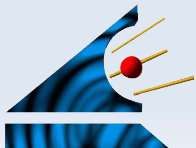


Filtrado y Cache de Contenido Web con GNU/Linux y Squid

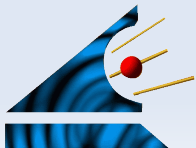
Jorge Armando Medina
Computación Gráfica de México

Abril de 2010



Agenda

- Presentación
- Introducción a los Proxy HTTP y sus ventajas
- Aceleración Web mediante cache de contenido
- Introducción al Proxy HTTP Squid
- Controles de Acceso basados en ACLs
- Listas Negras de URLs y SquidGuard
- Monitoreo, reportes y estadísticas del acceso web
- Preguntas y Respuestas
- Demostración



Introducción a los Proxy HTTP

Que es un Proxy HTTP?

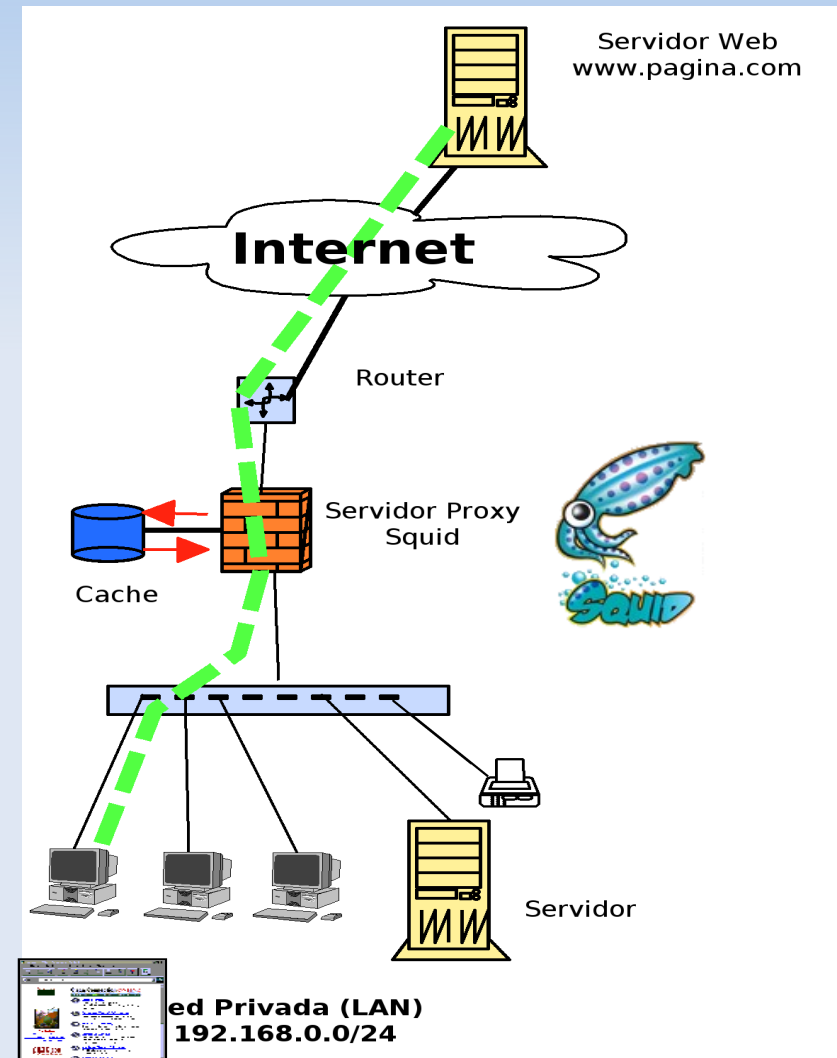
- Programa tipo servidor, recibe peticiones desde los clientes y re direcciona dichas peticiones al servidor original en medio del usuario
- Intercepta las peticiones de paginas web solicitadas por los clientes

Características principales de un Proxy HTTP

- Controlar el acceso a sitios y contenidos Web
- Cachea el contenido estático Web solicitado
- Registra las peticiones echas por los clientes

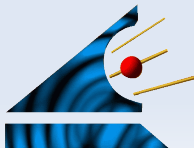
Soluciones de software libre para Proxy HTTP

