Servidor Proxy y filtrado de contenido (SQUID+DANSGUARDIAN)

El servidor Proxy sirve para permitir el acceso a Internet a todos los equipos de una organización cuando sólo se puede disponer de un único equipo conectado, esto es, una única dirección IP. En éste apartado se tratará de explicar los pasos a seguir para instalar un servidor proxy y que exista filtrado de contenido para regular las visitas de páginas web.

Sitio: Campus Virtual UEES
Curso: Sistema Operativo Ubuntu
Libro: Servidor Proxy y filtrado de contenido (SQUID+DANSGUARDIAN)
Imprimido por: Invitado
Fecha: miércoles, 9 de septiembre de 2009, 12:18

Tabla de contenidos

1 Introducción

La palabra proxy se usa en muchas situaciones en donde tiene sentido un intermediario. El uso más común es el de servidor proxy, que es un ordenador que intercepta las conexiones de red que un cliente hace a un servidor de destino. De ellos, el más famoso es el servidor proxy de web (comúnmente conocido solamente como «proxy»). Intercepta la navegación de los clientes por páginas web, por varios motivos posibles: seguridad, rendimiento, anonimato, etc.

Ventajas
En general, los proxies hacen posibles varias cosas:

- *Control*. Sólo el intermediario hace el trabajo real, por tanto se pueden limitar y restringir los derechos de los usuarios, y dar permisos sólo al proxy.
- *Ahorro*. Por tanto, sólo uno de los usuarios (el proxy) ha de estar equipado para hacer el trabajo real.
- *Velocidad*. Si varios clientes van a pedir el mismo recurso, el proxy puede hacer caché: guardar la respuesta de una petición para darla directamente cuando otro usuario la pida. Así no tiene que volver a contactar con el destino, y acaba más rápido.
- *Filtrado*. El proxy puede negarse a responder algunas peticiones si detecta que están prohibidas.

Desventajas
En general (no sólo en informática), el uso de un intermediario puede provocar:

- *Abuso*. Al estar dispuesto a recibir peticiones de muchos usuarios y responderlas, es posible que haga algún trabajo que no toque. Por tanto, ha de controlar quién tiene acceso y quién no a sus servicios, cosa que normalmente es muy difícil.
- *Carga*. Un proxy ha de hacer el trabajo de muchos usuarios.
- Intromisión. Es un paso más entre origen y destino, y algunos usuarios pueden no querer pasar por el proxy. Y menos si hace de caché y guarda copias de los datos.
- *Incoherencia*. Si hace de caché, es posible que se equivoque y dé una respuesta antigua cuando hay una más reciente en el recurso de destino.
- *Irregularidad*. El hecho de que el proxy represente a más de un usuario da problemas en muchos escenarios, en concreto los que presuponen una comunicación directa entre 1 emisor y 1 receptor (como TCP/IP).

Funcionamiento básico:

Un proxy permite a otros equipos conectarse a una red de forma indirecta a través de él. Cuando un equipo de la red desea acceder a una información o recurso, es realmente el proxy quien realiza la comunicación y a continuación traslada el resultado al equipo inicial. En unos casos esto se hace así porque no es posible la comunicación directa y en otros casos porque el proxy añade una funcionalidad adicional, como puede ser la de mantener los resultados obtenidos (p.ej.: una página web) en una cache que permita acelerar sucesivas consultas coincidentes.

Proxy de web / Proxy cache de web

Se trata de un proxy para una aplicación específica: el acceso a la web. Aparte de la utilidad general de un proxy, proporciona una cache para las páginas web y los contenidos descargados, que es compartida por todos los equipos de la red, con la consiguiente mejora en los tiempos de acceso para consultas coincidentes. Al mismo tiempo libera la carga de los enlaces hacia Internet.

Funcionamiento

El cliente realiza una petición (p.e. mediante un navegador web) de un recurso de Internet (una página web o cualquier otro archivo) especificado por una URL. Cuando el proxy caché recibe la petición, busca la URL resultante en su caché local. Si la encuentra, devuelve el documento inmediatamente, si no es así, lo captura del servidor remoto, lo devuelve al que lo pidió y guarda una copia en su caché para futuras peticiones.

