

World Wide Web

Laurent Réveillère

Enseirb
Département Télécommunications

`Laurent.Reveillere@enseirb.fr`
`http://www.enseirb.fr/~reveille`

(D'après le cours d'Emmanuel Nauer)

Plan

- ◆ Généralités sur le Web
- ◆ Principes de fonctionnement
 - HTTP
 - HTML
 - URL
 - Proxy et Cache
- ◆ Accéder au Web

Généralités

- ◆ Origine
 - CERN, 1989, projet World Wide Web
 - Explosion en 1993 avec l'apparition du premier navigateur : Mosaic
- ◆ Le web, c'est quoi donc ?
 - Système d'information universel
 - » **hypertexte** : documents reliés entre eux par des *liens* accessibles à partir d'*ancres*
 - » **multimédia** (hypermédia) : composé de différents média (textes, sons, images, vidéo, ...)
 - » **sur Internet**
 - Accès gratuit et facile

Intérêts

◆ Intérêts du web

- Gigantesque espace d'information résultant de la grande facilité de publication d'information :

Tout le monde peut mettre des informations à disposition sur le Web !

- Facilité d'utilisation (à priori !) et convivialité


Différents problèmes ...

- ◆ Organisation des informations
 - Accroissement du nombre de serveurs, du nombre de pages, du volume des données
- ◆ Utilisation du réseau
 - Important échange de données : données volumineuses (images, sons, vidéos, ...)
- ◆ Maîtrise de la présentation des documents
- ◆ Domaine en pleine expansion / maturation
 - Développements importants de logiciels clients, de logiciels serveurs, de logiciels d'édition, de langages ...
 - Développement de nouveaux services : commerce électronique, WAP
- ◆ Vers la fin de la gratuité ?
 - Produits commerciaux, informations payantes

- ◆ World Wide Web Consortium (W3C)
 - Nombreux organismes
 - » privés (Microsoft, Netscape, Sun, IBM, ...)
 - » publics (INRIA, MIT, ...)
 - Objectifs :
 - » développement et promotion du web
 - » travaux de standardisation (HTML, HTTP...), développement, ...
 - <http://www.w3c.org>

Grands principes régissant le Web

- ◆ Modèle client / serveur
 - Le client demande un document
 - Le serveur fournit (ou non) le document
- ◆ Protocole utilisé entre clients et serveurs web :
HyperText Transfert Protocol (HTTP)
- ◆ Langage de définition de document :
HyperText Markup Language (HTML)
- ◆ Schéma de nommage des ressources :
 - **Uniform Ressource Locator (URL)**



HyperText Transfert Protocol

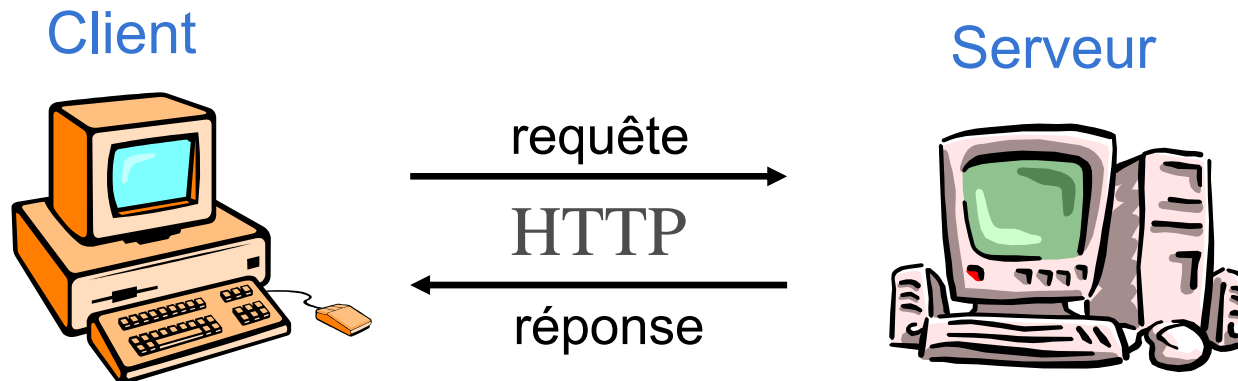
HTTP

◆ HyperText Transfert Protocol

- Version actuelle : HTTP 1.1
Spécification décrite dans le RFC 2616.
- Protocole client-serveur :
 - » le navigateur demande une page HTML (requête)
 - » le serveur répond à cette demande (réponse).
- Ce protocole est sans état (en anglais, *stateless*) :
le serveur traite les demandes indépendamment.

Dialogue entre client / serveur web

1. Le client demande un document via une URL
 - établissement de la connexion avec le serveur spécifié dans l'URL
 - demande du document sur le serveur
2. Le serveur traite la demande
 - il envoie le document ou un message d'erreur au client
 - puis ferme la connexion



◆ Remarque

- Chaque élément d'un document composé provoque ce type de dialogue (image dans une page, ...).

