L'Internet et ses outils

Thomas LEDUC

Thomas.Leduc@masi.ibp.fr

http://quartz.dgs.jussieu.fr:8080/

Mars 1997

Contents

INTERNET signifie INTERconnection of NETworks...

1 Bref historique

1.1 La petite histoire

1964	RAND Corporation, projet d'un réseau de machines informatiques décentralisé, à transmission par paquets.
1969	Defense Advanced Research Projects Agency (DARPA) crée l'ARPANet, le premier réseau informatique, il utilise le Network Control Protocol (NCP) protocole de communication proche de X25 (le protocole de Transpac)
1973	création de l'Internet et de son protocole (non propriétaire) Transmission Control Protocol/Internet Protocol (TCP/IP). Premières recherches d'interconnexion de réseaux
1980-81	création des réseaux BITNET et CSNET (IBM). Réseau parallèle connectant des réseaux IBM. Mise sur pied du réseau des LISTSERVs. Utilisation de TCP/IP sur les réseaux commerciaux X25.
1983	fondation du IAB qui prend la relève de la DARPA pour la recherche et se divise en deux groupes IETF et IRTF
1986	la National Science Foundation (NSF) décide de relier les réseaux de ses 5 principaux centres de recherche et crée ainsi le NFSNet à base de l'IP comme l'ARPANet. Les réseaux émergent aussi en Europe.
1990	croissance rapide du réseau des réseaux...les serveurs doublent chaque année depuis 88.

1.2 Petit comparatif

Teletel est un réseau (presque) limité à la France, à serveurs vidéotex, à faible débit, équipement spécifique nécessaire (Minitel ou émulateur sur ordinateur), contrôle de l'utilisation de tous les services, dépourvu de la notion de dialogue direct entre utilisateurs et à facturation non forfaitaire pénalisante.

Le Net : réseau immense par le nombre d'utilisateurs, de stations, de liaisons et par l'importance du débit...Aucun modèle ou constructeur d'équipement n'est imposé...sinon le protocole TCP/IP ! Abolition des distances et des frontières. Égalité des ordinateurs et des individus, en particulier dans leur désignation...qui sait que quartz.dgs.jussieu.fr est une Axil 311, kaos.idris.fr un Cray T3D ? Pas de signe extérieur de richesse...pas de contrôle ou surveillance des services échangés entre stations et individus. Liberté totale d'expression, information directe souvent sans filtrage ni réécriture, logiciels gratuits...pas de facturation individuelle (en milieu universitaire du moins...), pas de facturation au volume mais au forfait

1.3 Vue du côté de la recherche française

projet Cyclades, projet RCP (mise en oeuvre des premiers logiciels X25, développement par le biais de la société commerciale Transpac), tentative REseau des UNIversités et de la Recherche (REUNIR)...

A partir de 1984, IBM finance un European Academic Research Network (EARN) basé sur des protocoles propriétaires IBM (NJE, RSCS...). Équipements et logiciels chers, administration lourde et centralisée, nombre de stations limité, seule la messagerie était utilisable entre machines hétérogènes...IBM se désengage financièrement en 1988 et l'arrivée de RENATER et des réseaux régionaux achève le mourant !

Autres tentatives telles que FNET, ARISTOTE à la fin des années 80...

A la même époque, l'informatique des campus se décentralise et éprouve le besoin d'un réseau. Les réseaux Ethernet de campus apparaissent alors, se fédèrent en réseaux régionaux (RERIF = plaque francilienne, SYRHANO = plaque haute-normande...) puis en réseau national. Protocoles supports envisagés : TCP/IP, DECNET, SNA, X25... Naissance du Réseau NATional pour l'Enseignement et la Recherche (RENATER) dont l'exploitation est confiée à France Telecom. voir http://www.urec.fr/Renater/

1.4 L'influence du monde UNIX

Unix n'est pas à l'origine de TCP/IP mais l'a tout simplement intégré comme mode privilégié de communication réseau

2 Topologie

Pas d'organisme central unique, pas de propriétaire unique mais une communauté d'intérêt, une association d'utilisateurs qui défraient eux-mêmes les cots d'entretien et de modernisation des infrastructures. Réseau de réseaux, décentralisé, composé de tronçons de capacités diverses
Sur ce réseau, il existe des artères principales à gros débit (les backbones).
Le Net a intégré partiellement plusieurs réseaux plus anciens comme Bitnet (d'IBM) ou UUCP (technologie Unix) qui utilisent d'autres conventions d'adressage que l'IP
La transmission des données dans le Net se fait en mode commuté par paquets (l'information est découpé en fragments ou datagrammes ou packets et le trajet des packets est réévalué à chaque instant de sa trajectoire pour prendre la décision la plus rapide). Chaque datagramme contient, en plus de ses centaines ou milliers d'octets de données, l'adresse du destinataire, de l'émetteur, un numéro de port...

