PROYECTO ASR

Servidor Proxy: Squid

Jose Ángel Gumiel

Introducción:

El principio operativo básico de un servidor proxy es bastante sencillo: se trata de un servidor que actúa como "representante" de una aplicación efectuando solicitudes en Internet en su lugar. De esta manera, cuando un usuario se conecta a Internet con una aplicación del cliente configurada para utilizar un servidor proxy, la aplicación primero se conectará con el servidor proxy y le dará la solicitud. El servidor proxy se conecta entonces al servidor al que la aplicación del cliente desea conectarse y le envía la solicitud. Después, el servidor le envía la respuesta al proxy, el cual a su vez la envía a la aplicación del cliente.

Los servidores proxy tienen muchas utilidades, por ejemplo, se pueden emplear para gestionar el ancho de banda, controlar los accesos a páginas web o distribuir contenido de la caché. Los cuatro principales tipos de proxy son Web, el almacenamiento en caché, reverse y transparente. Algunos servidores tienen proxies que incluyen más de una función, es decir, se pueden combinar según las necesidades.

Proxy Web: En este caso el servidor proxy actúa como un intermediario entre un cliente y otro servidor. Es decir, si un cliente A quiere solicitar un recurso a un servidor C a través de un proxy B, el cliente le hará una petición a B, que a su vez trasladará la petición a C. El servidor C tendrá constancia de que ha habido una solicitud por parte del servidor B, pero no sabrá que le petición procedía de A.

Esta configuración tiene dos enfoques, donde el principal es el del administrador del sistema. A continuación se explican las utilidades de este proxy:

- Restringir el acceso a determinadas páginas o servicios. Un administrador podría querer hacer esto por varios motivos.
    - **Seguridad:** Hay páginas que pueden ser potencialmente peligrosas y se podrían querer bloquear para evitar posibles daños. Por ejemplo páginas que sean fuente de virus informáticos.
    - **Productividad:** Si los empleados dedican más tiempo a visitar páginas o utilizar servicios no relacionados con su trabajo el rendimiento de la empresa será menor. Es por ello que se suelen bloquear algunas webs o protocolos.
    - **Contenido:** En centros de educación como colegios e institutos existen políticas de filtrado de contenido, bloqueando así el material inadecuado.

Desde el punto de vista del usuario, también se puede usar un servidor proxy web para saltarse los bloqueos anteriormente mencionados, y existen proxys públicos que se pueden usar para estos fines. Nos permiten lo siguiente:

- **Acceder a sitios donde nuestra IP está bloqueada:** Si el responsable de la seguridad de un servidor detecta **intentos de intrusión** en el sistema o en un foro un usuario se dedica a dejar **spam**, es muy probable que se bloquee al infractor. Si este bloqueo ha sido por IP, se puede saltar a través de un servidor proxy.
- **Anonimato:** Es posible que el usuario quiera dejar el mínimo de información posible.El proxy permite ocultar la ubicación del usuario. Es más, concatenando distintos servidores proxy podemos dificultar la tarea de trazado de la ubicación real.

Proxy Caché: En esta configuración también actúa como intermediario entre el usuario y el servidor al que desea conectarse. Un proxy de almacenamiento en caché, sin embargo, almacena todos los datos no cifrados en un medio de almacenamiento. Si se le pasa al proxy una petición para acceder a un sitio previamente visitado, el proxy caché enviará los datos (estáticos) que tiene almacenados, en lugar de conectar con el servidor solicitado. Este sistema permite una reducción significativa tanto en el tiempo de acceso y ancho de banda para la empresa que implementa un proxy caché. El proxy debe comparar los datos que almacena en la memoria caché con los datos remotos de manera regular para garantizar que los datos en caché sean válidos.

Tiene las siguientes ventajas:

- Acceso más rápido a los recursos solicitados y ahorro de banda ancha. Al tener la información almacenada en el servidor, si el recurso ya se ha solicitado con anterioridad, no habrá que volver a pedir todos los elementos estáticos a un tercer servidor. Esto hace que las peticiones se tramiten más rápidamente y que no haya que volver a pedir los mismos recursos una y otra vez, ahorrando ancho de banda.
- Se pueden establecer controles. El servidor puede tener una lista de sitios web a los que no está permitido acceder. Además se pueden hacer estadísticas sobre el uso que dan los usuarios a Internet y qué sitios son los más visitados.
- Tener almacenada la información en una caché permite proporcionar información al usuario aunque el servidor de origen esté fuera de conexión.