Exemple de dialogue HTTP

```
$> telnet www.enseirb.fr 80
Trying 147.210.18.146...
Connected to grandgousier.enseirb.fr.
Escape character is '^]'.
GET /~reveille/
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml">
<head>

...

Connection closed by foreign host.
```


◆ Étapes

- connexion TCP sur le port 80
- requête du document (GET)
- fermeture de la connexion

HTTP... et ensuite ?

◆ Traitement du document

- Réalisé par le client (navigateur)
- HTML
 - » mise en forme du texte
 - » insertion d'images, de sons, d'animations, etc.
 - » spécification de liens hypertextes :
 - ◆ à travers les URL
 - ◆ un navigateur traite spécifiquement les liens : déclenchement de l'accès au document lorsque le lien est activé
 - » documents non supportés par le client :
 - ◆ appel de modules de visualisation externes



U niform R essource L ocator

Introduction

◆ Définition

- Schéma de nommage des ressources qui permet d'adresser de façon précise toute ressource accessible sur l'Internet

◆ Pourquoi les URL ?

- Le web = système hypertexte sur Internet.
- Nécessité de pouvoir référencer des ressources :
 - » indépendamment du type d'accès : serveurs Web, FTP, WAIS, ...
 - » indépendamment de la localisation sur le réseau (machine physique ou logique)
 - » de manière non ambiguë

◆ Comment ?

- Extension de la notion de désignation de fichier

URL

◆ Format d'une URL

```
<protocole>://[user[:password]@]<machine>[:port][/ <path>[#label|?liste paramêtres]]
```

- <protocole> : méthode d'accès au document
- <machine> : adresse de la machine
- <port> : numéro de port sur la machine
- <path> : chemin d'accès au document sur la machine

◆ Exemples d'URL

- <http://www.enseirb.fr/>
- <http://www.enseirb.fr:80/formations/>

Exemples de protocoles

- ◆ **http** : ressource accessible sur un serveur Web.
- ◆ **file** : fichier local sur la machine.
- ◆ **ftp** : ressource accessible sur un serveur FTP.
- ◆ **telnet** : ressource accessible via une session TELNET
- ◆ **gopher** : ressource accessible sur un serveur GOPHER.
- ◆ **wais** : ressource accessible via la version WAIS du protocole Z39.50.
- ◆ **news** : ressource accessible sur serveur de news (protocole NNTP).

Exemples plus variés d'URLs

- `file://localhost/users/nauer/essai.html`
- `file:/usr/local/lib/mailcap`
- `ftp://ftp.inria.fr/`
- `gopher://gopher.univ-rennes1.fr/`
- `http://www.enseirb.fr/`
- `http://www.paris.org:80/Monuments/Eiffel/info.html`
- `telnet://info.cern.ch`
- `telnet://www@info.cern.ch`
- `wais://zenon.inria.fr:210/directory-zenon-inria-fr`
- `news:fr.comp.infosystemes`
- ...

Caractères spéciaux



- % : caractère de contrôle permettant de coder certains caractères
%20 = *espace*, %26 = *&*, %2B = *fin de ligne*, etc.
- # suivi d'une chaîne de caractères :
 - » permet de référencer un *fragment* de document
<http://web.urec.fr/docs/www/web.1.html#chap2>
- ? suivi d'une chaîne de caractères :
 - » liste de paramètres pour l'exécution d'un programme
<http://www.enseirb.fr/ens/web/exemples/foo?nom=laurent>


Désignation d'un document

- URL absolue

```
protocole://serveur/répertoire/fichier  
http://www.enseirb.fr/enseignement/telecom/index.htm  
https://www.enseirb.fr/horde/  
mailto:reveillere@enseirb.fr  
ftp://ftp.rpmfind.net/linux/  
...
```

- URL relative (résolution relative au document courant)

```
Document courant = http://www.enseirb.fr/enseignement/  
telecom  http://www.enseirb.fr/enseignement/telecom  
../extranet.html  http://www.enseirb.fr/extranet
```



