Squid

Olivier Aubert

Liens

- `http://www.squid-cache.org/`
- `http://squid.visolve.com/`

Principe

- Squid : cache web logiciel
- Fonctionne sur un OS classique (Linux, BSD, Solaris, Windows, ...)
- Logiciel libre, développement actif
- Protocoles cachés : http, ftp, gopher
- SSL, contrôle d'accès, coopération

Exigences matérielles

- Disques
  - Taille de l'espace de stockage
  - Vitesse d'accès aléatoire aux disques
  - Taux de transfert soutenu
- Quantité de mémoire disponible
- Puissance CPU

Taille de l'espace de stockage

- Détermination du trafic quotidien : mesure ou extrapolation (par la bande passante du site)
- Expériences :
  `http://wwwcache.ja.net/servers/squids.html`
  (Squid Sizing for Intel Platforms)

Configuration des clients

- Manuelle : préférences de l'application ou variable d'environnement `http_proxy`.
- Automatique : fichier javascript `proxy.pac`, WPAD
- Cache transparent : pas de configuration des clients à effectuer.

Autoconfiguration

```
function FindProxyForURL(url, host) {
 if (host=="127.0.0.1"      ||
     isPlainHostName(host) ||
     dnsDomainIs(host,".univ-lyon1.fr")) {
        return("DIRECT");
     } else {
        return ("PROXY 134.214.88.23:8080; "
               + "PROXY 134.214.103.196:8080;"
               + "DIRECT; ");
} }
```


Configuration de Squid

- Réseau
- Stockage
- Programmes externes
- Réglages
- Contrôle d'accès
- Delay pools

Configuration réseau

- `http_port` : port utilisé par le cache. Généralement 8080.
- `icp_port` : par défaut, 3130.
- `cache_peer` : proxys parents ou cousins. Syntaxe : `cache_peerhostnametypehttp_porticp_portoptions`
- `request_size` : taille maximale des requêtes.
- `hierarchy_stoplist` : permet de spécifier quelles URL ne doivent pas passer par la hiérarchie, mais être traitées directement. Typiquement `cgi-bin`, `jsp`.

Configuration de stockage

- `cache_mem` : taille du cache mémoire (défaut : 8 Mo), la plus grande possible. Comptez 1/3 de la mémoire consacrée à Squid. Balance entre `cache_mem_low` (75%) et `cache_mem_high` (90%).
- `cache_dir` : répertoire de données du cache. Syntaxe : `cache_dir Type Maxobjsize Directory-Name Mbytes Leve`
- `cache_replacement_policy` : politique de remplacement appliquée au cache.
- `maximum_object_size` : taille maximale de l'objet qui sera sauvegardé sur le disque.

Configuration programmes externes

- `dns_children` : nombre de requêtes DNS maximum.
- `authenticate_program` : programme d'authentification
- `authenticate_ttl` : durée de validité de l'authentification
- `redirect_program` : programme de redirection

Réglages fins

- `refresh_pattern` : Réglage des expirations suivant le nom des fichiers. Syntaxe :
  `refresh_pattern [-i] regex min percent max [options]`
  Exemple :
  `refresh_pattern -i \.jpg\$ 1440 700\% 262800.`
  Voir *squij*.
- `reference_age` : définit l'âge maximum d'un document pour l'expiration LRU.
- `quick_abort_(min|max|pct)` : comportement à adopter sur les requêtes stoppées par l'utilisateur.

Contrôle d'accès

- `acl aclname acltype string1...` : définition de paramètres de contrôle d'accès. `acltype` peut être
  - `src` : adresse IP du client
  - `srcdomain` : domaine DNS du client
  - `dst` : adresse IP du serveur
  - `dstdomain` : domaine DNS du serveur
  - `time` : définition d'une date
  - `url_regexp` : expression régulière appliquée à l'URL
  - `proxy_auth` : authentification des utilisateurs

Contrôle d'accès

- `http_access allow|deny` : contrôle l'accès via HTTP
- `icp_access` : contrôle l'accès via ICP
- `cache_peer_access` : redirection des requêtes correspondant à l'acl vers un cache particulier.
- `always_direct` : connexion directe aux serveurs correspondant à une ACL particulière

Exemple

Restriction de l'accès de la machine 192.168.2.0 aux heures de bureau.

```
acl ip_acl src 192.168.2.0/24
acl time_acl time M T W H F 9:00-17:00
http_access allow ip_acl time_acl
http_access deny all
```


Configuration admin

- `cache_mgr` : adresse mail de l'administrateur du cache
- `cache_effective_user` : utilisateur possédant le processus.
- `cachmgr_passwd` : mot de passe contrôlant l'accès via l'interface web

Delay pools

- Moyen de contrôler l'utilisation de la bande passante en fonction de paramètres fixés par acl.
- 3 classes de *delay pools* selon que les clients sont traités séparément ou en groupes.
- Exemple: limitation du trafic de Squid

```
acl all src 0.0.0.0/0.0.0.0
delay_pools 1
delay_class 1 1
delay_access 1 allow all
delay_parameters 1 64000/64000
 # 512 kbits == 64 kbytes per second
```


Fichiers de log

- `access.log` : liste des requêtes qui ont transité par Squid
- `cache.log` : état du serveur lors de son démarrage
- `store.log` : liste des objets stockés dans le cache

Fichier access.log

- Codes HTTP : `TCP_HIT`, `TCP_MISS`, `TCP_REFRESH_(HIT|MISS)`, `TCP_CLIENT_REFRESH`, ...
- Codes ICP : `UDP_HIT`, `UDP_MISS`, ...
- Codes hiérarchiques : `DIRECT`, `FIRST_UP_PARENT`, `SIBLING_HIT`, ...

```
1019767444.637   1007 127.0.0.1 TCP_MISS/200 33943 \
GET http://www.squid-cache.org/Doc/FAQ/FAQ.html - \
DIRECT/206.168.0.9 text/html
```