SQUID – PROXY CACHES

Objetivos

Instalar un Proxy-Cache (SQUID) y comprender su funcionamiento. Se usaran diversas posibilidades que admite este programa como son la restricción de accesos a determinadas URL's, crear jerarquías de caches y verificar los ficheros de log.

Proxy-Cache

Un Proxy-Cache es un servidor que acelera la navegación por web, descarga de ficheros por FTP... Utiliza el sistema de "cache", guarda las páginas más visitadas y si estas se vuelven a pedir no hace falta que se acceda al servidor remoto, simplemente el proxy tiene la página guardada en disco y la entrega. Para su correcto funcionamiento hay que configurar el navegador para que no se conecte "directamente" a internet, sino que haga todas sus peticiones al proxy.

En configuraciones grandes no se usa un único cache, se tienen diversos servidores corriendo en diversas máquinas y si un cache no tiene la página solicitada se la pide a otro (normalmente se usa una estructura de árbol).

También a veces se coloca uno o varios en la proximidad de un servidor (proxy inverso) para amortiguar la carga de peticiones sobre el servidor principal.

Squid

Squid es un proxy cache para entornos UNIX-LINUX. Usaremos la versión Squid/2.3.STABLE2. Se puede conseguir gratuitamente de `http://squid.nlanr.net` o `ftp://squid.nlanr.net` junto con documentación, FAQ...

Para traerlo, se puede usar Netscape o lynx, pero configurándolo para que use el proxy del cesca (`proxy.cesca.es` puerto `8080`) o en Netscape usar la autoconfiguración (preferencias avanzadas de configuración): `http://www.cesca.es/proxy.pac`
De esta manera sólo se traerá en realidad una copia para toda la clase (¡a ver quién es el primero!).

Configuración Básica

Squid tiene un fichero "`squid.conf`" de configuración en el que se indica en que puerto escucha, a que URL´s se puede acceder, qué tamaño tiene el cache...

Puertos de escucha

Squid escucha en un puerto (por defecto el 3128) al que se conectan los browsers para solicitarle páginas web. Para configurar este puerto hay que indicar el número de puerto en el fichero "`squid.conf`":

```
http_port 9000
```

En este puerto Squid escucha las peticiones que hacen los navegadores u otros proxys vecinos. Existe otro puerto destinado especificamente a gestionar peticiones de otros proxys, este es el ICMP, se configura el número de puerto añadiendo la siguiente línea al fichero "`etc/squid.conf`":

icp_port 10000

Hay que tener en cuenta que estos puertos de escucha no tienen que estar en uso ni reservados.

Configuración de la jerarquía de caches

Squid puede formar parte de una jerarquía de caches. Diversos proxys trabajan conjuntamente sirviendo las peticiones de las páginas. Un navegador solicita siempre las páginas a un sólo proxy, si este no tiene la página en la cache hace peticiones a sus hermanos, que si tampoco las tienen las hacen a su/s padre/s... Estas peticiones se pueden hacer mediante dos protocolos, el HTTP normal (como si de un navegador se tratase) o uno más específico (y ligero, rápido: usa UDP): el ICMP. Para especificar cuáles son los proxys hermanos y cuál es el padre de un proxy Squid se añade la siguiente línea al fichero:

	hostname	type	port	port	options
cache_peer	parent.foo.net	parent	3128	3130	[proxy-only]

El primer campo indica el IP (o alias) de la máquina donde corre el cache, el segundo el tipo de cache, pude ser "parent" (padre) o "sibling" (hermano), el tercer y cuarto campo indican los puertos de escucha de ese proxy (http y icmp respectivamente) y el último diversas opciones (ver `etc/squid.conf` para mayor informacion).

Si no se quiere que se realicen peticiones ICMP se pone ese puerto a 0.