- Squid



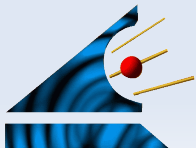
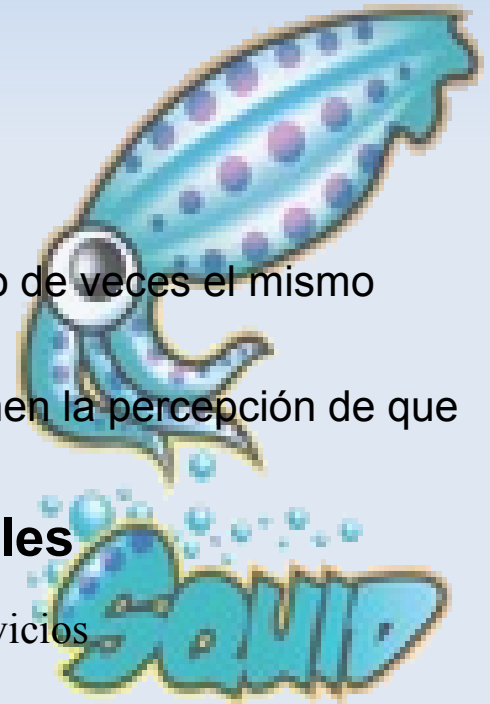
Ventajas de los Proxy HTTP

- **Control de acceso a sitios y contenido web**
 - Políticas de acceso y filtrado de contenido
 - Mejoras en la productividad de la organización
 - Asegura la red local al no permitir acceso a contenido inseguro
- **Aceleración web mediante el cache de paginas web y contenidos descargados**
 - Los usuarios tienen la percepción de que el acceso a Internet es más rápido
 - Ahorro de ancho de banda
- **Registro de accesos Web**
 - Se puede saber quien entra a que sitios y en que momento
 - Generar reportes para conocer el uso de los servicios de Internet
 - Útil para crear reglas y políticas de uso de servicios web



Aceleración Web mediante cache de contenido

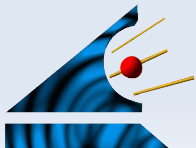
- **Que es un Proxy Cache y para que sirve?**
 - Almacena los objetos web solicitados para su re utilización
 - Cuando otro usuario solicita el mismo objeto es entregado desde el cache
 - Existe cache en caliente que es almacenado en memoria
 - El cache en frío es cuando el objeto se obtiene del disco duro
- **Porque necesitamos un Acelerador Web?**
 - Ahorrar el uso del ancho de banda, porque descargar X número de veces el mismo objeto?
 - Dar mayor servicio a los usuarios que usan el Proxy ya que tienen la percepción de que el Internet es más rápido en sus sitios habituales
- **El 20% de los objetos en paginas web son cacheables**
 - Esto se traduce en un mayor ancho de banda disponible para otros servicios



Aceleración Web mediante cache de contenido

Como mejorar el rendimiento del cache

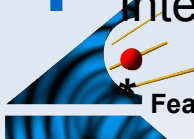
- Instale memoria RAM suficiente, entre más memoria RAM más objetos en cache de Memoria podemos usar
- Use discos duros rápidos SCSI, SATA, SAS ya que la mayoría de las operaciones son de escritura
- Para sistemas con mucha demanda no use discos IDE ya que se convierten en un cuello de botella
- Crear diferentes caches con diferentes discos duros en diferentes canales de datos o controladoras
- Squid esta optimizado para **JBOD** ó *Just a Bunch Of Drives*
- **NO** se recomienda usar arreglos de disco RAID



Introducción al Proxy HTTP

Squid: Características Generales

- Liberado bajo la Licencia GNU General Public License (GPL)
- Viene incluido y soportado en la mayoría de distribuciones GNU/Linux
- Soporta los protocolos IPv4 e IPv6
- Proxy para los protocolos **HTTP**, **HTTPS**, **FTP** y **GOPHER**
- Soporte otros Protocolos como **ICP**, **ICAP** y **WCCP**
- Cache de contenido para aceleración web con soporte de diferentes sistemas de archivos para el almacenamiento del cache
- Controles de acceso avanzados basados en ACLs
- Soporta diferentes esquemas de autenticación
- Soporta diferentes métodos de autorización
- Registro de Logs y soporte SNMP
- Soporte de plugins para autenticación de usuarios y grupos
- Integración de filtros de URLs y contenido como **SquidGuard** y **DansGuardian**



Características del Proxy HTTP

Squid: ACLs origen

Listas de Control de Acceso: Squid provee soporte para controles de accesos para los clientes basados en varios criterios de autenticación, como:

