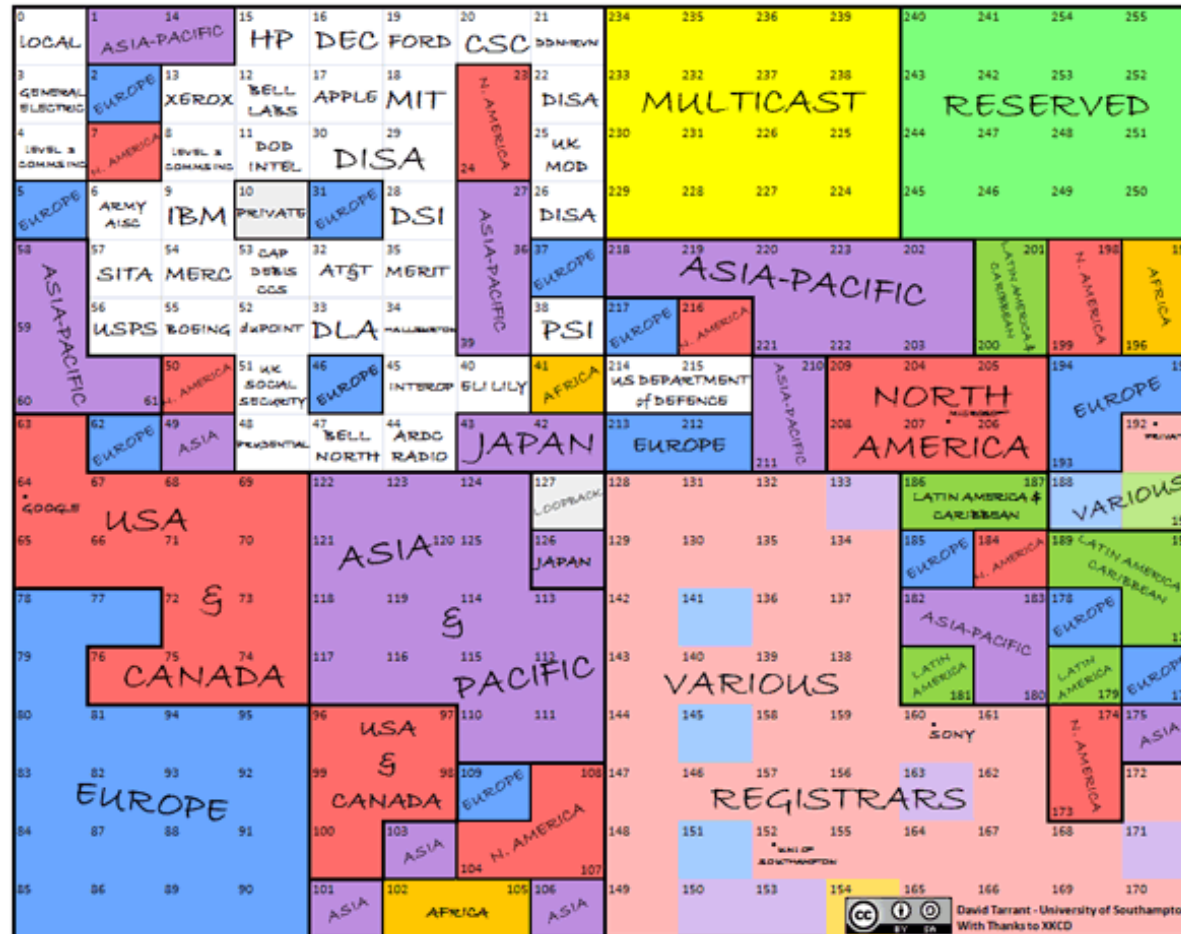
NIVEAU RÉSEAU : ADRESSES IP

- IP version 4 : chaque interface réseau est identifiée par une adresse 32 bits
- Cette adresse a une structure interne de type hiérarchique
 - Les bits les plus significatifs identifient un sous-réseau
 - Les bits les moins significatifs identifient une interface dans le sous-réseau
- En notation décimale, lisible aux humains :
 - Quatre entiers décimaux, séparés par un point (par exemple, 134.59.204.0)
 - Chaque entier représente un octet de l'adresse IP
 - Par exemple : 10000110 00111011 11001100 00001001
 - Valeur maximale d'un octet ?