No obstante, también tiene algunas desventajas:

- Lentitud para contenido nuevo. Si el recurso solicitado no se encuentra en la caché el rendimiento será peor.
- Desactualizado. Es posible que se solicite una página que esté en la caché y que haya sido actualizada en el servidor de origen. Si la memoria no ha sido actualizada, es posible que se muestre la versión antigua.
- Privacidad. Al almacenar información y existir un servidor de por medio que almacena logs, la privacidad de los usuarios es cuestionable.

Reverse Proxy: Un proxy inverso es un servidor proxy-caché "al revés". Es un servidor proxy que, en lugar de permitirles el acceso a Internet a usuarios internos, permite a usuarios de Internet acceder indirectamente a determinados servidores internos.

El servidor de proxy inverso es utilizado como un intermediario por los usuarios de Internet que desean acceder a un sitio web interno al enviar sus solicitudes indirectamente. Con un proxy inverso, el servidor web está protegido de ataques externos directos, lo cual fortalece la red interna. Además, la función caché de un proxy inverso puede disminuir la carga de trabajo del servidor asignado, razón por la cual se le denomina en ocasiones acelerador de servidor.

Finalmente, con algoritmos perfeccionados, el proxy inverso puede distribuir la carga de trabajo mediante la redirección de las solicitudes a otros servidores similares. Este proceso se denomina equilibrio de carga, y es importante porque permite nivelar el tráfico de la red, es decir, distribuir la carga total en diferentes equipos, y también asegurar la

disponibilidad, haciendo peticiones a los servidores más cercanos o qué menos carga de trabajo tengan, con el objetivo de obtener un mejor tiempo de respuesta.

Es una configuración que puede resultar útil en los siguientes casos:

- Un servidor que quiera hacer que todo su tráfico pase primero por el proxy.
    - Si estamos frente a un sitio web grande y con millones de visitas al día, un solo servidor no será suficiente para soportar todo el tráfico. La opción que tiene dicho servidor es configurar múltiples servidores y establecer un servidor proxy que redireccionará a los usuarios al servidor más próximo a su ubicación cuando traten de acceder.
- El administrador de un servidor no quiere exponerlo directamente al público, ya que su función es cuestionable.
    - Es el caso de servidores usados para distribuir spam. No se tiene acceso al servidor principal porque se desconoce su ubicación, se pueden desconectar los servidores públicos, pero habrá otros nodos que sigan en funcionamiento, de modo que se hace difícil combatirlos. Uno de los casos más conocidos es el de "Canadian Pharmacy".

Transparente: También denominado proxy interceptor o proxy forzado. Un proxy transparente es también un proxy caché, pero intercepta las comunicaciones en la capa de red sin necesidad de ninguna configuración especial en el cliente, es decir, los clientes no tienen por qué conocer la existencia del proxy. Este tipo de proxy suele encontrarse normalmente entre el cliente e Internet, y efectúa algunas funciones propias de la puerta de acceso o del router.

Definición estándar: Un proxy transparente es aquel que no modifica la petición o la respuesta más allá de lo referente a la autenticación e identificación del proxy.

Utilidades:

- Acelerar el acceso y reducir el consumo de ancho de banda. Es usado por algunos ISP's para ahorrar ancho de banda ascendente y para mejorar el tiempo de respuesta de cara al consumidor.
- En empresas. Para establecer políticas de uso, y para aliviar los gastos generales administrativos, ya que no requiere ninguna configuración en el navegador.
  Gastos generales administrativos hace referencia a la pérdida de beneficio que supone que los empleados hagan un mal uso de la conexión, y empleen el tiempo en otras actividades no relacionadas con su trabajo.

Herramienta seleccionada:

Squid es uno de los proxies HTTP más utilizados, y puede usarse como "forward proxy" o "reverse proxy". Su característica más distinguida es su flexibilidad y personalización. Es un software libre bajo licencia GNU que está mantenido por el "Squid Development Team".

Se puede configurar de varios modos, depende de las necesidades del usuario. Los soportados son los siguientes:

- Forward Proxy o Proxy Web.
- Proxy Caché.
- Proxy transparente.
- Reverse Proxy.
- Offline o modo agresivo.
- Procesador ESI

Anteriormente ya hemos explicado en qué consistían los 4 primeros puntos.

- Modo Agresivo: Esta modalidad hace un uso mayor de la caché, se recopila más información, esto permite ver un mayor número de páginas sin tener conexión a Internet. Cada vez es menos relevante, ya que por defecto, el almacenamiento de páginas web por Squid se ha ido incrementando por defecto, y a su vez, cada vez se usan más las páginas web dinámicas.
- ESI Processor: ESI es un lenguaje de marcado para el ensamblado de contenido de web dinámico. Algunas páginas web incluyen estas etiquetas, que un procesador ESI puede leer, e indican la acción que debe ser tomada para completar el ensamblado. Squid no permite combinar esta característica con otros modos.

