

# Structure et évolutions de l'Internet

## *quelles conséquences pour la documentation en ligne ?*

JOËL MARCHAND

[jma@math.jussieu.fr](mailto:jma@math.jussieu.fr)

GDS 2754 Mathrice

# Où en est l'Internet ?

- le réseau de tous les réseaux IP
- 35 ans
- environ 1 milliard d'internautes
- de l'ordre de 300 millions de machines connectées
- plusieurs dizaines de millions de serveurs Web
- plus de 4 milliards de pages Web indexées dans Google
- tendance à l'ubiquité de l'accès (sans-fil)
- le problème du débit est quasiment réglé (Renater, ADSL, sans-fil)

# L'adressage IP

- le même que celui conçu il y a 35 ans et donc en partie obsolète
- une adresse est composée de 4 octets, soit 4 nombres compris entre 0 et 255
- exemple : 134.157.13.2
- on définit un réseau IP comme étant un ensemble d'adresses IP consécutives
- exemple 1 : 134.157.13.0/24
- exemple 2 : 134.157.12.64/26

# Le nommage des machines

- nécessité d'un nommage des machines autrement que par leur numéro
- structure : des mots composés de lettres et de chiffres, séparés par des «.», et terminés par un Top Level Domain (TLD)
- .com, .net, .org, .info, .biz sont des TLD internationaux
- les autres sont nationaux : .fr, .de, .uk
- les Etats-Unis sont une exception : .edu, .gov, .mil

- c'est l'outil totalement indispensable à l'Internet
- il assure en permanence la correspondance entre un nom et un numéro IP
- attention : ce n'est pas bijectif
- une machine peut avoir plusieurs adresses IP, ce qui permet d'augmenter le nombre de services sur la même machine

- inversement, à un nom peuvent être associées plusieurs adresses IP, pour répartir la charge d'un service sur plusieurs machines
- notion de DNS direct : nom  $\longrightarrow$  adresse IP
- notion de DNS inverse : adresse IP  $\longrightarrow$  nom

- une application serveur (un *daemon*) utilise un couple constitué d'un port de connexion et d'une adresse IP, sur lesquels elle va répondre à une requête vers ce couple
- on parle alors de port TCP, nombre compris entre 1 et 65536
- cela permet d'avoir plusieurs applications sur la même machine, pour réaliser plusieurs services (Web, messagerie, telnet ...)
- lorsque l'internaute demande une connexion vers un service d'une machine, il indique en fait le port vers lequel il souhaite parler à la machine

- exemples
  - ◆ le Web : 80
  - ◆ relever sa messagerie par POP : 110
  - ◆ envoyer du courrier : 25
- ces ports TCP sont normalisés internationalement pour la partie basse (< 1024) et libres d'usage pour la partie haute (entre 1024 et 65536)
- attention : un serveur Web peut être relié à un port autre que 80



# Nouveaux problèmes de l'Internet

- souci principal : besoin accru de sécurité
- réponse principale : la fermeture des ports en entrée, voire en sortie, pour limiter les machines visibles et les services accessibles
- conséquence : une machine sans visibilité directe depuis l'Internet ne peut pas héberger de service
- en cas de fermeture également en sortie, elle ne peut pas accéder directement à certains services
- autre souci récurrent : manque d'adresses IPv4

# Les réponses à ces problèmes

- de manière générale, ce sont les CRI qui fixent la politique
- filtrage des ports (filtrage IP) par mise en place de pare-feu (firewall)
- système de traduction d'adresses avec usage d'adresses privées
- imposition de l'usage de systèmes *proxy*
- concentration sur les serveurs du CRI des services ouverts sur tout l'Internet
- restriction de ces services vers le strict minimum : LE serveur Web, LE serveur de messagerie
- étude du passage en adressage IPv6

# Autres problèmes du moment

- le spam
- les usurpations d'adresses de messagerie électronique
- les virus, les chevaux de troie

- le Web n'est **pas** l'Internet - c'est **un** service de l'Internet
- les protocoles : HTTP, HTTPS
- les serveurs HTTP : Apache, Internet Information Service (IIS) ...
- possibilité de mettre plusieurs sites sur la même machine : notion de VirtualHost

- les navigateurs Web : Internet Explorer, Netscape, Mozilla ...
- notion de plug-ins, applications externes
- norme MIME associant un type à un fichier (lors de l'envoi) et permettant d'associer une application pour l'ouvrir (lors de la réception)
- le langage HTML et ses nombreuses évolutions (XHTML)
- le Web dynamique : les langages de programmation Perl, PHP, Java, ASP, .net ...
- le couplage avec les bases de données : MySQL, PostgreSQL, Oracle ...