2.1 adresses logiques et résolution de noms

Le protocole IP est un protocole intermédiaire entre les protocoles applicatifs et les protocoles de transmission de bas niveau comme FDDI, Ethernet...Il peut fonctionner sur des liaisons à faible comme à haut débit.

Principes de base du protocole IP chaque station a une adresse (l'adresse IP), celle-ci est unique au monde. Il y a donc communication possible entre 2 machines sans ambiguïté, garce à ces fameuses adresses. une adresse IP (v4) se présente comme concaténation de 4 entiers naturels inférieurs strictement à 256 (un octet) et séparés par des points. Elle est stockée sur 4 octets. une partie des 4 octets de l'adresse est l'adresse de réseau et l'autre l'adresse locale.
Note : le premier groupe de numéro peut-ître plus ou moins grand (on parle de classe A, B ou C) de façon à réserver plus ou moins de place pour le restant de l'adresse, en fonction de la taille des sous-réseaux.

noms Fully Qualified Domain Name (FQDN) adresses IP sont peu mnémoniques, il y a donc un système de conversion automatique nom de machine-adresse IP. Noms domainisés : quartz.dgs.jussieu.fr, cestac.masi.fr...pour chaque "top level domain", il y a un gestionnaire officiel des noms (le NIC) ce qui garantit l'unicité des noms.

résolution de nom (ou correspondance nom-adresse IP) consultation et mise à jour manuelle du fichier /etc/hosts sous Unix...limité ! Automatisation par le Domain Name System (DNS) :

chaque domaine de niveau 2 (organisation) stocke une base de données sur une machine du domaine, le serveur de noms du domaine, contenant le nom et l'adresse de toutes les machines du parc. Cette base est gérée localement par le superviseur du Local Area Network (LAN) et consultable depuis toute station du Net.
Chaque "top level domain" possède un serveur de noms qui recense les noms et adresses IP des serveurs de noms de tous les domaines de niveau 2 qui dépendent de sa "juridiction"... En France, c'est le rôle de la machine inria.inria.fr (192.93.2.1).
A la racine, une machine connaît les noms et adresses IP des serveurs de noms de chaque "top level domain"... En fait, toutes ces bases de données sont dupliquées pour assurer la continuité du service même en cas de problème !

2.2 Les "top level domains"

à 3 lettres com entreprises commerciales
edu institutions universitaires
gov institutions gouvernementales
org organisations non lucratives
mil institutions militaires
net entreprises de réseaux

à 2 lettres fr, uk, jp, de... (codes ISO-3166 des pays)

2.3 Fournisseurs d'accés

voir http://www.nic.fr/Prestataires/
voir http://www.urec.fr/France/Classement/Themes/Internet/Fournisseurs%20d'acc
%26egrave;s/Fournisseurs%20d'acc%26egrave;s.html
ou encore (UNGI - Carte des providers au niveau national) :
http://www.imaginet.fr/~ gmaire/france.htm

types de liaisons

- Liaison Spécialisée (LS) : connexion via une liaison louée à France Telecom (et un équipement de routage à votre extrémité), d'un débit compris entre 9,6 Kbit/s et 34 Mbit/s.
- X25 : connexion via une liaison louée à Transpac (et un équipement de routage à votre extrémité), d'un débit compris entre 9,6 Kbit/s et 2 Mbit/s.
- Réseau téléphonique commuté (RTC) : Connexion via un modem et le réseau téléphonique commuté ou Numeris, les paquets IP étant transportés jusqu'à votre extrémité (protocole SLIP ou PPP).
- Applicatif : connexion via un modem et le réseau téléphonique, un applicatif particulier est mis à votre disposition par le fournisseur du service pour accéder à son serveur. Ce type de connexion permet d'accéder à un serveur situé chez le prestataire, ce dernier étant relié à l'Internet.
- Numéris (RNIS) : Connexion via une liaison numéris à France Telecom (et un équipement de routage à votre extrémité), d'un débit compris entre 2 canaux de 64Kbits à 30 canaux selon le choix.

