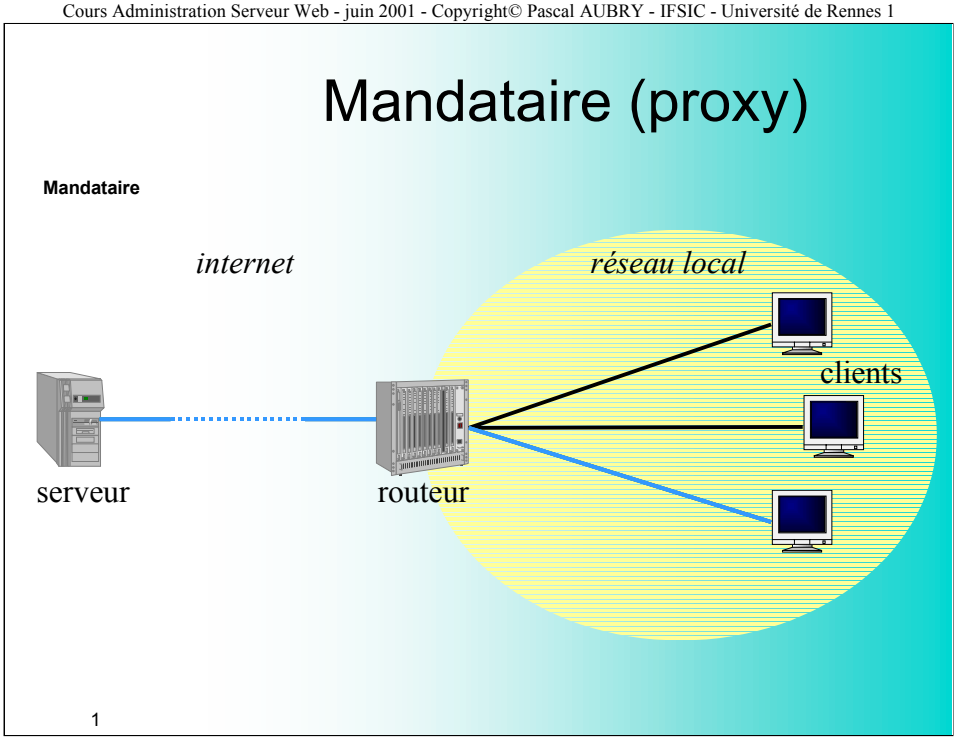
Mandataires, caches et filtres

Pascal AUBRY
IFSIC - Université de Rennes 1
Pascal.Aubry@univ-rennes1.fr

Plan :

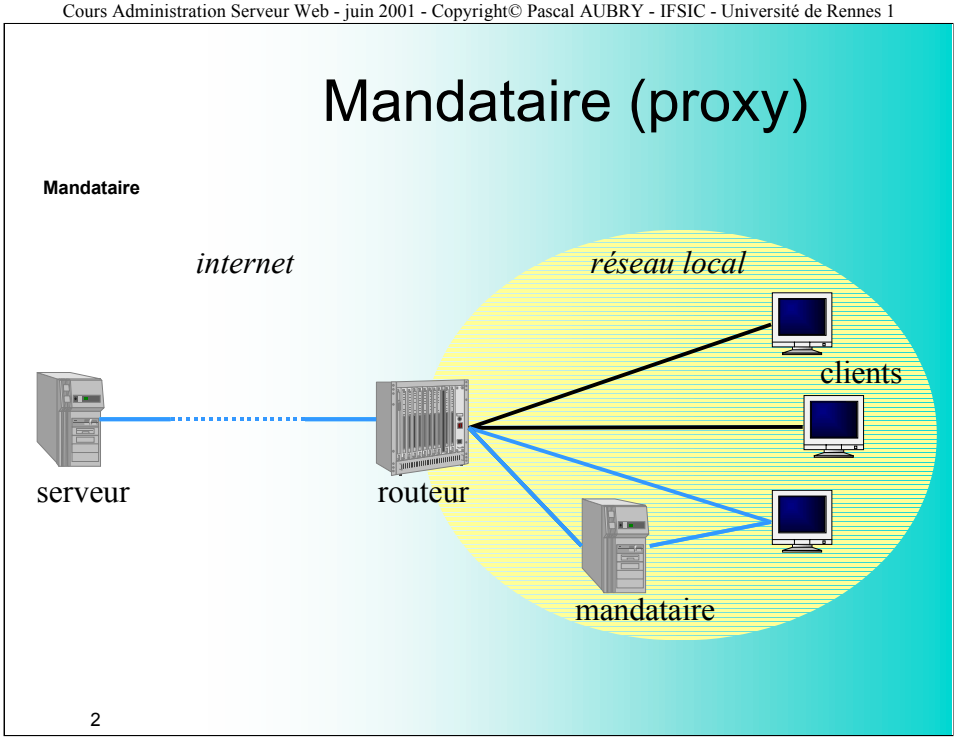
- mandataires
- caches
- filtrage
- serveur de proxy
- exemple de mise en œuvre



Sans mandataire, les machines clientes d'un réseau accèdent directement aux serveurs de l'internet, à travers un routeur.

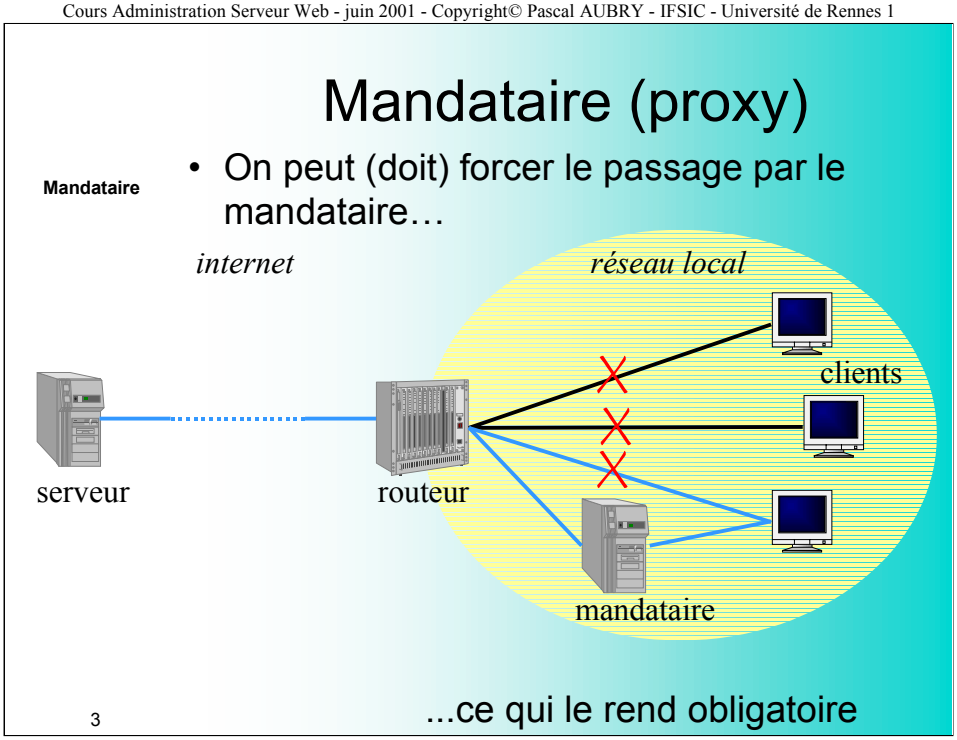
Les services accédés sont nombreux : http, ftp, wais, gopher, …

Dans la pratique, http et ftp sont les deux les plus utilisés (de lojn).