- Direcciones IP: listas, rangos, subredes
- Direcciones MAC: Solo redes locales
- Usuarios locales NCSA
- Usuarios y Grupos LDAP: OpenLDAP, Active Directory
- RADIUS
- Kerberos
- Active Directory (Single Sign On)
- NTLM (Single Sign On)
- PAM (Linux)
- Horarios

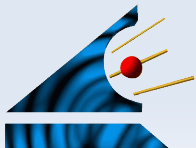
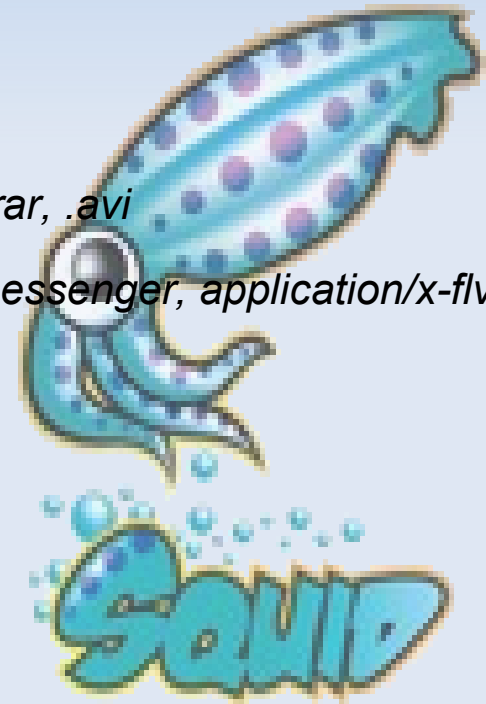


Características del Proxy HTTP

Squid: ACLs Destino

Listas de Control de Acceso: Squid provee soporte robusto y extensible para controlar las peticiones basadas en el destino y contenido, creando reglas para:

- URLs destino, ejem: *http://porn.com/downloads/free/*
- Nombres de dominio DNS destino, ejem: *dl.fileshare.com*
- Direcciones IP, ejem: *http://18.1.3.22/downloads/*
- Expresiones regulares para las URLs destino, ejem: *.mp3, .torrent, .rar, .avi*
- Tipos MIME para contenido multimedia: *audio/mpeg, application/x-messenger, application/x-flv*
- Control de acceso para peticiones y respuestas HTTP: *POST, GET*



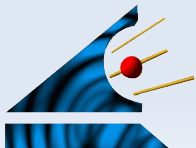
Características del Proxy HTTP

Squid: Logs

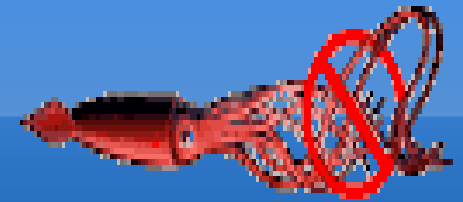
Logs Avanzados: Squid provee soporte avanzado de registro de las peticiones, en los logs podemos ver información como:

- Hora de la petición
- Dirección IP origen
- Tamaño de la petición
- URL completo de la petición
- Nombre de usuario
- Tipo MIME del contenido

Logs de objetos cacheados: Squid permite registrar metadata de cada objeto puesto en el cache, como URL, fecha de expiración entre otra información.

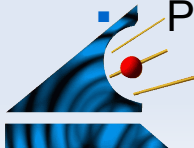


Listas Negras de URLs y SquidGuard



SquidGuard es un plugin para Squid, su función es analizar los URLs solicitados por los usuarios y compararlos contra una base de datos de listas negras y en base al resultado obtenido permitir o denegar el acceso al URL.

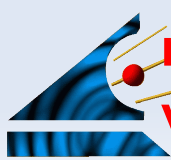
- Controlar el acceso a URLs para protocolos HTTP y HTTPS
- Control basado en:
 - Direcciones IP
 - Usuarios NCSA
 - Usuarios y Grupos LDAP
 - Usuarios y Grupos MySQL
- Soporte diferentes listas negras ya sea comerciales o de uso libre
 - **Shalla's Blacklists:** Gratis para uso privado o no comercial. Más de 1.6 millones de entradas con 70 diferentes categorías. Actualizada regularmente: <http://www.shallalist.de/cgi-bin/stat.cgi>
 - **URLBlacklist.com:** Lista comercial, con más de 2 millones de entradas. Actualizada regularmente
- Permite el uso de listas blancas para excluir sitios bloqueados (falsos positivos) por alguna categoría de lista negra
- Permite re direccionar las peticiones denegadas a una página HTML informativa