El caché utiliza normalmente un algoritmo para determinar cuándo un documento está obsoleto y debe ser eliminado de la caché, dependiendo de su antigüedad, tamaño e histórico de acceso. Dos de esos algoritmos básicos son el LRU (el usado menos recientemente, en inglés "Least Recently Used") y el LFU (el usado menos frecuentemente, "Least Frequently Used").

Los proxies web también pueden filtrar el contenido de las páginas Web servidas. Algunas aplicaciones que intentan bloquear contenido Web ofensivo están implementadas como proxies Web. Otros tipos de proxy cambian el formato de las páginas web para un propósito o una audiencia específicos, para, por ejemplo, mostrar una página en un teléfono móvil o una PDA. Algunos operadores de red también tienen proxies para interceptar virus y otros contenidos hostiles servidos por páginas Web remotas.

Proxies transparentes

Muchas organizaciones (incluyendo empresas, colegios y familias) usan los proxies para reforzar las políticas de uso de la red o

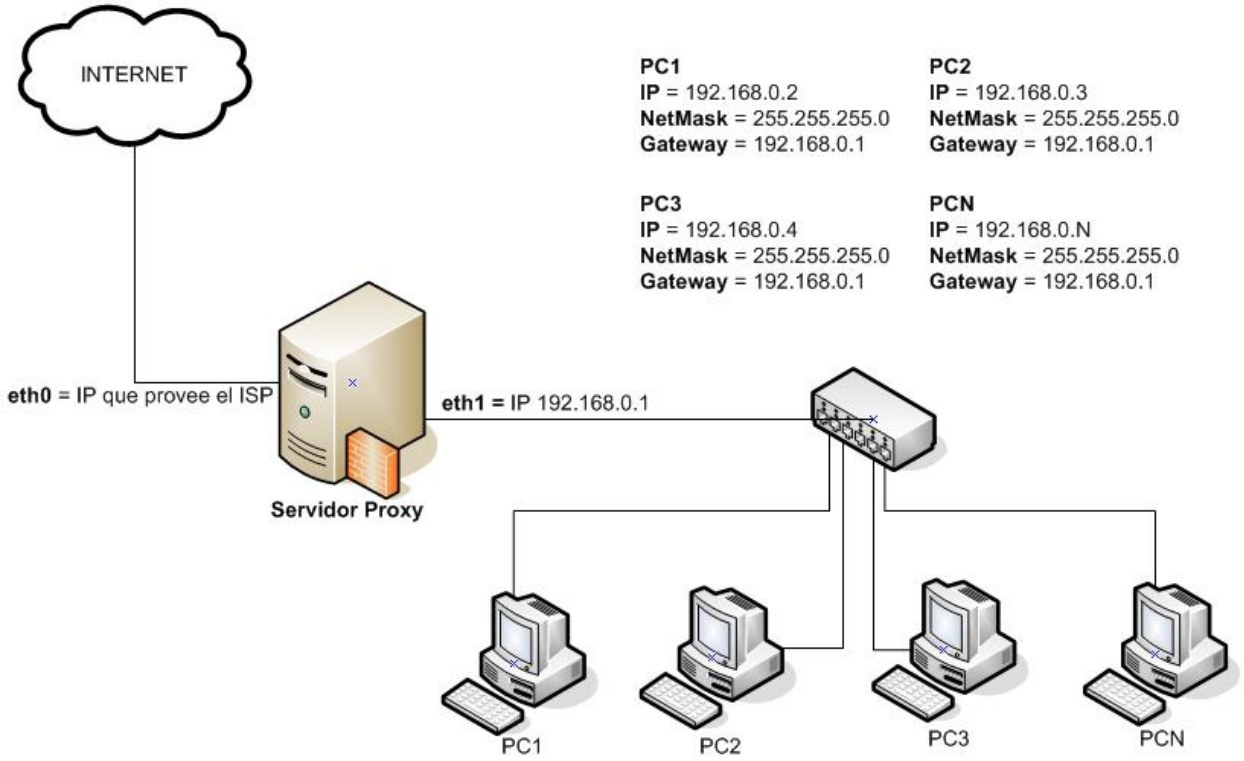
para proporcionar seguridad y servicios de caché. Normalmente, un proxy Web o NAT no es transparente a la aplicación cliente: debe ser configurada para usar el proxy, manualmente. Por lo tanto, el usuario puede evadir el proxy cambiando simplemente la configuración. Una ventaja de tal es que se puede usar para redes de empresa.