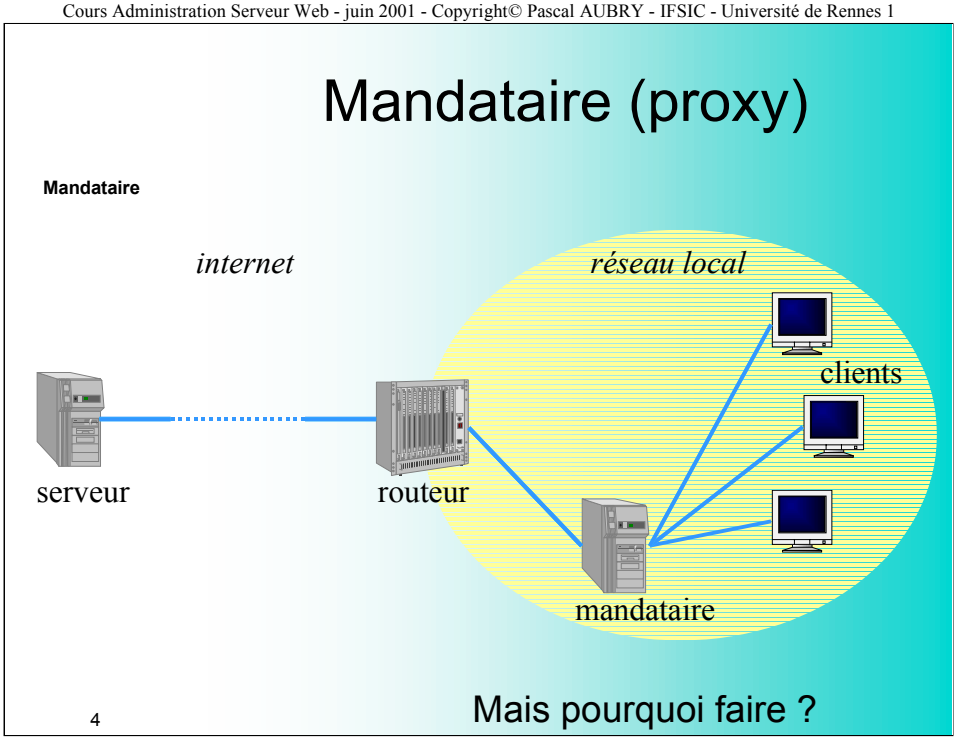
Un mandataire est un serveur qui fait l'interface entre les clients d'un réseau et les serveurs web.

Pour dialoguer avec les serveurs (accéder aux services qu'ils proposent, les clients s'adressent au mandataire, qui relaie les requêtes vers les serveurs.



La mise en place d'un mandataire répondant à un besoin (cf plus loin), on force en général leur utilisation pour qu'ils puissent remplir leur rôle. Pour cela, on configure le routeur pour l'empêcher de laisser passer les requêtes des clients ; seul le mandataire est autorisé à sortir du réseau.

Un mandataire peut ainsi être utilisé pour passer un firewall.



Un mandataire, pour quoi faire ?

Pour cacher les requêtes des clients

- •diminution du traffic
- •accès à des serveurs indisponibles

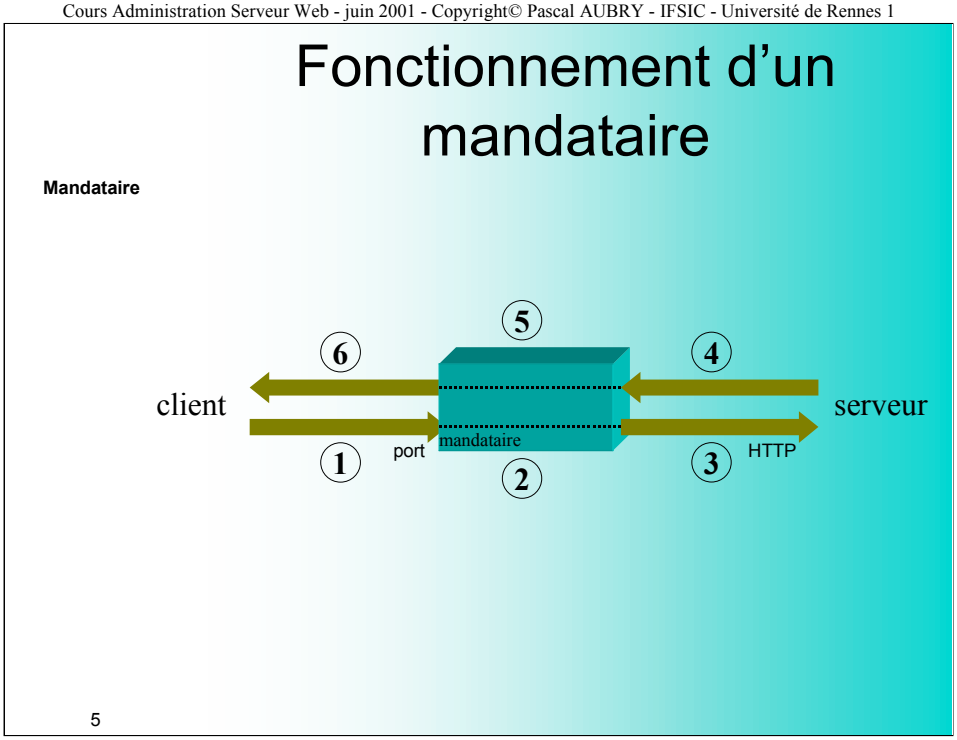
Pour surveiller le trafic

- •statistiques d'utilisation des ressources

Pour filtrer les requêtes

- •sur différents critères

Pour passer un firewall



Le client :

- se connecte sur un port du mandataire (1)

Le mandataire :

- filtre la requête (2) *
- redirige la requête (2) *
- effectue la requête à la place du client (3 & 4)
- mémorise le résultat de la recherche (5) *
- renvoie le résultat de la requête au client (6)

Les actions marquées d'un astérisque sont optionnelles (selon le mandataire).