NIVEAU RÉSEAU : ADRESSES IP

- Au début : attributions de toutes les adresses IP commençant par un octet à une entreprise, une université, une organisation (« Class A », « /8 blocks »)...
 - Par exemple : toutes les adresses en 17.x.x.x appartiennent à Apple !
 - À ce rythme là, <255 entreprises dans le monde pourraient utiliser internet !
- Maintenant, on est beaucoup plus économe avec les adresses IP
 - Stanford a rendu son bloc 36.x.x.x
 - On préfère attribuer des sous-réseaux de classe B (deux premiers octets fixes), de classe C (trois premiers octets fixes), ou des solutions intermédiaires
- Certaines adresses IP réservées pour les réseaux locaux
 - Par exemple, 192.168.0.x, ou 10.0.0.x sont très souvent utilisés
 - Machines pas directement accessibles de l'extérieur (sauf redirections de ports)

NIVEAU RÉSEAU : ADRESSES IP

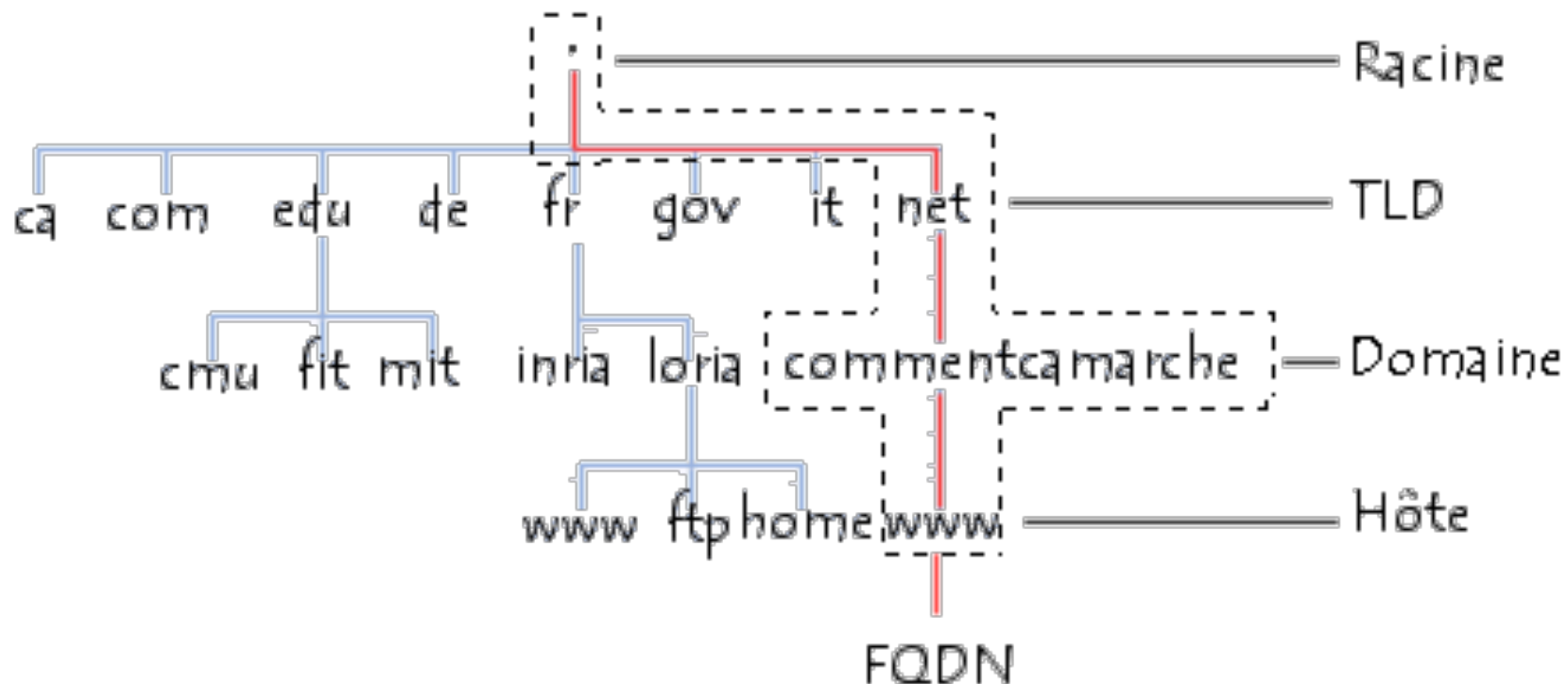


DNS: DOMAIN NAME SYSTEM

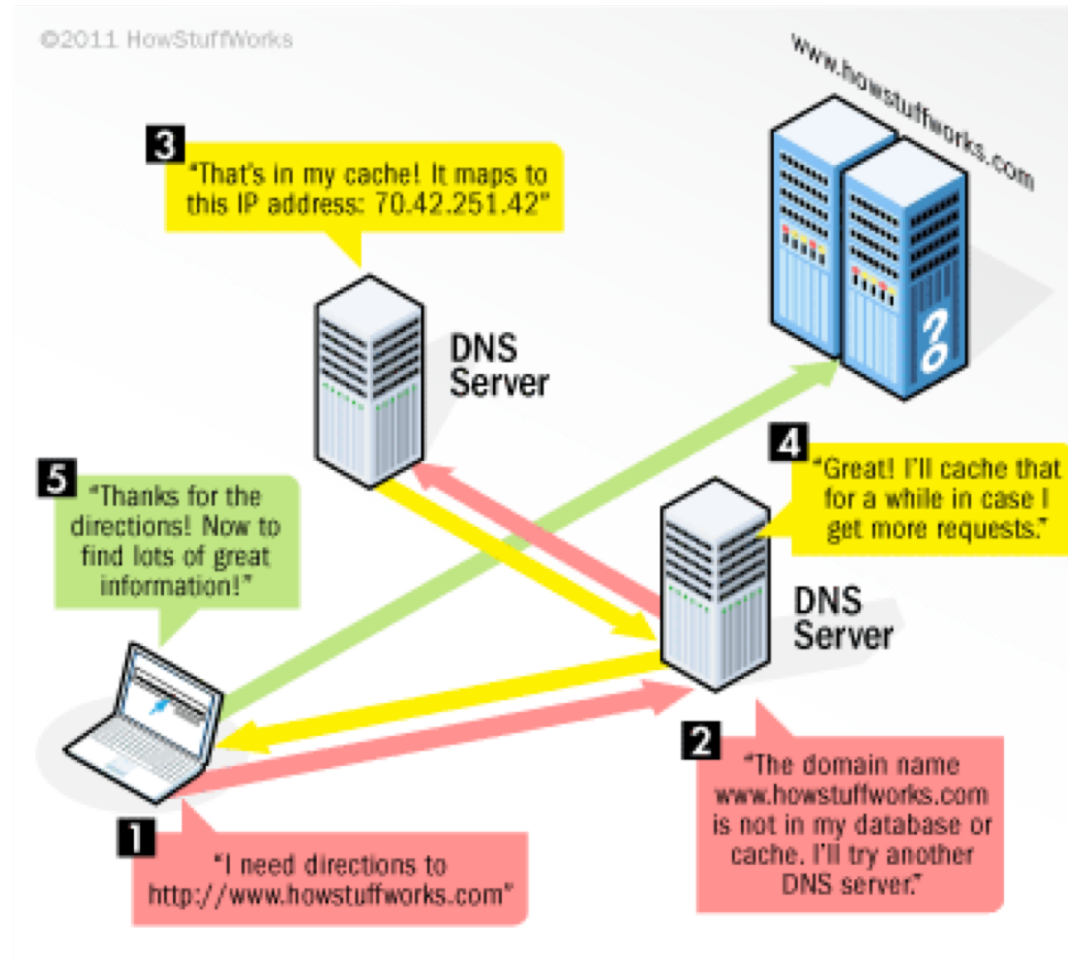
- Internet constitué de sous-réseaux (dizaines de milliers)
- IP permet l'accès aux machines par leur adresse IP
- Très difficile pour les humains de se souvenir des IPs des machines !
- DNS permet d'identifier une machine par un nom structuré représentatif de la machine et du réseau sur laquelle elle se trouve, par exemple : <http://deptinfo.unice.fr>
- Système mis en œuvre par une base de données distribuée au niveau mondial
- Les noms sont gérés par un organisme mondial : l'InterNIC et les organismes délégués : RIPE (Europe), NIC France, NIC Angleterre, etc.