Un proxy transparente combina un servidor proxy con NAT de manera que las conexiones son enrutadas dentro del proxy sin configuración por parte del cliente, y habitualmente sin que el propio cliente conozca de su existencia. Este es el tipo de proxy que utilizan los proveedores de servicios de internet (ISP).

1.1 El propósito

El objetivo principal de este manual es de proporcionar información a los administradores de sistemas GNU/Linux para instalar y configurar un servidor proxy (squid + dansguardian), y así minimizar el riesgo de usuarios que pretendan ingresar al internet para obtener contenido sexual explícito.

Para poder instalar un servidor proxy debemos estructurar una red alámbrica similar a la que se muestra a continuación:



1.2 Configuraciones de red

Para comenzar a trabajar en el servidor Proxy, procederemos por configurar las interfacez de red según el diagrama anterior.

1. Configuración del servidor proxy

- Abrir la aplicación de configuración de la red que se encuentra en el menu “Sistema” -> “Administración” -> “Red”

- Seleccionar la interface “Conexión alámbrica (eth1)” y hacer click en el botón “Propiedades”.

- Hacer click en “Activar esta estación”
- En la opción “Configuración” seleccionar “Dirección de IP estática”
- En la opción “Dirección IP” se colocará la IP 192.168.0.1 (ver Esquema)
- En la opción “Máscara de subred” 255.255.255.0

- Hacer click en el botón “Aceptar”

En la pestaña DNS anotamos el valor que se encuentra en “Servidores DNS”, esto servirá para configurar las estaciones de trabajo.

- Hacer click en el botón “Cerrar”.

2. Configuración de las estaciones de trabajo

- Abrir la aplicación de configuración de la red que se encuentra en el menu “Sistema” -> “Administración” -> “Red”
- Seleccionar la interface “Conexión alámbrica (eth0)” y hacer click en el botón “Propiedades”.
- Hacer click en “Activar esta estación”
- En la opción “Configuración” seleccionar “Dirección de IP estática”
- En la opción “Dirección IP” se colocará la IP 192.168.0.N (Donde N se cambiará por algún valor entre 2-254 ó según la enumeración de los equipos en la red, ver esquema)
- En la opción “Máscara de subred” 255.255.255.0
- En la opción “Dirección de la puerta de enlace” 192.168.0.1
- Hacer click en el botón “Aceptar”

En la pestaña DNS en la opción “Servidores DNS”, colocar el valor que se anotó en el paso 2.3 del tema Configuración del servidor proxy

- Hacer click en el botón “Cerrar”.

2 SQUID

Squid es un popular programa de software libre que implementa un servidor proxy y un demonio para caché de páginas web, publicado bajo licencia GPL. Tiene una amplia variedad de utilidades, desde acelerar un Servidor Web, guardando en caché peticiones repetidas a DNS y otras búsquedas para un grupo de gente que comparte recursos de la red, hasta caché de web, además de añadir seguridad filtrando el tráfico. Está especialmente diseñado para ejecutarse bajo entornos tipo Unix.

Squid ha sido desarrollado durante muchos años y se le considera muy completo y robusto. Aunque orientado a principalmente a HTTP y FTP es compatible con otros protocolos como Internet Gopher. Implementa varias modalidades de cifrado como TLS, SSL, y HTTPS.