Le mandataire de référence : Squid

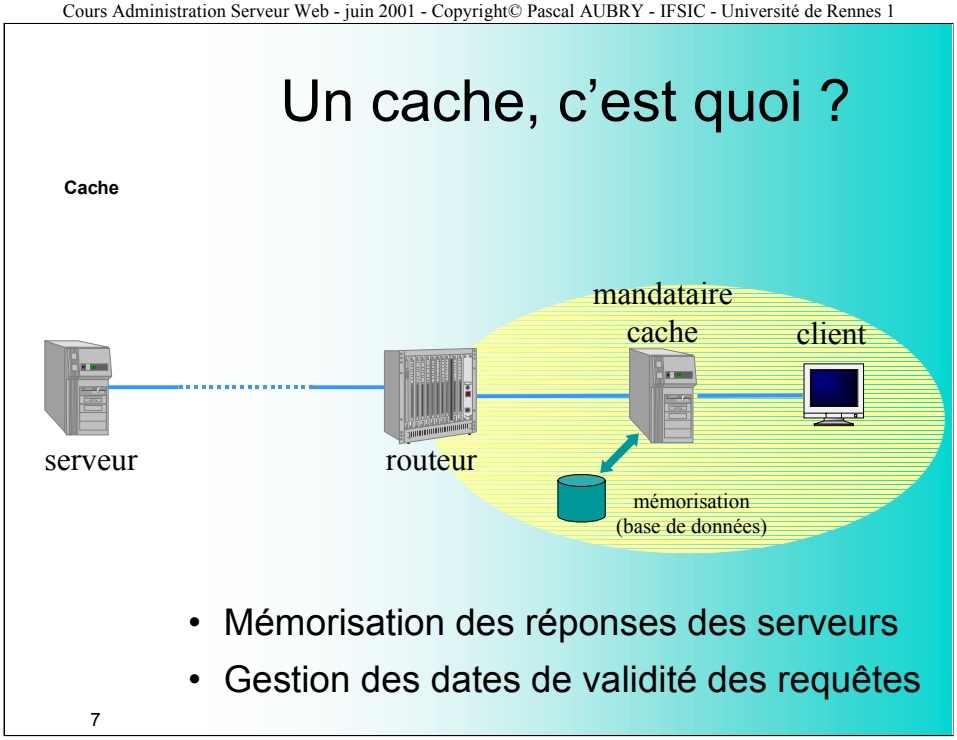
Squid

- Caractéristiques
  - Multi-plateformes (Unix, Windows, …)
  - Libre (GPL)
  - Paramétrable
  - Performant (cache DNS, garbage collector)
  - Évolutif
- Fonctionnalités
  - mandataire
  - cachant
  - filtrant (squidGuard)
- http://www.squid-cache.org

6

D ’autres mandataires existent, mais Squid est de loin le plus employé.

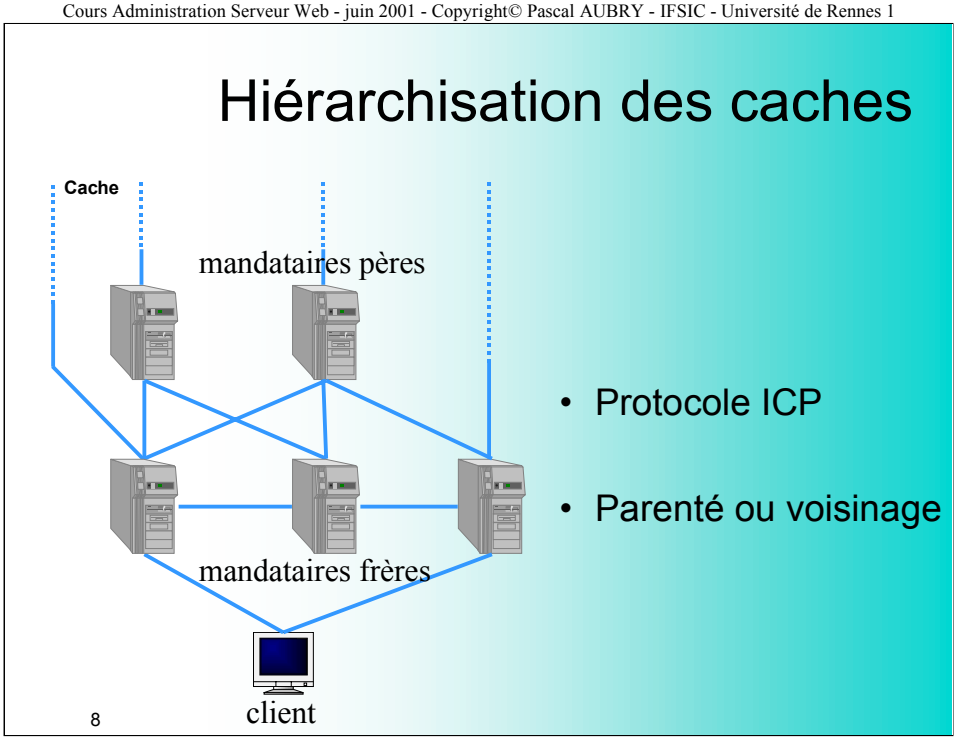
En une phrase : il y a moins bien, mais c’est plus cher.



Le cache est une fonctionnalité des mandataires (c'est parce que cette fonctionnalité est offerte par Squid, le mandataire le plus utilisé, que l'on confond très souvent mandataire et cache).

Cacher une requête, c'est en mémoriser le résultat (dans une base de données) avant de le renvoyer au client. De cette manière, le mandataire peut servir les autres clients (ou le même) qui font la même requête ultérieurement plus rapidement (en renvoyant le résultat stocké dans la base de données au lieu de le redemander au serveur).

Note : dans la pratique, le mandataire interroge tout de même le serveur pour savoir si le document n'a pas changé depuis la dernière requête, mais sans lui demander le contenu du document (ce qui est beaucoup plus rapide).



Pour plus d'efficacité, les caches peuvent être organisés de manière hiérarchique, en utilisant le protocole ICP (Internet Cache Protocol).