DNS: DOMAIN NAME SYSTEM

- Exemple de www.commentcamarche.net :

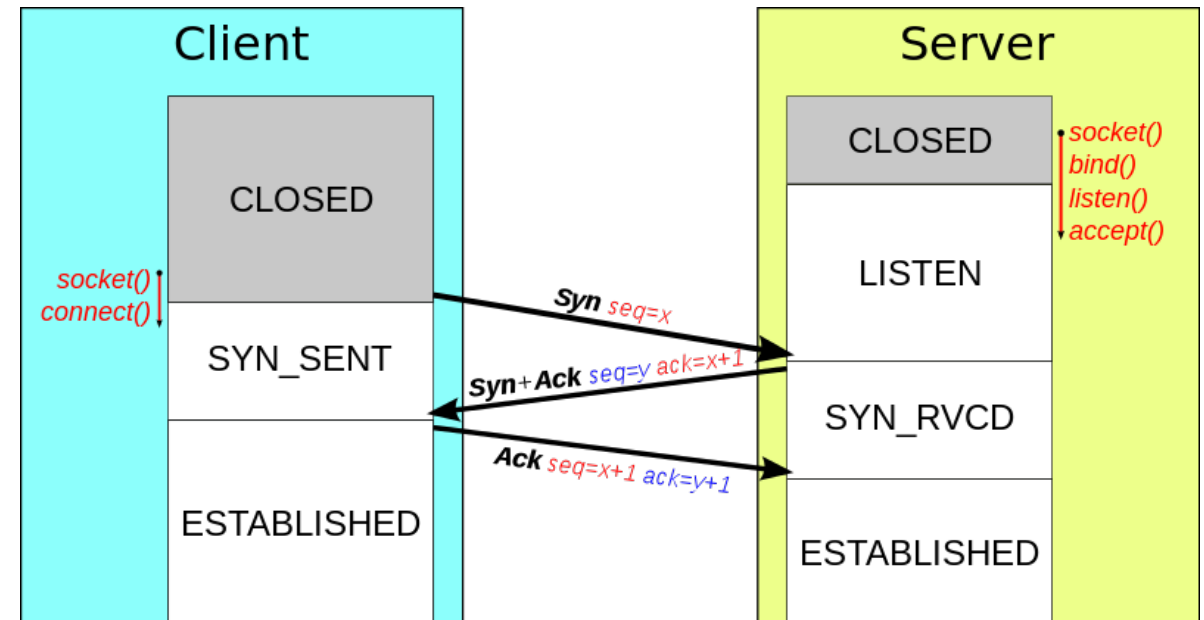


DNS: DOMAIN NAME SYSTEM

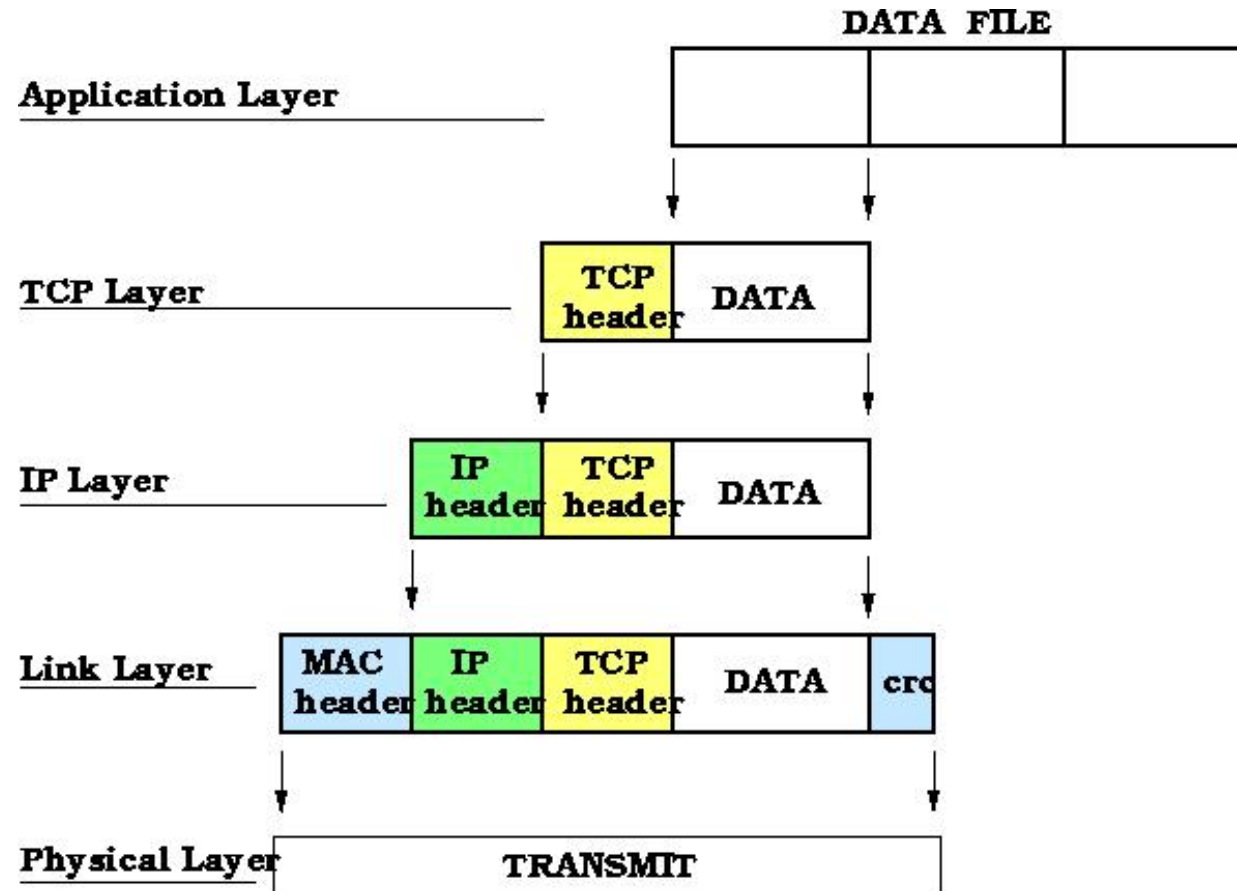


NIVEAU TRANSPORT: TCP

- Transmission Control Protocol
- Protocole de transfert fiable (paquets non dupliqués, dans le bon ordre,
- En mode connecté : idée de session
- Aussi idée de ports (e.g. 80 pour HTTP)



ETHERNET + IP + TCP : LES PAQUETS



LE WEB : PROTOCOLE HTTP

PROTOCOLE HTTP

- HTTP = HyperText Transfer Protocol
- Fonctionne au-dessus des protocoles TCP et IP
- Permet aux serveurs Web de transmettre des pages (ou images, etc.) aux clients
 - Serveur : hôte qui contient des pages web
 - Client : navigateur (Google Chrome, Mozilla Firefox, Safari, Internet Explorer...)
- Un serveur HTTP = un serveur HTTP (par exemple Apache) qui tourne sur une machine hôte (appelée aussi serveur !)
- Un client HTTP = un navigateur qui tourne sur la machine de l'utilisateur
- Une page web est un document, écrit en HTML (HyperText Markup Language)