Lista de categorías en listas negras publicas Shalla

Las listas negras de uso publico Shalla incluyen las siguientes categorías:

- **adv chat downloads** finance homestyle
jobsearch movies politics religion searchengines **spyware**
warez **webradio**
- aggressive drugs fortunetelling government hospitals library
music porn remotecontrol **sex tracker**
weapons **webtv**
- alcohol costtraps dynamic **forum** hacking
imagehosting military news recreation ringtones **shopping**
updatesites webmail
- automobile dating education gamble hobby isp
models podcasts **redirector** science **socialnet**
violence webphone



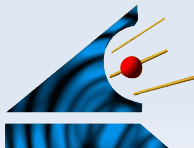
Monitoreo, reportes y estadísticas del acceso web

Generación de gráficas y estadísticas con SARG

SARG es una herramienta de análisis de logs de Squid

Mediante los reportes de uso web usted podrá obtener la siguiente información:

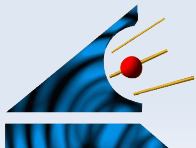
- Top Ten de sitios más visitados
- Reportes diarios, semanales y mensuales
- Accesos por usuarios
- Tiempos de navegación
- Descargas



Conexiones la Proxy Squid en tiempo real usando Squidview

- Squidview es una interfaz en línea de comando para visualizar las conexiones activas del proxy squid, algunas de sus funcionalidades son:
 - Ver quien (usuario/host) esta navegando en tiempo real
 - Que sitios/urls son los que se están visitando en el preciso momento

```
93.190 www.mercadolibre.com.ar/ja/img?
93.190 R www.mercadolibre.com.ar/org-img/jsapi/iteSrvScriptMiddle.js
93.190 H www.mercadolibre.com.ar/org-img/listados/std/arr_c.gif
93.190 H www.mercadolibre.com.ar/org-img/new2.gif
93.190 H www.mercadolibre.com.ar/org-img/calif/estr_5.gif
93.190 H www.mercadolibre.com.ar/org-img/calif/estr_4.gif
93.190 H www.mercadolibre.com.ar/org-img/beta.gif
93.190 www.mercadolibre.com.ar/ja/ml.pas.admin.pixel.cache.AdaPixelsCacheController?
93.190 oscecx-en.url.trendamicro.com/A35512038628F162E4039C7305A78C19E8782E990E28E1173A1485A8A30061BD12BC56840EBAA302579A19E368998A
93.190 www.mercadolibre.com.ar/ja/ml.advbanners.advBanners?
93.190 oscecx-en.url.trendamicro.com/E4039C7305A78C90CA378CC2A44493175E3DCDAD835A0BE24C338DE74181262368B5A08C738A4868F569856859A2F6
93.139 I www.hkinventory.com/Passport/HKIPassport.asp?
93.190 articulo.mercadolibre.com.ar/ja/ites?
93.190 articulo.mercadolibre.com.ar/ja/ites?
93.190 H www.mercadolibre.com.ar/org-img/calif/estr_6.gif
93.139 www.hkinventory.com/public/AdvancedResult.asp?
93.190 c oscecx-en.url.trendamicro.com/A1AE643E6939B710ACDFBBA01C87C32E237BF5A42298388FD1D0814964DC6CC434311F8A73A22302C48CC08DE50B
93.190 H www.mercadolibre.com/org-img/advertising/oas.js
93.190 oscecx-en.url.trendamicro.com/672092ECB0654DC27544167DF93EB1382A9317E00ADC477689EC114E9771E51C7847C28998ACF58218C276E5A88747
93.190 oas.adserving1.com/RealMedia/ads/adstream_ejx.ads/ML_MEXICO/ITM/1540/1898/9016/1715083193qTop.Right1?
93.190 H www.mercadolibre.com/org-img/advertising/JMA-RM/JMA-RM-0.3.js
93.190 M oas.adserving1.com/RealMedia/ads/Creatives/default/empty.gif
93.190 H www.mercadolibre.com.ar/org-img/advertising/alibaba09/alibaba_mobile_728_esp.jpg
93.190 c oscecx-en.url.trendamicro.com/A1AE643E6939B721A9E39A37830841AD8777D85FB38C538DC9CEE98D17804D812F315280BE16638212708D8B11C380
93.190 H www.mercadolibre.com.ar/org-img/jsapi/paspixel.js
93.190 H www.mercadolibre.com/org-img/jsapi/dejavu.js
93.190 w H www.google-analytics.com/ga.js
93.190 oscecx-en.url.trendamicro.com/1A33080C7572B0298D7231568D55848D159436F7D3ED96C9A8E5321AFED687F0E0684D06366507F03F32DE976F3FB
93.190 oscecx-en.url.trendamicro.com/A7E092A71C07384637C8D6961176C769EC90F8F79D25C17E46067823CBF9034647D25A3E256901ACAF22096C3AE5F2
93.190 paspxl.mercadolibre.com.ar/ja/PasPixel?
93.190 w www.google-analytics.com/_utm.gif?
93.190 I www.mercadolibre.com/org-img/advertising/JMA-RM/JMA-RM-0.3.js
93.190 c oscecx-en.url.trendamicro.com/949C6A0783F9949128CEDDFD1CD54884B486723385D0CB8FF18A8D5FF261B0053778FE9AA63E60739F3EBAEDCF7
93.190 H articulo.mercadolibre.com.ar/favicon.ico
93.139 I www2.hkinventory.com/public/OfferInventResult.asp?
100.00% Wed Sep 2 11:15 2009 Mon Pri | h = help
```