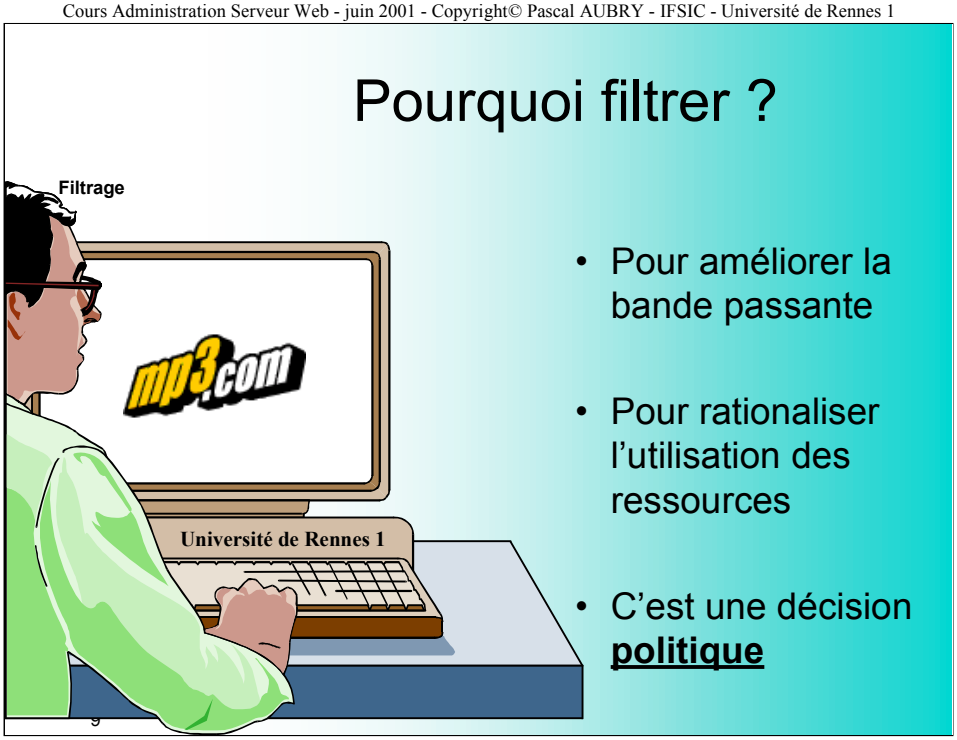
Avantages des caches :

- clients servis plus rapidement
- réduction du trafic sortant
- accès à des serveurs indisponibles
- masquage des adresses IP

Inconvénients des caches :

- masquage des adresses IP
- de plus en plus inutiles (pages dynamiques)

Les hiérarchies de cache sont ainsi de moins en moins utilisées (arrêt du serveur de cache national de Jussieu il y a plus d'un an).



En filtrant certains sites, on diminue le trafic réseau spécifiquement lié à la navigation sur l'internet. On libère ainsi de la bande passante pour les autres services.

La rationalisation de l'utilisation des ressources consiste essentiellement en la libération de machines occupées par des « surfeurs », qui constatent assez rapidement, avec le système mis en place, que l 'IFSIC n'est pas un cyber-café. Ces machines peuvent alors être utilisées par d'autres utilisateurs, pour des utilisations plus en rapport avec les enseignements dispensés à l'IFSIC.

Note : cette présentation ne décrit que les aspects techniques des mandataires. Les aspects philosophiques et éthiques sont laissés à des spécialistes. Dans tous les cas, même si la mise en place d'un filtrage est effectuée par les administrateurs du réseau, il s'agit bien là d'une décision politique.

Critères de filtrage

Filtrage

- L'origine de la requête
  - selon le client
- La destination de la requête
  - selon le serveur
- Le type de requête
  - selon le port, la méthode
- Le contenu
  - analyse de mots-clés
- Le temps
  - selon l'heure (créneaux différenciés)

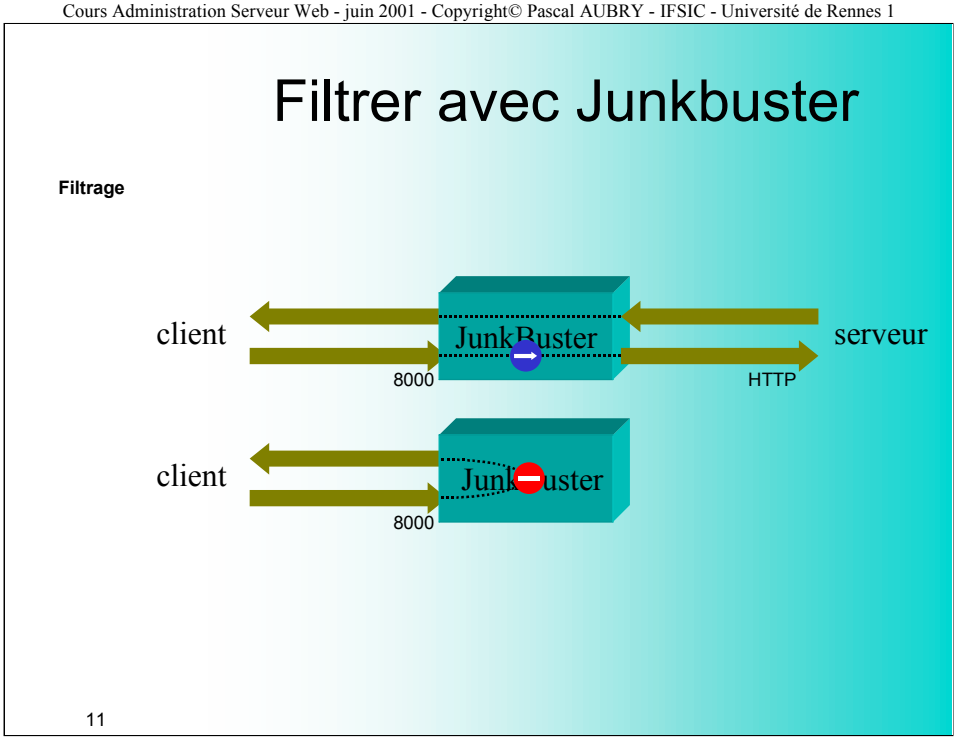
10

Le filtrage selon l'origine permet de distinguer différentes classes de machines clientes au sein d'un réseau.

Le filtrage selon la destination permet d'arrêter certaines requêtes sur des sites jugés impropres à la navigation.

Le filtrage selon le type de requête permet d'arrêter certaines requêtes correspondant à des utilisations indésirables. On arrêtera ainsi les requêtes de commerce électronique en bloquant le port 443 et les mises à jour de sites distants en interdisant la méthode PUT.

Le filtrage du contenu permet d'arrêter des requêtes dont le résultat n'est pas jugé admissible avec les règles en vigueur sur le réseau (détection de mots-clés par exemple).