Ej:
cache_peer proxy.cesca.es parent 3128 0
cache_peer pc1.fib.upc.es sibling 9000 10000
cache_peer pc2.fib.upc.es sibling 9669 3130

El proxy que estamos configurando tiene como padre a "`proxy.cesca.es`" al que le realiza peticiones HTTP (no ICMP) y tiene dos hermanos "`pc1.fib.upc.es`" y "`pc2.fib.upc.es`" a los que les hace peticiones HTTP o ICMP indistintamente.

Ficheros de Log

Squid genera 4 ficheros de log:

- `logs/access.log`: Aquí se guardan las peticiones que se le hacen al proxy, podemos saber cuanta gente usa el proxy, qué páginas son las más visitadas...
- `logs/cache.log`: Aquí se van guardanto los errores, mensajes de inicio...
- `logs/store.log`: Aquí se va guardando lo que pasa con el cache, que páginas (objetos) se añaden, cuales se quitan...
- `cache/log`: Aquí se mapean los objetos (páginas guardadas) al fichero en el que estan guardadas físicamente.

Para configurar donde se generaran estos ficheros de log se utilizan las siguientes líneas del fichero "`squid.conf`":

```
cache_dir ufs /home_i/users/alumnes/eng/x/xxxxxx/p6/cache 1 2 16
#Esta línea indica dónde se guardará nuestra cache (y el espacio en disco que ocupará), las
#siguientes líneas indican dónde se guardarán los ficheros de log.
cache_access_log /home_i/users/alumnes/eng/x/xxxxxx/p6/logs/access.log
cache_log /home_i/users/alumnes/eng/x/xxxxxx/p6/logs/cache.log
cache_store_log /home_i/users/alumnes/eng/x/xxxxxx/p6/logs/store.log
```

Memoria Usada

Para especificar a squid qué cantidad de memoria debe usar hay que añadir la siguiente línea al fichero de configuración:

```
cache_mem 1 MB
```

PID

Squid utiliza un fichero para guardar su propio pid, hay que especificarle un directorio:

```
pid_filename /home_i/users/alumnes/eng/x/xxxxxx/p6/logs/squid.pid
```

Controles de Acceso (ACL)

Squid es capaz de permitir o denegar accesos de navegadores a ciertas páginas, no permitir que los navegadores les hagan peticiones de ciertas páginas.... Utiliza listas de control de acceso ACL:

HTTP_ACCESS

Con este parámetro especificamos qué navegadores (u otros proxys) se podrán conectar a nuestro proxy para realizar peticiones HTTP:

```
ej
acl ourallowedhosts src 196.4.160.0/255.255.255.0
acl all src 0.0.0.0/0.0.0.0
http_access allow ourallowedhosts
http_access deny all
```

Con esta serie de "instrucciones" se definen primero las listas de control de acceso (acl ...) y segundo se permite acceso (http_access) a los hosts `196.4.160.*` y luego denegamos acceso a todos los demás. Es decir sólo pueden usar nuestro proxy unos determinados IP.

Esta "instrucción" también se puede usar para denegar acceso desde nuestro proxy a "ciertas" páginas web. Si no quisieramos que los navegadores que se conectan a nuestro proxy lo usen para conectarse a esas páginas:

```
acl adult dstdomain .playboy.com .sex.com
acl ourallowedhosts src 196.4.160.0/255.255.255.0
acl all src 0.0.0.0/0.0.0.0
http_access deny adult
http_access allow ourallowedhosts
http_access deny all
```

De esta forma nuestro proxy jamás solicitará páginas a dominios como `playboy.com` o `sex.com`, en su lugar mostrará un mensaje de error.

Nota: Hay que tener cuidado con el orden, si pusiéramos primero el "`deny all`" las demás líneas de "`http_access`" no tendrían efecto puesto que ya hemos denegado a todo el mundo (en cuanto una regla encaja dejar de mirar). Primero permitimos acceso y luego denegamos al resto.

ICP_ACCESS

De manera análoga al puerto HTTP funciona este parámetro, podemos permitir o denegar acceso a proxys vecinos para que nos hagan peticiones ICP.