- la syntaxe générique des URL

`protocole://user:passwd@serveur:port/dossier/fichier.html`

- exemple 1 :

`http://www.math.jussieu.fr/informatique/`

- exemple 2 :

`https://www.urec.cnrs.fr/securite/corres-secu/index.html`

# Rappels sur le Web - 4/4

- **exemple 3** : `http://www.site.org:8080/`
- **exemple 4** : `ftp://jma:bidon@ftp.math.jussieu.fr/`
- **avenir** : les bureaux virtuels, les environnements numériques de travail, les services Web

- un proxy-cache est un système de mandataire, d'intermédiaire, pour réaliser une connexion entre un poste client et un serveur
- on parle aussi de relai applicatif
- désormais la fonction *cache* est peu utile
- c'est la fonction *proxy* qui devient prépondérante
- vu du serveur, le client est le proxy, **pas** le poste client qui a émis la requête



- c'est une solution aux problèmes d'accès à des sites (et notamment pour des sites à accès restreint, comme ceux des revues) :
  - ◆ lorsqu'on veut mutualiser un accès Internet à bas débit
  - ◆ lorsqu'on veut protéger les machines clientes
  - ◆ lorsqu'on veut une seule machine en sortie vers l'Internet
  - ◆ lorsqu'on veut accéder depuis l'extérieur de son laboratoire
  - ◆ lorsqu'on manque d'adresses IPv4

- comment on peut limiter l'usage d'un proxy ?
  - ◆ par interdiction par expressions régulières sur les URL
  - ◆ par restriction des adresses IP autorisées
  - ◆ par authentification lors de l'usage du proxy
  - ◆ par authentification préalable à son usage

- c'est une notion complexe, qui grosso modo s'apparente à une carte d'identité électronique
- ils sont émis par des entités réputées de confiance et produits suivant une procédure normalisée et fiable, qui garantit in fine la validité du certificat
- ils se matérialisent par de petits fichiers, que l'on peut donc aisément transporter, pour prouver son identité
- ils peuvent être utiles pour signer et chiffrer des messages électroniques, ou pour prouver son identité lors d'un accès à un site Web

- cela pourrait être une piste d'avenir pour l'accès aux journaux électroniques, car ils permettraient d'authentifier de manière fiable (nominativement ou par établissement) les accès aux sites Web des éditeurs
- des contraintes
  - ◆ difficulté d'usage
  - ◆ difficulté de stockage
  - ◆ besoin d'évolution forte de la part des éditeurs
  - ◆ difficulté de mise en œuvre à grande échelle

## Quelques prévisions personnelles - 1/2

- dans de nombreux endroits, il va devenir impossible d'héberger un service (et donc un site Web) dans un laboratoire
- dans pas mal d'endroits, il sera imposé de passer par un proxy pour accéder au Web
- ceci est raisonnablement motivé par le fait que les problèmes majeurs de sécurité ne vont pas s'arranger et que bon nombre sont liés aux sites Web

- le système de restriction aux journaux électroniques par adresses IPv4 est en fin de vie
- il ne répond pas à l'envie croissante d'accès depuis d'autres lieux que le bureau : domicile, laboratoire d'une mission, congrès ...
- il y a peu de chances qu'à court terme, les certificats soient déployés pour l'accès aux revues
- il serait bien en revanche de s'orienter vers une solution de proxy, avec authentification
- c'est ce qu'on a fait dans la maquette Mathrice

# Démonstration

- utilisation d'un proxy
- utilisation d'un certificat
- utilisation de la maquette Mathrice