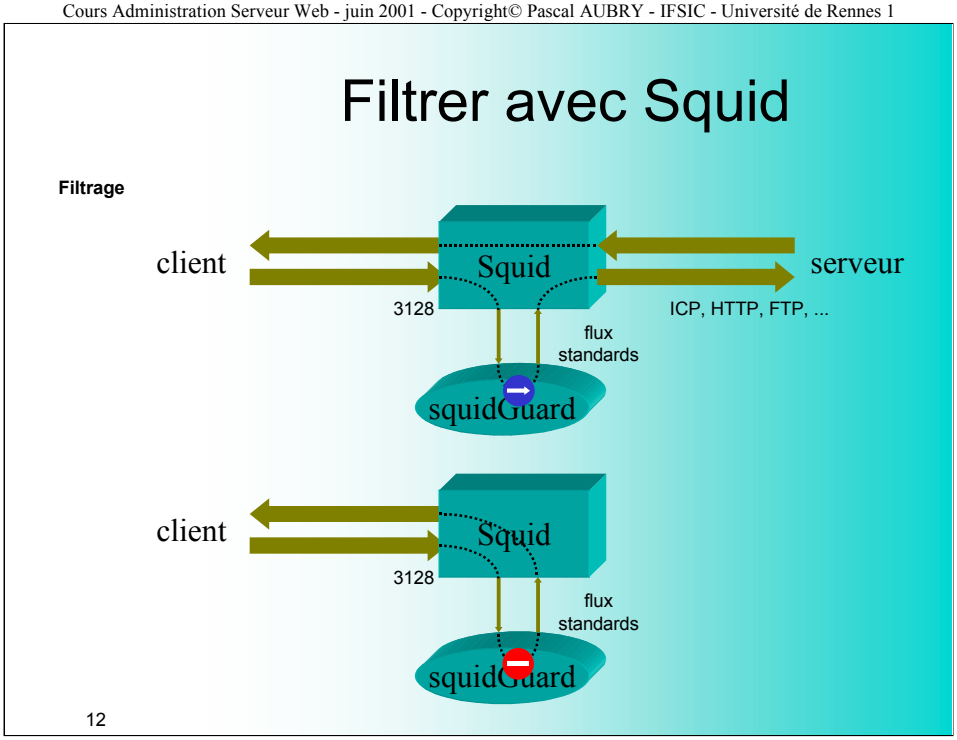
JunkBuster est un mandataire conçu pour le filtrage des publicités (remplacées par une image transparente ou un icône de JunkBuster) et des cookies (non transmission).

Le filtrage est basé sur des expressions régulières.

JunkBuster ne sait filtrer que le protocole HTTP, ce qui implique que les requêtes FTP (entre autres) ne doivent pas lui être adressées (cf configuration du serveur de proxy de l 'IFSIC).

JunkBuster est un outil libre.

Note : le blocage pur et simple des cookies peut être gênant car ils sont nécessaires pour certains sites.



Squid ne filtre pas les requêtes de ses clients de manière native (fonctionnalité de cache seulement). Le filtrage est réalisé par un redirecteur, programme externe qui est interrogé par Squid à chaque requête (un démon Squid n'utilise qu'un seul redirecteur).

Le redirecteur le plus souvent employé est SquidGuard, car le plus abouti et le plus performant.

Le filtrage de Squid Guard est basé des expressions régulières.

Les pages arrêtées par le filtrage sont réécrites, c'est-à-dire que l'on peut rediriger vers une URL où l'on explique le filtrage aux utilisateurs.

Squid et SquidGuard sont des outils libres.

Analyse du filtrage

Filtrage

- Filtres binaires
  - on passe / on ne passe pas
- Filtres basés sur des « listes noires »
  - impossibilité de les maintenir
  - confiance limitée
- On souhaite un filtrage conditionnel
  - responsabilisation des utilisateurs
  - dégagement de responsabilité

13

Le filtrage tel que celui proposé par les outils disponibles aujourd'hui est binaire : on passe ou on ne passe pas.

Or les filtres sont basés sur des listes d'URLs, sur lesquelles on ne peut pas toujours se fier : elle sont parfois trop restrictives. On ne peut alors avoir en elles qu'une confiance limitée et il est hors de question de maintenir ce genre de listes (par manque de temps).

L'idéal serait évidemment un filtrage conditionnel, où un utilisateur qui voudrait consulter une page indésirable se verrait signifier l'appartenance de la page voulue à une liste interdite, mais pourrait quand même l'accéder en confirmant son désir de consulter la page.

Cette politique de filtrage conditionnel aurait pour effet de responsabiliser les utilisateurs. Malheureusement, la mise en œuvre de cette solution est très délicate (mise en œuvre à l'IFSIC pendant quelques mois mais jugée non suffisamment stable).

Serveur de proxy

Serveur de proxy

- Configuration automatique des navigateurs
- Accès via HTTP (URL)
  - ex : http://www-proxy/
- Document « application/x-ns-proxy-config »
- Code JavaScript
- Fonction FindProxyForURL()

14

Exemple de configuration Apache :

```
AddType application/x-ns-proxy-autoconfig .proxy
<VirtualHost 148.60.4.30>
          ServerName www-proxy.ifsic.univ-rennes1.fr
          ServerAlias www-proxy
          ErrorLog /dev/null
          TransferLog /dev/null
          DocumentRoot /www/proxy
          DirectoryIndex ifsic.proxy
          <Directory /www/proxy>
                    Options Indexes
                    Order allow,deny
                    Allow from 148.60.
          </Directory>
</VirtualHost>
```

Exemple de configuration de proxy (ifsic.proxy) :

```
function FindProxyForURL(host,url)
{
          if ( IsPlainHostName(host)
            || DnsDomainIs(host,".ifsic.univ-rennes1.fr") )
                    return "DIRECT" ;
          else
                    return "cache.ifsic.univ-rennes1.fr:3128" ;
}
```

Pour information, voici une configuration de proxy qui permet d'équilibrer la charge entre deux mandataires.

```
<VirtualHost 148.60.4.30>
          ServerName www-proxy.ifsic.univ-rennes1.fr
          ServerAlias www-proxy
          ErrorLog /dev/null
          TransferLog /dev/null
          DocumentRoot /www/proxy
          DirectoryIndex index.php
          <Directory /www/proxy>
                    Options Indexes
                    Order allow,deny
                    Allow from 148.60.
          </Directory>
</VirtualHost>
```

Configuration de proxy (/www/proxy/index.php)

```
<?
   strtok($GLOBALS["REMOTE_ADDR"],".") ;
   strtok(".") ;
   strtok(".") ;
   $last_ip_byte = strtok(".") ;
   header("Content-Type: application/x-ns-proxy- autoconfig");
   echo "function FindProxyForURL(url,host) { return " ;
   echo ( ( $last_ip_byte % 2 )
          ? "\"www-cache1:3128\""
          : "\"www-cache2:3128\"" ) ;
   echo "}" ;
?>
```

Les clients impairs se branchent sur www-cache1, les client pairs sur www-cache2.