2.1 Configuración de SQUID

Configuración e instalación del servidor proxy (squid)

- Desde una terminal y como superusuario iniciar el proceso de instalación de squid y dansguardian.

  apt-get -y install squid3 dansguardian

- Antes de configurar el servidor squid es recomendable detener el servicio.

  invoke-rc.d squid3 stop

- Adquirir el nombre del host del servidor que se esté configurando. El valor obtenido de esta instrucción se utilizará para la configuración del servidor proxy.

  hostname
  <nombre de host>

- Abrir el archivo de configuración del servidor proxy squid con el editor de texto de su preferencia, este archivo se encuentra en el directorio /etc/squid3/squid.conf

  nano /etc/squid3/squid.conf

- Editar el archivo squid.conf con las siguientes opciones, se recomienda que las siguientes modificaciones se realicen en los TAGS apropiados.
- El puerto por donde escuchará el servidor proxy es el 3128

  # TAG: http_port
  http_port 192.168.0.1:3128 transparent

- ¿Cuanto desea almacenar de Internet en el disco duro?, Este parámetro se utiliza para establecer que tamaño se desea que tenga el caché en el disco duro para squid, la siguiente línea establece un caché de 1024 MB.

  # TAG: cache_dir
  cache_dir ufs /var/spool/squid3 1024 16 256

- El parámetro cache_mem establece la cantidad de memoria para los objetos en transito, objetos frecuentemente utilizados y objetos negativamente utilizados en la caché, este parámetro está limitado por la memoria RAM del sistema.

  # TAG: cache_mem
  cache_mem 32 MB

- Agregar el nombre del host (ver paso 3).

  # TAG: visible_hostname
  visible_hostname <nombre de host>

- Configurar squid para que muestre los mensajes de error en español

  # TAG: error_directory
  error_directory /usr/share/squid3/errors/Spanish

- Configurar la red local

  # TAG: acl
  acl localhost src 127.0.0.1/32
  acl to_localhost dst 127.0.0.0/8
  acl my_lan src 192.168.0.0/24 #Red local ver esquema

- Acceso a la red local

  # TAG: http_access
  http_access allow localhost

```
http_access allow my_lan
http_access deny all
```

- Guardar y salir del editor de texto.
- Iniciar el servicio de squid

```
# invoke-rc.d squid3 start
```

3 DANSGUARDIAN

DansGuardian es un filtro directo que se ubica entre el el cliente Web (web browser) y el Servidor Proxy Squid. Dansguardian acepta conexiones en el puerto 8080 y se conecta a squid en el puerto 3128. Por lo tanto, es importante que no haya otro servicio utilizando el puerto 8080.

Si lo que se desea es poder filtrar por contenido lo que se navega en tu red, DansGuardian te puede ayudar a poder poner las reglas de la navegación, solo se debe parametrizar al gusto y trabaja en conjunto con Squid.

La herramienta DansGuardian es código abierto, está desarrollada en C++ y permite una configuración flexible adaptándose a las necesidades del usuario.

Al instalar el paquete la configuración por defecto ya limita las visitas a páginas prohibidas para menores, pero dispone de gran cantidad de archivos de configuración para llevar a cabo un ajuste del servicio mas personalizado.

El mecanismo es el siguiente: los clientes mediante sus navegadores web hacen peticiones de páginas que son recibidas por DansGuardian y sólo son redireccionadas al servidor proxy SQUID aquellas que superan la fase de filtrado.

cliente web -> DansGuardian -> Squid -> servidor

En realidad DansGuardian se ejecuta como un demonio independiente del proxy, acepta peticiones en el puerto 8080 y las redirecciona al proxy SQUID, que escucha en el puerto 3128.

Por lo tanto, cuando una petición entra por el puerto 8080, DansGuardian la filtra y la pasa al proxy SQUID por el puerto 3128. Es importante, en consecuencia, que ningún otro servicio esté utilizando el puerto 8080.

Si el resultado del filtrado (dependiendo de los filtros configurados) es una denegación de acceso a una determinada página web se muestra al usuario el mensaje correspondiente al 'Acceso Denegado'.

Si DansGuardian está en la máquina que hace de cortafuegos y se configura un proxy transparente1 en SQUID, habrá que redireccionar todo el tráfico saliente en el cortafuegos del puerto 80 al puerto 8080. Es decir, se capturan todas las peticiones que se hagan a un servidor http (petición de páginas web) y se envían a DansGuardian (8080) para que se encargue del filtrado.