2.4 RENATER

Voir http://www.urec.fr/Renater
RNI = Réseau National d'Interconnexion de RENATER (2 à 34 Mbit/s)
NTI = Noeud de Transit International de RENATER
plaques régionales en 1995 : Rerif (IdF), Syrhano (Rouen), Vikman (Cherbourg), OR (Rennes), Aquarel (Bordeaux), Remip Octares (Toulouse), R3LR (Montpellier), R3T2 (Marseille), Aramis (Lyon), Lothaire (Metz, Nancy), Noropale (Lille) liaisons externes vers US NSFNet, Corse et DOM-TOM par X25, Europe (EBONE), pays méditerranéens...
Carte du RNI, voir http://www.urec.fr/Renater/auj/reseaunational.html

Services réseaux proposés Raccordement IP permanent à des débits de 64 kbit/s à 2 Mbit/s
Pas de raccordement par réseau commuté (Numeris...), puisqu'il existe des opérateurs privés.
Tarifs : voir http://www.urec.fr/Renater/auj/procedureraccordt.html

Petits exemples un étudiant grenoblois surfe sur le Web du MIT à Boston depuis son campus, cela correspond à : information transite sur l'Ethernet 10 Mbits/s du laboratoire, passe par routeur sur le réseau du campus (fibre optique), puis sur le réseau régional ARAMIS au travers d'un routeur ; à Lyon , le réseau régional est connecté au RNI de RENATER via un autre routeur, le message passe alors sur RENATER et remonte vers Paris (liaison optique 34 Mbit/s) pour accéder au noeud central de RENATER. Là, l'information passe alors dans la partie NTI puis emprunte une liaison transatlantique (câble optique Bretagne-Washington)...
un particulier surfe sur le FTP anonymous de l'IBP depuis chez lui PC + modem + liaison téléphonique standard, accès au fournisseur d'accès (ex : Oleane...), puis au NTI de RENATER (liaison louée par France Telecom au "provider" Oleane)...

Traceroute

```
hector{leduc}3: traceroute j16.inktomi.com
traceroute to j16.inktomi.com (209.1.12.113), 30 hops max, 40 byte packets
 1  hera.ibp.fr (132.227.61.135)  2.749 ms  1.003 ms  0.985 ms
 2  kerbere.ibp.fr (132.227.60.3)  2.918 ms  4.288 ms  3.258 ms
 3  r-jusren.reseau.jussieu.fr (134.157.254.254)  5.541 ms  4.729 ms  4.61 ms
 4  r-rerif.reseau.jussieu.fr (192.44.54.1)  3.925 ms  4.005 ms  3.499 ms
 5  danton1.rerif.ft.net (193.48.58.121)  10.812 ms  3.735 ms  4.599 ms
 6  stlambert.rerif.ft.net (193.48.53.49)  3.856 ms  6.379 ms  5.95 ms
 7  stamand3.renater.ft.net (195.220.180.9)  5.562 ms  6.208 ms  4.928 ms
 8  stamand1.renater.ft.net (195.220.180.43)  5.523 ms  3.802 ms  4.49 ms
 9  rbs1.renater.ft.net (195.220.180.50)  5.188 ms  4.959 ms  5.811 ms
10  raspail-ip.eurogate.net (194.206.207.18)  5.539 ms  7.116 ms  4.399 ms
11  Reston.eurogate.net (194.206.207.5)  84.939 ms  81.32 ms  82.228 ms
12  gsl-sl-dc-fddi.gsl.net (204.59.144.198)  266.642 ms  359.409 ms  90.759 ms
13  sl-dc-8-F/T.sprintlink.net (198.67.0.8)  86.975 ms  177.197 ms  242.392 ms
14  sl-mae-e-H2/0-T3.sprintlink.net (144.228.10.42)  85.556 ms  87.819 ms  84.43 ms
15  mae-east.agis.net (192.41.177.145)  165.462 ms  99.606 ms *
16  h0-0.1.losangeles1.agis.net (204.130.243.45)  370.59 ms  327.641 ms  368.798 ms
17  a1-0.1003.santaclara2.agis.net (206.62.13.245)  370.716 ms  346.893 ms  327.248 ms
18  pb-nap.exodus.net (198.32.128.25)  376.104 ms  403.66 ms  413.682 ms
19  * * *
```