Accès à l'internet depuis l'IFSIC

L'IFSIC

- Une problématique complexe
  - on doit limiter l'accès à l'internet
  - on ne peut pas tout filtrer
  - on ne veut pas restreindre le personnel
  - on doit préserver un accès complet
- La solution mise en oeuvre
  - un découpage en quatre zones
  - une bonne lisibilité
  - un bon compromis

16

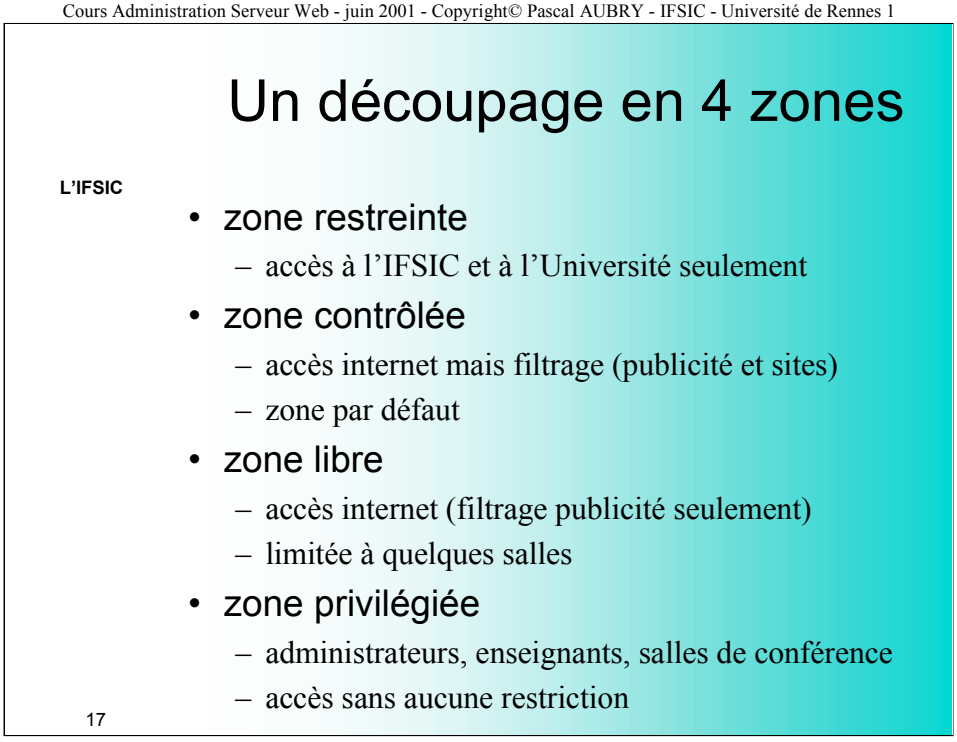
Problématique :

•On doit limiter l'accès à l'internet pour diminuer le trafic réseau dû à la navigation et rendre des machines plus disponibles.

•Il n'existe pas de solution toute faite qui permette d'imposer un filtrage depuis tous les clients de l'IFSIC ; il faudrait dans ce cas être sûr des règles de filtrage, ce qui est impossible (cf analyse du filtrage).

•Le personnel doit (veut ;-) être traité de manière distincte.

•On trouvera toujours des cas précis où les étudiants peuvent avoir besoin de consulter un site, dans le cadre de leurs enseignements.



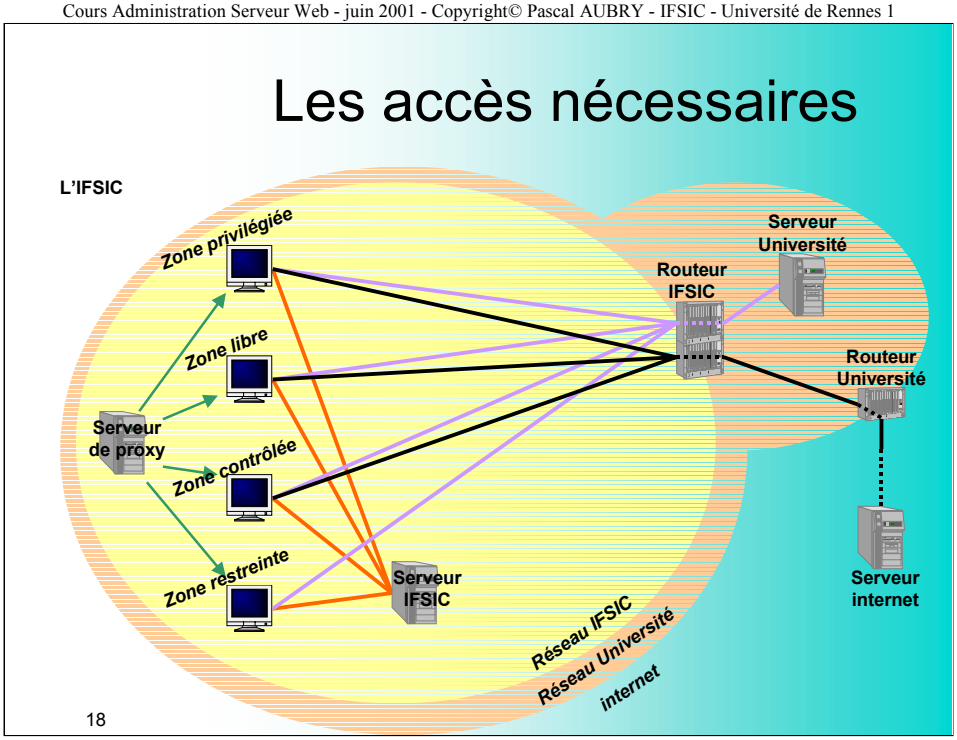
La zone restreinte correspond à deux salles de terminaux X servies par un serveur MetaFrame qui ne supporterait pas 20 navigateurs web en parallèle.

La zone contrôlée est la zone par défaut des machines de l'IFSIC. Un filtrage contraignant est appliqué aux accès à l'internet depuis les clients de cette zone.

La zone libre est constituée de 4 salles, d'où les étudiants peuvent accéder à toutes les ressources de l'internet, sans exception.

La zone privilégiée correspond aux machines des personnels et celles des salles de conférence.

Ce découpage a été proposé par les administrateurs de l'IFSIC à une commission réunissant enseignants, direction, administrateurs et étudiants.