3.1 Configuración de DansGuardian

- Antes de comenzar se recomienda detener el servicio de dansguardian

  # invoke-rc.d dansguardian stop

- Abrir el archivo de configuración de dansguardian con el editor de su preferencia.

  # nano /etc/dansguardian/dansguardian.conf

- Editar el archivo de configuración dansguardian.conf con las siguientes opciones.
- Documentar la siguiente línea asignando el símbolo sharp (#).

  #UNCONFIGURED - Please remove this line after configuration

- Establecer el idioma de dansguardian.

  # language to use from languagedir.
  language = 'spanish'

- Verificar que las siguientes líneas se encuentren de la siguiente manera, de lo contrario modificarlas como se muestra a continuación.

  filterport = 8080
  proxyip = 192.168.0.1 #IP eth1 Ver esquema
  proxyport = 3128

- Modificar las siguientes líneas de configuración para desactivar antivirus.

  virusscan = off
  #virusengine = 'clamav'

- Guardar y salir del fichero dansguardian.conf
- Reiniciar el servicio dansguardian

  #invoke-rc.d dansguardian start

4 Configuración avanzada

Métodos de filtrado

DansGuardian utiliza un sistema de peso de las frases (**/etc/dansguardian/phraselists)** para mejorar el objetivo de bloqueo y permite filtrar por un gran número de criterios.

Los métodos utilizados son:

1. Realizar filtros utilizando el sistema de etiquetas **PICS** (Platform for Internet Content Selection).
2. Filtrar comprobando que las extensiones de los archivos y los tipos **MIME** no estén en una lista de extensiones y tipos MIME prohibidos.
3. Filtrar de acuerdo con las **URLs**, incluyendo expresiones regulares.
4. Trabajar con **listas blancas** y **listas negras.**

Compara el contenido de las páginas con el de una lista de palabras prohibidas. Esta lista contiene palabras asociadas con la pornografía y otros contenidos no deseados. Todos estos métodos se apoyan en la utilización de unos archivos de filtros que almacenan frases, palabras, URLs, etc, cuyo acceso queda prohibido.

Archivos de filtros en /etc/dansguardian/	
Archivo	Descripción
bannedphraselist	contiene una lista de frases prohibidas. Las frases deben estar entre <>. Por defecto incluye una lista ejemplo en inglés. Las frases pueden contener espacios. Se puede también utilizar combinaciones de frases, que si se encuentran en una página, serán bloqueadas.
bannedmimetypelist	contiene una lista de tipos MIME prohibidos. Si una URL devuelve un tipo MIME incluido en la lista, quedará bloqueada. Por defecto se incluyen algunos ejemplos de tipos MIME que serán bloqueados.
bannedextensionlist	contiene una lista de extensiones de archivos no permitidas. Si una URL termina con alguna extensión contenida en esta lista, será bloqueada. Por defecto se incluye un archivo ejemplo que muestra como denegar extensiones.
bannedregexpurllist	contiene una lista de expresiones regulares[3] que si se cumplen sobre la URL ésta será bloqueada.
bannedsitelist	contiene una lista de sitios prohibidos. Si se indica un nombre de dominio todo él será bloqueado. Si se quiere sólo bloquear partes de un sitio hay que utilizar el archivo **bannedurllist**. También se pueden bloquear los sitios indicados exeptuando los dados en el archivo **exceptionsitelist**. Existe la posibilidad de descargarse listas negras tanto de sitios como de URLs y situarlas en los archivos correspondientes. Están disponibles en http://dansguardian.org/?page=extras.
bannedurllist	permite bloquear partes específicas de un sitio web. **bannedsitelist** bloquea todo el sitio web y ésta sólo bloquea una parte.
banneduserlist	lista de los nombres de usuario que estarán bloqueados.

Archivos de excepciones en /etc/dansguardian/	
Archivo	Descripción
exceptionsitelist	contiene una lista de los nombres de dominio que no serán filtrados Es importante tener en cuenta que el nombre de dominio no debe incluir **http://** o **www**.