2.5 Le Multicast BackbONE (MBONE)

est un réseau virtuel qui utilise l'Internet pour véhiculer un trafic multicast IP (technique permettant à un même datagramme émis par une source d'être reçu par plusieurs destinataires) utile pour la diffusion d'informations à grande échelle (vidéoconférences de meetings IETF...)
Pour simuler le routage de paquets multicast IP, on a employé une technique d'encapsulation dans des paquets unicast IP (pour les transmettre au moyen d'une liaison dite tunnel). La machine cible, équipée elle d'un logiciel supportant le routage multicast IP se charge alors de la distribution en local. L'ensemble des ordinateurs reliés par des liaisons tunnel forme le MBone. L'information transmise par MBone est acheminée dans Internet de manière standard.
Note : FMBone est la version multipoint française.
voir http://www.nic.fr/images/mbone-topology.gif

3 Les institutions du Net

Réseau basé sur des principes coopératifs et décentralisés...Tout sous-réseau le constituant assure lui-même son fonctionnement et participe aux coûts de raccordements aux voisins. Par transitivité tout nouvel arrivant sur le réseau profite du maillage mondial...Tendance à la formalisation de l'organisation du Net.

3.1 Les instances mondiales

l'Internet SOCiety (ISOC) coordonne depuis 1991 le développement du réseau. Ensemble de grands opérateurs (dont France Telecom), de grands fournisseurs de matériels et logiciels réseaux, de grands constructeurs de stations, d'organismes de normalisation et d'adhérents individuels.

l'Internet Architecture Board (IAB) Groupe de quelques individus (les "sages", volontariat) pilotant le évolutions techniques de protocoles et d'architecture mondiale (dépend directement de l'ISOC). Christian Huitema en fut le président...

l'Internet Engineering Task Force (IETF) Groupes de volontaires (les "gourous") proposant des évolutions techniques et rédigeant les Request For Comment (RFC) spécifications techniques consensuelles. L'IAB coordonne l'activité des groupes de l'IETF qui se réunissent "physiquement" trois fois par an).

l'Internet Research Task Force (IRTF) Structure équivalente à l'IETF pour les évolutions à long terme, plus théorique, moins connue.

l'Internet Assigned Numbers Authorithy (IANA) Pilotée par l'IAB, elle gère l'attribution des numéros de ports (les "wellknown ports"), des types et sous-types MIME...Elle délègue l'attribution des numéros de réseau aux Network Information Centers (NIC) mais maintient l'annuaire des top level domains (à la racine).
voir le NIC France : http://www.nic.fr/ (jusqu'à 30 requêtes par secondes sur leur DNS)
charte de nommage de la zone .fr, voir http://www.nic.fr/Procedures/nommage.html

3.2 A l'échelle européenne : le Réseaux IP Européens (RIPE)

voir http://www.ripe.net/
C'est un "super NIC" un peu comme l'Internic (domaines internationaux, voir http://ds.internic.net/), l'APNIC (http://www.apnic.net/)...

3.3 Les RFC, FYI, FAQ, émoticons...

Indexation des RFC à l'Institut Pasteur, voir
http://www.pasteur.fr/other/computer/RFC/

4 Les types de services possibles - les fonctionnalités de l'Internet

4.1 TCP procure un service de communication fiable entre applications

il est complémentaire à l'IP. L'IP se charge d'acheminer l'information jusqu'à la machine cible, TCP établit un circuit virtuel entre applications distantes et introduit pour cela la notion de port. TCP joue un rôle de multiplexeur d'applications (plusieurs applications tournent souvent sur une même machine). Il assure également la fiabilité de la transmission (réordonancement des paquets, retransmission des paquets perdus). Le couple numéro IP/ port TCP permet d'adresser un service particulier sur une machine bien identifiée.