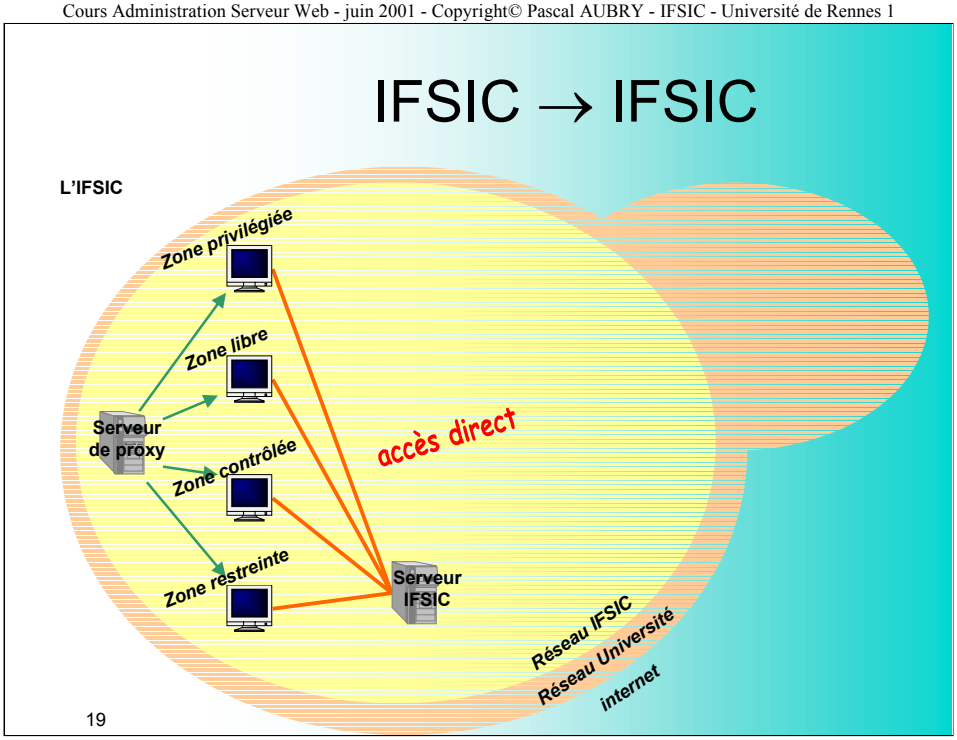


Les machines du réseau de l'IFSIC doivent pouvoir envoyer des requêtes à trois types de serveurs :

- des serveurs du réseau de l'IFSIC
- des serveurs du réseau de l'Université
- des serveurs externes

Pour que les mandataires mis en place pour réduire le trafic sortant jouent leur rôle, il faut que les clients les utilisent. Pour cela, on interdit (au niveau du routeur) toute requête provenant des machines du réseau de l'IFSIC de sortir (par la suite, on autorisera les mandataires à passer le routeur).

Tous les navigateurs du réseau sont configurés automatiquement par un serveur de proxy.

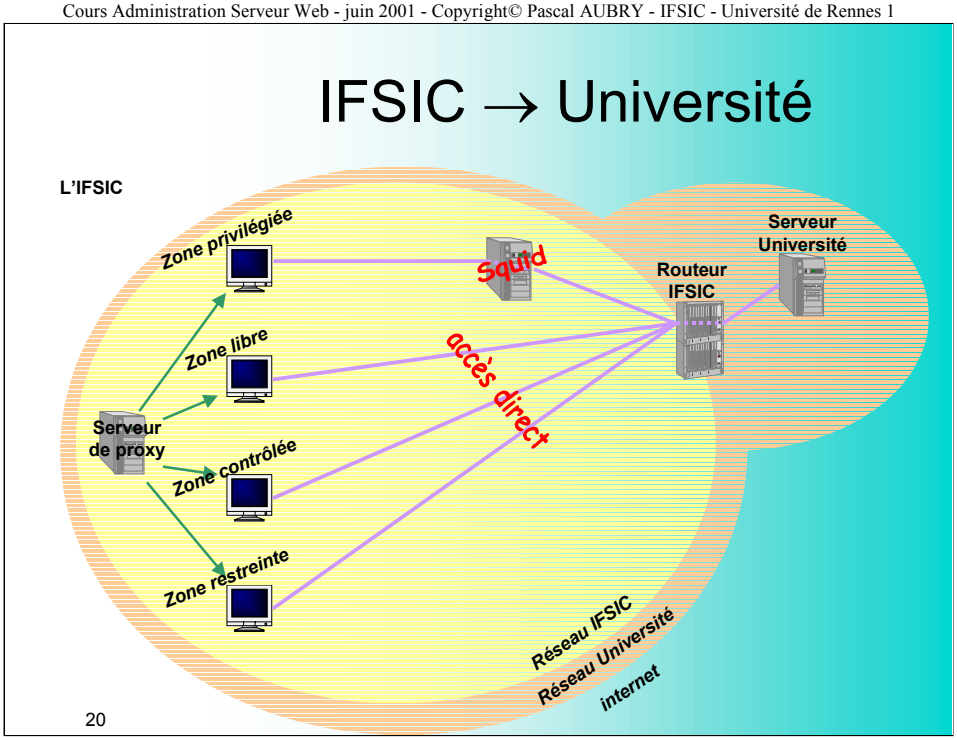


Les requêtes vers des serveurs du réseau de l'FSIC sont dirigées directement vers les serveurs cibles. Il est en effet inutile de cacher ce type de requête :

- aucune diminution du trafic
- gain de temps insignifiant

La configuration du serveur de proxy est la suivante :

```
if ( IsPlainHostName(host)
   || DnsDomainIs(host,".ifsic.univ-rennes1.fr") )
          return "DIRECT" ;
else
          ...
```



Toutes les machines du réseau de l'IFSIC doivent pouvoir accéder aux serveurs de l'Université. Pour cela, on les autorise (toutes) à traverser le routeur (seulement vers les machines de l'Université, pas vers celles extérieures au réseau de l'Université).