Archivos de excepciones en /etc/dansguardian/	
exceptioniplist	contiene una lista de las direcciones IP de los clientes a los que se permite el acceso sin restricciones. este sería el caso de la dirección IP del administrator.
exceptionuserlist	lista de los nombres de usuarios que no serán filtrados en el caso de utilizar control de acceso por usuario. Requiere autenticación básica o "ident".
exceptionphraselist	lista de las frases que, si aparecen en una página web, pasará el filtro.

4.1 Clasificación de contenidos

Existen diferentes sistemas de clasificación de contenidos. De ellos es muy conocido el sistema de etiquetas **PICS** (Plataforma para la Selección de Contenido de Internet) que permite que cualquiera pueda etiquetar un contenido. PICS utiliza dos métodos de clasificación:

- Clasificación llevada a cabo por los propios creadores de las páginas web.
- Clasificación llevada a cabo por terceros: en este caso la clasificación no está contenida en la propia página, sino en archivos o en servidores a los que debe acceder el usuario.

El archivo **/etc/dansguardian/pics** permite al usuario hacer un ajuste 'a la carta' del filtro de PICS. El archivo está estructurado en base a secciones PICS y cada sección contiene una descripción de las configuraciones permitidas. Las configuraciones predeterminadas de DansGuardian están pensadas para menores. Por ejemplo, los chats no están permitidos sino están moderados.

En el caso de la sección **ICRA**, valor 0 significa que no hay nada permitido en esta categoría y valor 1 está permitido.

Por ejemplo:

```
ICRAmoderatedchat = 1 #permite el chat moderado
```

La sección **RSAC** (versión antigua de ICRA, http://www.rsac.org/) contiene valores que varían de 0 (nada permitido) pasa por 2 (valor predeterminado) hasta 4, que permite todo en la categoría.

```
RSACviolence = 2
```

La sección **evaluWEB** utiliza un sistema de calificación del tipo de las películas inglesas:

- **0 = U** (**U**niversal, para todas las edades)
- **1 = PG** (recomendada la presencia de los padres)
- **2 = 18** (sólo para mayores de 18 años)

La sección **SafeSurf** (http://www.safesurf.com/) es parecida a RSAC, pero el rango de categorías es mas amplio (desde 0 para filtrar todo, hasta 9 para permitir todo).

```
SafeSurfintolerance = 3
```

Otras secciones son **Weburbia** (http://www.weburbia.com/safe/index.shtml), **Vancouver Webpages** (http://vancouver-webpages.com/VWP1.0/), etc, que utilizan otros sistemas de clasificación.

6 Archivos adicionales de filtros para DansGuardian

Además de las listas incluidas por defecto el usuario puede encontrar en http://urlblacklist.com/ archivos con filtros compatibles con DansGuardian organizados por categorías.

Vamos a descargar el archivo **bigblacklist.tar.gz** y lo copiamos a **/etc/dansguardian.** Ahora lo descomprimimos con:

```
$sudo tar xvzf bigblacklist.tar.gz
```

Por último modificamos los archivos de filtros **/etc/dansguardian/bannedsitelist** y **/etc/dansguardian/bannedurllist** para activar los filtros que nos interesen.

El usuario puede encontrar también información sobre filtros adicionales en http://dansguardian.org/?page=blacklist.

5 Configuración en Navegador (Clientes)

Configuración del proxy en el navegador web Firefox

Algunos navegadores web como Galeon y Epiphany toman la configuración del proxy del entorno gráfico de GNOME.

1.- Abrir el navegador web Firefox.

2.- Hace click en el menu “Editar” -> “Preferencias”.

3.- Hacer click en la pestaña “Red”.

4.- Hacer click en el botón “Configuración...”

5.- Seleccionar la opción “Configuración manual del proxy”

6.- Habilitar la opción “Usar el mismo proxy para todos los protocolos”

7.- En “Proxy para HTTP” escribir 192.168.0.1

8.- En “Puerto” escribir 8080

6.- Hacer click en el botón “Aceptar”.

7.- Hacer click en el botón “Cerrar”

(esta configuración tambien es compatible con el navegador web Iceweasel)