Ejecutando Squid

Creando el espacio de cache

Antes de ejecutar Squid hay que crear el "sistema de ficheros" de caches, para ello ejecutamos

```
>~aad/bin/squid –f/home_i/users/alumnes/eng/x/xxxxxx/p6/squid.conf –z
```

De esta forma se creará el directorio `~/p6/cache` donde se guardarán las páginas físicas.

Creando el directorio para los logs

Hay que crear antes el directorio para que squid guarde los ficheros de log:

```
>cd ~/p6
>mkdir logs
```

Iniciando y parando Squid

Existe un script para inicializarlo, hay que cambiar dos variables del mismo para que funcione correctamente (mirar comentarios en RunCache)

```
>RunCache &
```

Para pararlo:

> ./bin/squid –k shutdown

También hay que matar el script (RunCache), sino volverá a ejecutar Squid en 10 segundos.

Nota: Si se produce algún error iniciando squid se puede consultar el fichero `~/p6/squid.out` donde aparecerán los mensajes de error.

¿Qué hay que hacer?

1.- Hay que configurar Squid para que:

a) Solo permita conexiones desde los browsers que haya en la sala de prácticas

b) No permita que esos browsers que usan Squid se conecten a `playboy.com` ni `sex.com`

2.- Hay que configurar Squid para que pertenezca a la siguiente jerarquía de caches:

a) El padre será el proxy que hay en el CESCA: `proxy.cesca.es` (puerto `8008`). Solo se le harán peticiones HTTP a este proxy (no ICMP).

b) Tenga como hermanos a los proxys de los demás alumnos, en clase se especificará una política de puertos para que no haya confusiones.

Aconsejamos primero montar squid con un padre (a) y probarlo, si todos los puntos funcionan correctamente volver a repetirlos con hermanos (b).

3.- Configurar el Lynx para que use nuestro proxy (hay que poner a la variable de entorno "http_proxy" la dirección de nuestro proxy).

ej
% **`setenv http_proxy http://pc123456.fib.upc.es:9001/`**

4.- Comprobar el correcto funcionamiento del Lynx y ver que no podemos acceder a `www.playboy.com` ni a `www.sex.com`.

5.- Estudiar los ficheros de log para entender su contenido, ver qué páginas se han solicitado, qué páginas están guardadas en el cache... Estudiar qué páginas dan `TCP_MISS`, `TCP_HIT`, `TCP_MEM_HIT`...

6.- Comprobar que hemos configurado correctamente la jerarquía mediante el comando `TRACE`.

Este comando sirve para ver el camino que hace nuestra petición por todos los caches existentes. Si ejecutamos:

> **telnet pc123456.fib.upc.es 9001** *(el nombre y puerto de nuestro cache)* **| tee temp.txt**
TRACE http://www.bc.edu HTTP/1.1
Host: www.bc.edu

(Apretar 2 veces Intro)
.
.
.
> more temp.txt
Trying 147.83.41.110...
Connected to pc123456.fib.upc.es.
Escape character is '^]'.
HTTP/1.0 500 Internal Server Error
Server: Netscape-Enterprise/3.6 SP3
Date: Tue, 28 Nov 2000 11:16:41 GMT
Content-Length: 305
Content-Type: text/html
X-Cache: MISS from sd.us.ircache.net
X-Cache-Lookup: NONE from sd.us.ircache.net:3128
Age: 0
X-Cache: MISS from proxy.cesca.es
X-Cache-Lookup: NONE from proxy.cesca.es:8080
X-Cache: MISS from pc123456.fib.upc.es
Proxy-Connection: keep-alive

<HTML><HEAD><TITLE>Server Error</TITLE></HEAD>
<BODY><H1>Server Error</H1>
This server has encountered an internal error which prevents it from fulfilling your request. The most likely cause is a misconfiguration. Please ask the administrator to look for messages in the server's error log.

Vemos que ha intentado buscar en nuestro proxy (`pc123456.fib.upc.es`) después ha preguntado al padre (`proxy.cesca.es`) y luego ha seguido preguntando al padre del padre... (`sd.us.ircache.net`)