4.2 Les "wellknown port numbers"

ftp (21, TCP), telnet (23, TCP), Simple Mail Transfer Protocol (25, TCP), gopher (70), finger (79, TCP), www (80)...
(voir /etc/services sous Unix)

4.3 Terminal distant (telnet...)

4.4 Transfert de fichiers (ftp anonymous ou non)

voir http://www.urec.fr/doc/ftp_fr.html

4.5 Courrier électronique, serveurs SMTP

SMTP (Simple Mail Transfert Protocol) est le protocole de base de messageries Internet ; il est souvent implanté en natif sur les messageries propriétaires. POP3 (RFC 1725) et Imap4 (RFC 1730) sont d'autres protocoles plus spécifiques. Post Office Protocol permet de faire du traitement de messages pour clients dépourvus de connexion permanente). Internet Message Access Protocol est plus récent et plus complexe...

4.6 Listes de diffusion (LISTSERV)

Françaises : voir http://sir.univ-rennes1.fr/LISTES/

4.7 Groupes de discussion : USENET et les News (communication asynchrone)

Plus de 8000 groupes de news (newsgroups) au mois de juin 95...chaque serveur de news peut décider quel groupe ou branche de groupes il désire recevoir...La propagation des news se fait de proche en proche, pas de serveur central. Principe très simple : tout le monde peut envoyer une contribution à condition d'avoir le client adhoc...et de ne pas se tromper de forum de discussion !
Usenet est un ensemble de serveurs supportant ces forums, créé en 1979 par 2 étudiants de l'UNC et de l'Université Duke.
Voir (pour les listes francophones) http://www.fr.net/news-fr/liste.html
Voir http://www.w3.org/pub/DataSources/News/Groups/

4.8 Internet Chat Relay (IRC) (communication interactive, principe du talk à plusieurs)

Moins formel que les news.

4.9 Network Time Protocol (NTP)

Mise à jour automatique des horloges internes à partir d'un serveur du Net

4.10 Les Services d'Informations Répartis (SIR)

4.10.1 Archie

catalogue de serveurs ftp anonymous, outil développé par l'université McGill du Canada. Plus de 3000 serveurs Ftp anonymous au mois de juin 95 (voir http://archie.univ-rennes1.fr ou http://www.urec.fr/archie.html)

4.10.2 Wais

BD indexée dont l'annuaire mondial était centralisé sur quake.think.com

4.10.3 Gopher

BD indexée et distribuée - les serveurs se connaissent et savent sur lequel d'entre eux se trouve l'information recherchée - Veronica est un service d'indexation Gopher comme Archie est un service d'indexation de ftp anonymous système de distribution hiérarchisé de l'information (voir gopher://info3.scs.unr.edu:8117/7)
Conçu par l'université du Minnesota pour rendre accessible l'information au niveau d'un campus, il s'est très vite répandu à l'échelle planétaire. Le "Gopher space" est constitué de documents qui peuvent être consultés en parcourant une arborescence. Chaque nom ou partie de l'arborescence possède un nom et une adresse réseau...on saute de serveur en serveur de façon transparente.

4.10.4 WWW

ici, ce ne sont plus les machines qui se connaissent mais les documents de tous types qui ont des liens entre-eux indépendamment des machines hôtes.

5 L'avènement du Web

Le service qui popularisa l'Internet est incontestablement le World Wide Web. Version réseau du concept d'hypertexte.

5.1 Tim Berners-Lee et la petite histoire

Le fondateur du Web Tim Berners-Lee (diplômé d'Oxford University, actuellement en poste au Laboratory for Computer Science du MIT et directeur du World Wide Web Consortium W3C) présente en 1989, au sein du CERN (Genève), un "projet hypertexte global"...qui prendra par la suite le nom de World Wide Web. Il développe le premier client (un éditeur-navigateur hypertexte wysiwyg) et le premier serveur en environnement NeXTStep à partir du mois d'octobre 1990. De 1991-1993, profitant du retour des utilisateurs et avant de rejoindre le MIT, il affine les spécifications de la définition de l'URL, du protocole HTTP et du langage HTML.

L'idée maîtresse du W3 telle que définie sur le serveur du W3C "The World Wide Web is conceived as a seamless world in which all information, from any source, can be accessed in a consistent and simple way."