PROTOCOLE HTTP

- On a vu tout à l'heure ce qu'était une URL. Rappel, par exemple :
[protocole://sous-sous-domaine.sous-domaine.domaine:port/chemin/vers/fichier/fichier#ancree?param=valeur¶m=valeur](mailto:sous-sous-domaine.sous-domaine.domaine:port/chemin/vers/fichier/fichier#ancree?param=valeur¶m=valeur)
 - Ancre = position dans la page, paramètres = nom d'utilisateur par exemple
 - Paramètres utiles pour les pages générées dynamiquement, en PHP par exemple
- Après le nom de domaine/port, chaque élément entre les slashes est une séquence de caractères alphanumériques, plus les deux caractères « - » et « _ ».
- Les espaces sont remplacés par « + » ou « %20 »
- Séquences d'échappement pour les caractères spéciaux : %xx où xx est le code hexadécimal du caractère en ASCII. Par exemple, %20 = espace, %7E = « ~ », %2B = « + », %25 = « % »...
- Attention ! Masqué par les navigateurs récents

PROTOCOLE HTTP

- Plus de séquences d'échappement dans les URLs :

Espace	%20
“	%22
#	%23
%	%25
&	%26
(	%28
)	%29
+	%2B
,	%2C

.	%2E
/	%2F
:	%3A
;	%3B
<	%3C
=	%3D
>	%3E
?	%3F
@	%40

[	%5B
\	%5C
]	%5D
^	%5E
‘	%60
{	%7B
	%7C
}	%7D
~	%7E

PROTOCOLE HTTP

- HTTP est le protocole le plus utilisé depuis 1990
- Version 0.9 : version très basique, seulement une commande GET pour récupérer un document, sans recevoir aucune information sur son type.
- Version 1.0, en 1996 : permet de transférer des messages avec des en-têtes décrivant le contenu du message en utilisant un codage de type MIME (« Multipurpose Internet Mail Extensions »). [RFC 1945]
- Version 1.1, en 1997 : quelques améliorations, dont une meilleure gestion du cache
- Version 2 : toute récente (2015), peu utilisée

PROTOCOLE HTTP

- Transfert de fichiers identifiés grâce à une URL entre un navigateur (le client) et un serveur web (Apache) :
 - Le navigateur effectue une requête HTTP
 - Le serveur traite la requête puis envoie une réponse HTTP
- Commandes HTTP :

Commande	Description
GET	Requête la ressource située à l'URL spécifiée (éventuellement données par URL avec ?a=b&...)
HEAD	Requête seulement de l'en-tête de la ressource située à l'URL spécifiée
POST	Envoi de données au programme situé à l'URL spécifiée
PUT	Place la ressource à l'URL spécifiée (en pratique rarement utilisé)
DELETE	Suppression de la ressource située à l'URL spécifiée (en pratique rarement utilisé)

PROTOCOLE HTTP

■ Format d'une réponse HTTP :

```
VERSION-HTTP CODE EXPLICATION<crlf>
EN-TETE : Valeur<crlf>
. . .
EN-TETE : Valeur<crlf>
Ligne vide<crlf>
CORPS DE LA REPONSE
```

■ Un exemple :

```
HTTP/1.0 200 OK
Date : Sat, 15 Jan 2000 14:37:12 GMT
Server : Microsoft-IIS/2.0
Content-Type : text/HTML
Content-Length : 1245
Last-Modified : Fri, 14 Jan 2000 08:25:13 GMT
```

PROTOCOLE HTTP

■ Codes réponse HTTP :

Code	Message	Description
10x	Message d'information	Codes non utilisés dans les versions 1.0/1.1 du protocole
20x	Réussite	La transaction s'est bien passée (page trouvée...)
30x	Redirection	La ressource ne se trouve plus à l'emplacement indiqué
40x	Erreur due au client	La requête était incorrecte
403	FORBIDDEN	Accès à la ressource interdit
404	NOT FOUND	Un classique ! Le serveur n'a pas trouvé la ressource à l'adresse indiquée
50x	Erreur due au serveur	Un erreur interne au serveur s'est produite

PROTOCOLE HTTP : RÉSUMÉ