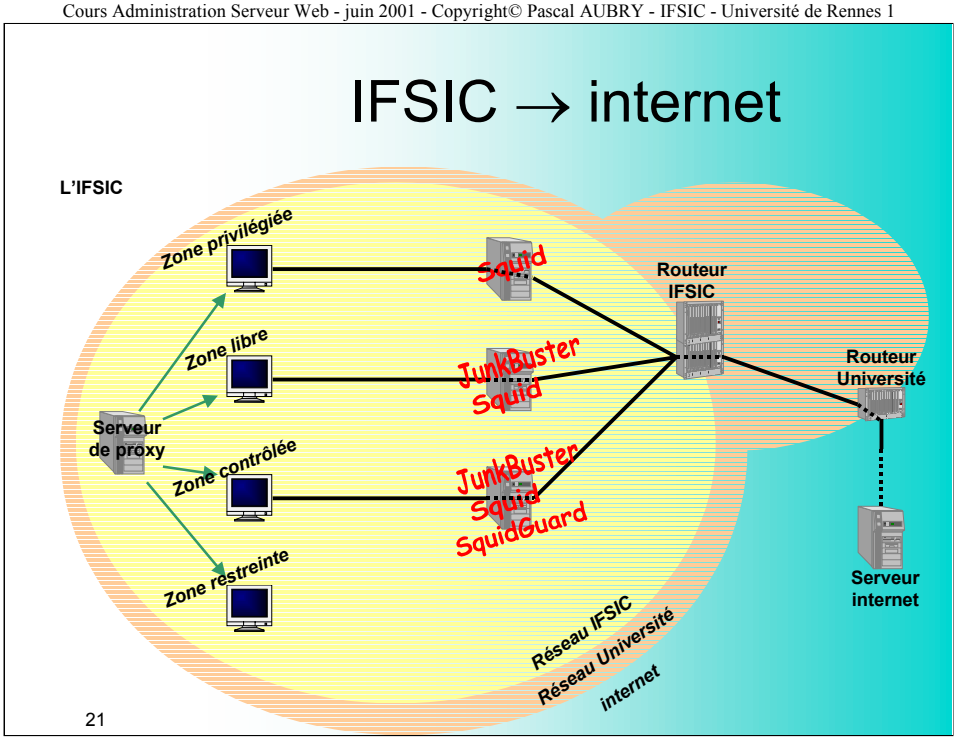
Les machines de la zone privilégiée doivent pouvoir accéder aux serveurs de l'Université avec des droits particuliers (accès à l'intranet de l'Université pour les personnels de l'IFSIC). Pour cela, on met en place un mandataire autorisé (par les serveurs de l'Université) à accéder à l'intranet de l'Université. Afin de limiter ce droit, l'accès au mandataire est interdit aux machines hors de la zone privilégiée.

Configuration de proxy pour la zone privilégiée :

```
if ( DnsDomainIs(host,".univ-rennes1.fr") )
          return "cache-admin.ifsic.univ-rennes1.fr:3128" ;
else
          …
```

Configuration de proxy pour les autres zones :

```
if ( DnsDomainIs(host,".univ-rennes1.fr") )
          return  "DIRECT" ;
else
          …
```



Afin que les machines de la zone privilégiée puissent accéder à l'internet, le routeur laisse passer les requêtes du mandataire de cette zone vers l'internet. Les zones contrôlée et libre ayant des politiques de filtrage différentes, on met en place deux mandataires supplémentaires (un pour chaque zone).

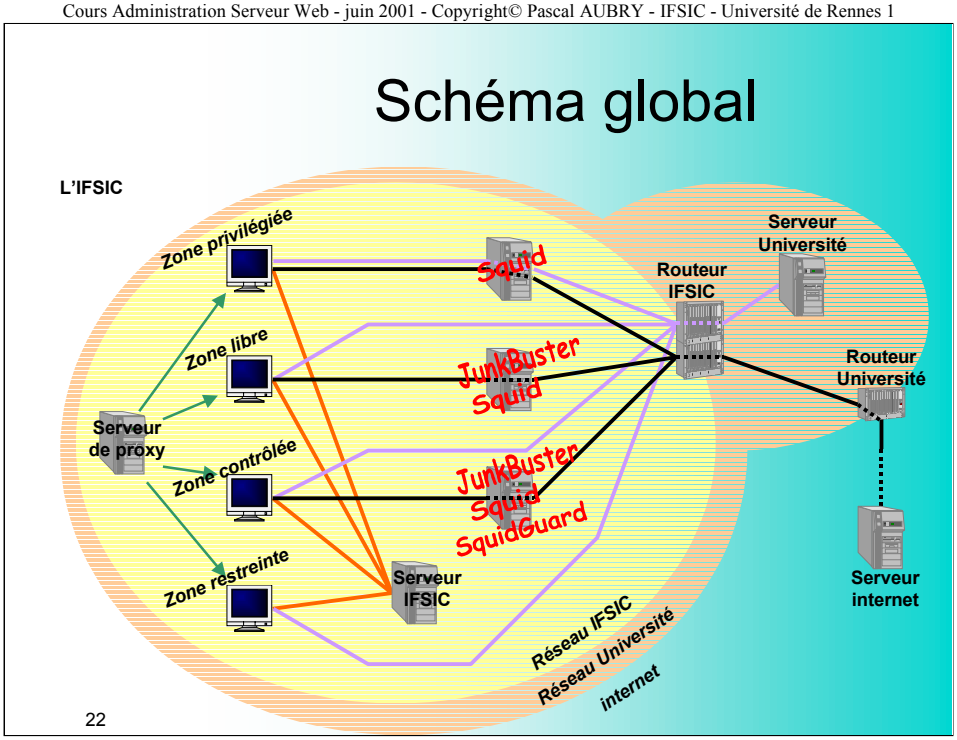
Le mandataire de la zone libre est muni de JunkBuster (pour les publicités) et Squid (pour cacher). Afin de limiter le droit de consultation sans restriction à la zone libre, on limite l'accès au mandataire de la zone libre (réservé aux machines de la zone libre).

Le mandataire de la zone contrôlée est muni de JunkBuster, Suid et SquidGuard (pour les sites interdits). L'accès au mandataire de la zone contrôlée est interdit aux machines de la zone restreinte.

De cette manière, la zone restreinte n'est pas autorisée à accéder aux serveurs de l'internet (le routeur ne laisse passer que les mandataires des trois autres zones, mandataires dont l'accès leur est interdit).

Note : il faut pour JunkBuster distinguer les accès accès HTTP (les seuls que peut mandater Junkbuster des autres, cf filtrer avec JunkBuster) :

```
if ( shExpMatch(url,"http://") )        return "cache:8000" ;
else                                    return "cache:3128" ;
```



Le serveur de proxy doit délivrer à chaque client une configuration qui lui dit, pour chaque destination, s'il faut utiliser un mandataire, et si oui lequel.

Deux solutions sont alors possibles :

- délivrer à tous les clients la même configuration de proxy (les clients étant différenciés au moment de la consultation des URLs). Cela implique un fichier de configuration plus long (mais un seul) car la fonction FindProxyForURL doit prendre en compte tous les cas, en différenciant les machines selon leur numéro IP (qui détermine leur zone).
- délivrer aux client un proxy correspondant à la zone dans laquelle il se trouvent (différenciation au moment du délivrement de la configuration de proxy par le serveur de proxy). Cela implique une configuration de proxy dynamique (recalculée à chaque requête en fonction du numéro IP du client).

La première solution a été choisie dans une optique de lisibilité et de transparence.