Historique rapide de l'hypertexte et du Web...

1945	Vannevar Bush (conseiller de Roosevelt) publie une note concernant des toiles conceptuelles d'information
1965	Ted Nelson donne naissance à l'hypertexte (principe : un texte sur support informatique n'a pas à être présenté de façon linéaire, il peut être composé de fractions de documents reliés entre eux par des liens sémantiques. Il y a donc plusieurs entrées ou grille de lecture possibles pour un même document...cf. le principe des index) et à un logiciel de navigation hypertexte qui ne fonctionnera jamais...
1987	logiciel d'Apple (Hypercard) basé sur l'hypertexte
mars 89	Tim Berners-Lee publie l'article "Hypertexte et le CERN" - Conseil Européen pour la Recherche Nucléaire - Genève.
oct. 91	premier serveur Web fonctionne au CERN (mode texte)
jan 93	une cinquantaine de serveurs Web dans le monde. Le CERN sort le premier navigateur graphique (tkWWW, violaWWW, midasWWW)
fév 93	Marc Andreessen édite Mosaic (pour Unix) au sein du National Center for Supercomputing Applications (NCSA, université de l'Illinois)
mars 93	Marc Andreessen et Jim Clark (fondateur de Silicon Graphics) développent Netscape
juillet 93	le CERN, le MIT et l'INRIA fondent le W3C (WWW Consortium)
début 94	qui du gopher ou du Web va gagner la guerre du Net (devenir le service le plus populaire) ?
oct 94	Netscape est lancé en béta-test sur PC, Mac et Unix
fév 95	4 millions d'utilisateurs de Netscape (75% des navigateurs sont des Netscape)
nov 95	Netscape v 2.03b est lancé, il supporte les news, le courrier électronique en réception et émission, Java et JavaScript...
déc 95	Microsoft Internet Explorer 2.0
oct 96	MSIE 3.0 et Netscape 3.0 se font une guerre acharnée...ActiveX contre Netscape One

5.2 Universal Resource Locator : l'adressage universel des ressources

Une URL est une adresse plus complète qu'une adresse IP ou FQDN en ce sens qu'elle comprend : l'adresse FQDN, le type de service, l'emplacement sur le serveur, le nom du fichier.
un accès unifié à l'information
forme générale schématique :

$service : //nomUtilisateur : motDePasse@serveur : port/chemin/nomDuDocument?arg1+arg2\ldots$

Protocoles possibles : http, https, ftp, gopher, mailto, news, newsrc, nntp, telnet, rlogin, tn3270, wais, file...
Exemples :
snews://secnews.netscape.com/netscape.navigator
news:fr.petites-annonces.immobilier
Attention file (qui permet d'accéder à un fichier "en local") n'est pas normalisé, ainsi :

```
sur PC file:///c|tmp/fichier.txt
sur Unix file:/home/quartz/info/leduc/fichier.txt
```

5.3 HyperText Transfer Protocol

Protocole sans état (proche de gopher)

5.3.1 HTTP v 0.9

Principe client effectue une demande de connexion TCP au serveur sur un port particulier (port 80 par défaut) client émet sa requête (commande GET suivie de l'URL demandée, terminée par les caractères CR et LF), serveur retourne le résultat ou contenu de la ressource demandée et coupe ensuite la connexion
Exemple explicite (soumission manuelle de requête par telnet) telnet quartz.dgs.jussieu.fr sur le port 8080 puis GET /TMP/

5.3.2 HTTP v 1.0

extension de la version précédente. Nouvelles fonctionnalités : HEAD (permet d'obtenir des informations sur un document précis), PUT (permet de mettre à jour un document précis, rarement implémentée), DELETE (suppression de documents, non implémentée), POST (permet de créer de nouveaux documents, poster un article pour un forum de discussion, traitement de réponse à un formulaire...)
La réponse HTTP en provenance du serveur contient un certain nombre d'informations dont le statut, l'en-tête, le descriptif ressource (type d'encodage, longueur, type Multipurpose Internet Mail Extension = MIME...).