Proxy & Cache

Proxy

◆ Définition

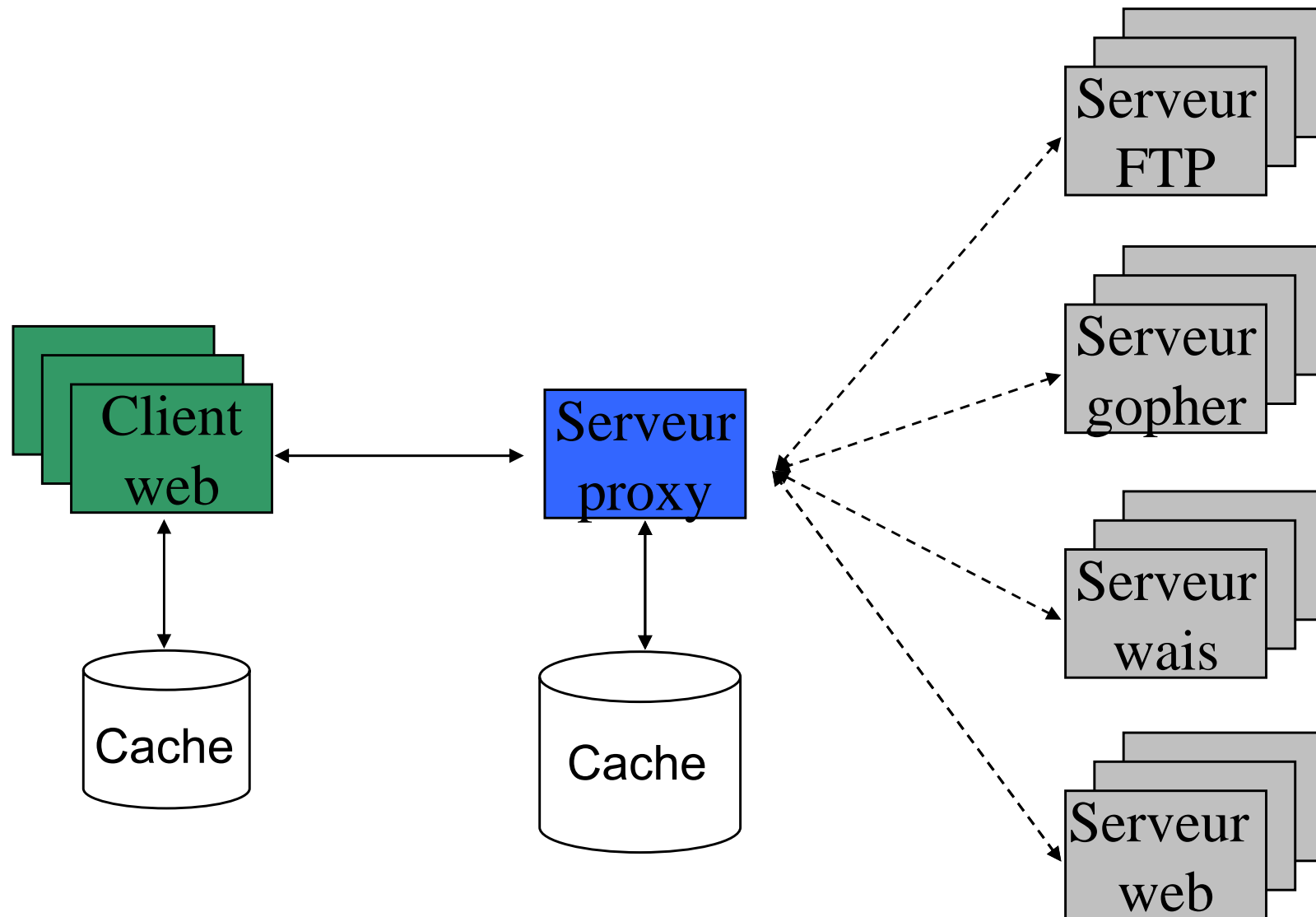
- Intermédiaire entre clients et serveurs web



◆ Objectifs

- Sécurité : mise en place de gardes-barrière (*fire-wall*)
- Communication : permet d'accéder à des serveurs dont les clients web ne supportent pas le protocole (WAIS par exemple)
- Réduction du trafic réseau : couplé à un cache, il permet d'optimiser l'utilisation du réseau et de diminuer les temps de réponse des requêtes

Cache et proxy cache





Accéder au web

Outils de navigation

- ◆ Client web : Netscape, Internet Explorer, Opéra, Mozilla, ...
- ◆ Interface d'accès au web permettant
 - d'accéder aux documents du web
 - de parcourir les liens hypertextes
 - des gestions spécifiques (stockage des URL intéressantes, ...)
- ◆ Les clients web intègrent l'accès aux services d'informations autres que les serveurs Web :
 - gopher, ftp, wais, news, ...

Clients Web

- ◆ De très nombreux clients différents et ... des comportements également différents
- ◆ 3 grandes familles de moteurs de rendu HTML
 - Internet Explorer 83.29 %
 - Mozilla Firefox 13.53 %
 - Apple Safari 01.41 %

Taux de pénétration Monde en Janvier 2007
- ◆ Différences
 - Interprétations variables du même code HTML :
 - » Positionnement des objets, capacité différente de traitement d'un code HTML erroné, etc.
 - Code HTML spécifique à un navigateur (certaines balises sont spécifiques à certains navigateurs)
 - Intégration de fonctionnalités
 - » Documents Microsoft, VBscript intégrés dans IE

➔ Contraintes pour le développement de sites web