Calamaris: Analizador de logs de Squid

Calamaris genera reportes y estadísticas del uso del proxy, sus principales características son:

- Reportes web y por correo
- Total de peticiones realizadas al proxy
- Total de usuarios que usan el proxy
- Total de ancho de banda usado
- Cantidad de peticiones en cache
- Cantidad de ancho de banda ahorrado
- Porcentaje de ancho de banda ahorrado
- Otras estadísticas sobre dominios visitados, tipos de archivos descargados

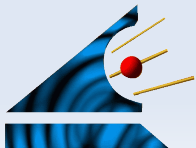
```
-----  
Proxy statistics  
-----  
Total amount: requests 96799  
unique hosts/users: hosts 73  
Total Bandwidth: Byte 1267M  
Proxy efficiency (HIT [kB/sec] / DIRECT [kB/sec]): factor 2.74  
Average speed increase: % 6.54  
TCP response time of 91.43% requests (requests > 2000 msec skipped):  
msec 340  
-----
```

```
-----  
Cache statistics  
-----  
Total amount cached: requests 32859  
Request hit rate: % 33.95  
Bandwidth savings: Byte 122M  
Bandwidth savings in Percent (Byte hit rate): % 9.66  
Average cached object size: Byte 3904  
Average direct object size: Byte 18770  
Average object size: Byte 13724  
-----
```

```
# Incoming requests by method  
method request % sec/req Byte % kB/sec  
-----  
GET 89530 92.49 2.36 1182745K 91.17 5.59  
POST 3128 3.23 3.24 109142K 8.41 10.77  
NONE 3005 3.10 0.00 5000583 0.38 3057.08  
HEAD 1091 1.13 0.22 512151 0.04 2.05  
OPTIONS 27 0.03 2.14 29143 0.00 0.49  
CONNECT 17 0.02 60.27 887 0.00 0.00  
PUT 1 0.00 3.60 0 0.00 0.00  
-----  
Sum 96799 100.00 2.30 1297300K 100.00 5.82
```

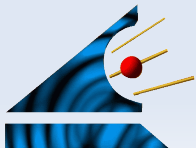
Demostración

- Uso básico de Proxy Squid
- Control de acceso por IP
- Control de acceso por usuarios y grupos
- Filtrado de URLs con SquidGuard
- Monitoreo conexiones
- Generación de reportes con SARG



Recursos adicionales

- **Squid Cache:** <http://wiki.squid-cache.org/>
- **Feature Comparison Map for Squid:** <http://wiki.squid-cache.org/FeatureComparison>
- **Squid Wiki:** <http://wiki.squid-cache.org/>
- **Squid FAQ:** <http://wiki.squid-cache.org/SquidFaq>
- **Manual Squid 3.0:** <http://www.squid-cache.org/Versions/v3/3.0/cfgman/>
- **Configuraciones de ejemplo:** <http://wiki.squid-cache.org/ConfigExamples>
- **Squid: The Definitive Guide:** <http://oreilly.com/catalog/9780596001629/>



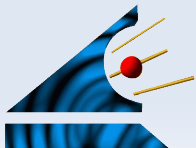
Preguntas y Respuestas

??????



Jorge Armando Medina

jmedina@e-compugraf.com



Gracias