Statut de la réponse statut de la transaction informe le client du déroulement de la procédure de traitement de la requête par le serveur...
Statuts de classe 2 : requête reçue comprise et traitée (ex : 200 OK, 201 Created - méthode post de création de nouvelle ressource, 204 No Content)
Statuts de classe 3 : traitement incomplet de la requête (nouvelle adresse connue pour la ressource, 301 Moved Permanently...)
Statuts de classe 4 : erreur client (400 Bad Request - erreur de syntaxe, 401 Unauthorized - identification préalable du client, 403 Forbidden, 404 Not Found...)
Statuts de classe 5 : erreur serveur (500 Internal Server Error - erreur propre au serveur, 501 Not Implemented - fonctionnalité non supportée par le serveur, 503 Service Unavailable - surcharge ou maintenance...)

5.4 HyperText Markup Language

langage utilisé pour la diffusion de documents par les serveurs Web. Il s'agit d'une classe de documents au format Standard Generic Markup Language (SGML). SGML est une méthode normalisée (norme ISO en 1986) pour représenter l'information contenue dans un document indépendamment du système de saisie et de traitement, qui propose un métalangage pour construire des langages de balisage.

5.5 Quelques navigateurs - logiciels de lecture du service Web (clients)

Lynx (mode texte), Arena, Mosaic, Netscape, Hotjava...
Quelques trucs et astuces pour l'utilisation du navigateur Netscape raccourci pour certains types d'URL, série des about (+ view- source), petit interpréteur en ligne de commandes JavaScript (mocha, LiveScript ou Javascript), raccourcis clavier et combinaisons de touches "spéciales"

5.6 Quelques serveurs du service Web (httpd)

le NCSA httpd, le CERN httpd (APACHE), Website, MS Internet Information Server, NetSite (Netscape Communication Corporation)
voir le site : http://webcompare.iworld.com

5.7 CGI (Common Gateway Interface), SSI (Server Side Includes, fonctionnalité d'inclusion de documents à la volée)...

5.8 Les nouvelles technologies : JavaScript, Java, Visual J++, Active X, VRML

6 Recherche dans le cyberespace

augmentation exponentielle du nombre de documents de tous types sur l'Internet...risque d'asphyxie. Juin 96, Matthew Gray (MIT) annonce environ 230 000 sites Web référencés (à raison de 40 serveurs en moyenne par site). En janvier 94, on en dénombrait moins de 1000 ! Hotbot annonçait alors 54 millions de documents référencés...

6.1 Un peu d'histoire

En 1994, J Yang et D Filo étudiants à Stanford ont mis leurs bookmarks (classés par catégorie) sur le réseau, et ont appelé leur site : YAHOO. Point de passage vite devenu obligé, les webmasters de sites non référencés ont écrit aux étudiants pour se retrouver sur le fameux site...Création d'une société privée au début 95 (amélioration des algo), conservation de la gratuité de l'accès aux pages, mais publicité (avec hyperliens vers le site de l'annonceur) payante, coté en bourse en mars 96 ! En octobre 96, 60 personnes travaillent sur l'annuaire et les associations se multiplient : avec Digital (et Alta-Vista), Reuters (cours de la bourse en ligne). 9 millions de pages par jour en juin 96, 14 millions de pages par jour à la fin octobre 96 !

6.2 Robots (moteurs de recherche ou search engines)

Lycos (l'ancêtre du temps de Yahoo, venu de Carnegie Mellon), Alta-Vista, Excite, Hotbot, Infoseek, WebCrawler...

6.3 Annuaires

problème : comment construire une "hiérarchie" de classement correcte ?
Yahoo, Magellan (www.mckinley.com), Gamelan (référence Java)...

6.4 Exemple AltaVista

www.altavista.digital.com dispose d'un index impressionnant et une panoplie d'options pour la recherche.

mode simple de recherche clef non accentuée, en minuscule (recherche indépendante de la casse), sinon, recherche en fonction de l'accent ou de la majuscule. +clef1+clef2 sans espace force la recherche de documents contenant les deux mots. "château de Versailles" force la recherche sur la phrase. wildcard. Limitation par site : clef host:ibp.fr

mode avancé de recherche options plus riches (AND, OR, NOT, NEAR...)

6.5 La balise $< META >$

$$< META\ \ name = "description"\ content = "cours\ sur\ l'Internet" >$$

$$< META\ \ name = "keywords"\ content = "Internet, web, search\ engines" >$$

6.6 L'option link - qui pointe sur ma page ?

link:http://www.ibp.fr

6.7 Fichier robots.txt - Robots Exclusion Standard

Pour "empêcher" tous les robots de visiter toute l'arborescence d'un site, placer le fichier robots.txt à la racine du serveur avec le contenu suivant :

```
User-agent: *
Disallow: /
```

Pour faire une sélection plus nuancée, procéder par exemple de la façon suivante :

```
# ceci est un commentaire
# tout est permis au robot WebCrawler
User-agent: webcrawler
Disallow:
# tout est interdit au robot Lycra
User-agent: lycra
Disallow: /
# les repertoires TMP et DOC ne doivent pas etre indexes par les
autres robots
User-agent: *
Disallow: /TMP
Disallow: /DOC
```

7 Conseils pour la gestion des "méga-sites Web"

hypertexte = invention géniale et diabolique... elle consiste à lier des documents entre-eux par le biais de signes, mots, phrases, images...en nombre infinis ! Attention à la péremption desdits liens et aux sites humainement ingérables !

Répertorions les types de liens possibles jardin (papillonner de page en page à son gré), promenade (ordre de visite préétabli, mode d'emploi, chronologie...), index (listes de produits...), glossaire (par définitions des termes rencontrés), relatifs (page suivante, précédente...), thématiques (pour les sites multi-services)

Donc nécessité d'informatiser la gestion de la production du site (car le langage HTML n'est pas un outil de production...de même que ce document n'a pas été écrit en postscript directement)
Limiter le nombre de pages statiques par interfaçage à une base de données (accessible par formulaire + CGI)
Rationaliser la mise en page (pas forcément uniformser)
Générer automatiquement les index et les sommaires Traduire les contenus en plusieurs langues
Gérer automatiquement les liens de types glossaire, relatif, navigation...

Offrir la possibilité aux auteurs de soumettre un contenu directement par le Web en toute sécurité, indépendant de l'HTML (de la nécessité de drivers)
Supporter en standard les nouvelles technologies...

En toile de fond l'idée d'un SERVEUR UNIVERSEL (SGBD Relationnel-Objet, Internet, Intranet regroupant toutes les bases de données du site Web propre et de l'ensemble de l'entreprise...)

8 Perspectives

8.1 IPv6

1992	premières réflexion sur la saturation de l'adressage IP au sein de l'IETF
1994	premières recommandations à partir des RFC
1996	premiers prototypes éditeurs et constructeurs
1998	début de la migration IPv4 vers IPv6

La norme actuelle IP v4 est insuffisante : espace adressable trop restreint (sur 4 octets), pas de QoS (mauvaise adaptation au temps réel et donc au multimédia), manque de robustesse en matière de sécurité.
Le ST Datagram Mode ou IPv5 est resté inutilisé.
Contraintes pour l'IPng : régler le problème de l'espace d'adressage (16 octets ou 1024 adresses IP théoriques par mètre carré de la surface terrestre !), garantir une certaine QoS (champs "identification de flux" et "priorité" qui permettent d'envisager un traitement temps réel de certaines trames), extensions envisageables (routage imposé, intégrité et authenticité des messages, informations de fragmentation et réassemblage...), assurer l'adaptation dynamique (mobilité des machines connectées et réadressage dynamique), inclure en natif le multicast...
Tout en respectant les 2 grands principes de l'IP : pas de réservation de bande passante, autonomie de chacun des noeuds du réseau.
Nécessaire coexistence des 2 versions du protocole (pas de date D), le réseau doit pouvoir accepter et gérer les 2 types de trames

8.2 Asynchronous Transfer Mode

possibilité de garantir capacité et QoS par connexion (IP, lui, est un protocole sans connexion)

8.3 HTTP v 1.1

8.4 L'idée d'un client universel

Orientation résolumment groupware du Netscape Navigator 4 avec messagerie, gestion d'agendas, groupes de news, traitement de texte collaboratif (?)...intégrés !

9 TD

Téléchargement et installation d'un navigateur à partir d'un site FTP anonymous proche (que l'on identifiera d'abord à l'aide d'un service Archie - si l'on possède un tel client du moins !).
Paramétrage correct du client Web (bookmarks, notions de Proxy, d'helpers-apps, mailer et lecteur de news intégrés...)