Herramientas como Squid han sido usadas por los ISP's desde principios de los 90's con el fin de dar una mayor velocidad de descarga y reducir las latencias, en especial para contenido multimedia y streaming de video. Las redes de distribución de contenido y los medios de comunicación utilizan este tipo de servidores proxy para que sus usuarios tengan una mejor experiencia, en concreto para efectuar un balanceo de carga y controlar los picos de tráfico que surgen cuando un contenido es muy popular.

Instalación y configuración

Para este proyecto usaré el software VMWare que permite la creación de máquinas virtuales. Usaré una distribución de Linux basado en Debian y podré simular el funcionamiento del proxy mediante otras ejecuciones de máquinas virtuales. El programa de virtualización recrea también el entorno de red, por lo que también dispondré de la ayuda de un analizador de tramas, la herramienta wireshark podría ser la idónea para realizar algunas de las comprobaciones requeridas.

La aplicación Squid únicamente deberá estar instalada en una de las máquinas, la cual actuará como servidor.

Durante la configuración de cada uno de los casos a probar se irán tomando notas y dejando reflejados todos los comandos introducidos y cambios realizados en un registro. Tras comprobar el correcto funcionamiento se realizarán copias de seguridad de los ficheros de configuración, para que se puedan cambiar fácilmente el día de la demostración.

Demostración

En la demostración se enseñará cómo se configura la aplicación Squid en los siguientes casos:

- Proxy web. Es posible emplear una máquina física o virtual y configurarla para que funcione como un forward proxy. En mi configuración es probable que no sea posible saber si realmente estoy conectándome a internet mediante el proxy comprobando la dirección IP, ya que en la red de pruebas la puerta externa va a ser la misma para todos los dispositivos, sin embargo, se puede aplicar un filtro a una determinada página, si es posible efectuar una conexión a cualquier sitio menos a los bloqueados, significará que la conexión se realiza mediante el proxy.
- Proxy transparente. Se puede configurar un proxy transparente. Para las comprobaciones habrá que conseguir pruebas que demuestren que la conexión se realiza a través del proxy. Aunque se denomine transparente, hay formas de saber si se está detrás de un proxy o no, algunas de ellas son las siguientes:
    - Comparar la IP externa del cliente con la que es vista con un servidor externo o examinando las cabeceras HTTP que recibe el servidor.
    - Mediante el comando "traceroute" podemos saber por los nodos que pasa, uno de ellos tiene que ser el proxy.
    - Intentar establecer una conexión a un servidor inexistente. El proxy aceptará la conexión, y cuando se encuentre con el problema de que el servidor no existe, devolverá un mensaje de error o terminará la conexión. En el navegador podremos ver cómo el mensaje que recibimos es diferente en caso de usar proxy o de no usarlo.

- Proxy caché. Una configuración en este modo no debería tener más complicación que las anteriores. La comprobación se basará en acceder a distintas páginas web y tratar de volver a visualizarlas sin conexión a Internet. También se analizará el directorio donde se almacenan los datos de las páginas web.

Los modos de Reverse Proxy y Procesador ESI son más difíciles de probar. El reverse proxy consiste en una máquina que actúa como servidor final, pero que toma los datos de otros servidores según conveniencia. Es un escenario que parece más complejo. El modo procesador ESI podría ser factible en el caso de disponer de alguna página que use estas etiquetas, o incluso se podría crear un archivo HTML que las incorpore.

Estos dos escenarios se intentarán probar, pero serán en función del tiempo disponible.

Estimación de las horas de trabajo

Las horas que calculo voy a necesitar para la realización de este proyecto se desglosan en la siguiente tabla:

Puesta en marcha	Realización de pruebas	Entrevista	Preparación de la presentación
1’30h.	15h.	45 minutos	3-4h.

El tiempo es aproximado, el que estimo que más pueda variar es el de realización de pruebas, es la parte que está sujeta a más imprevistos y complicaciones, y es por eso que se puede predecir un aumento en las horas de dedicación que serán necesarias.

Bibliografía

http://www.squid-cache.org/ – Página oficial del proyecto Squid.

http://wiki.squid-cache.org – Wiki oficial de Squid. Contiene documentación fiable.

https://en.wikipedia.org/wiki/Proxy_server – Información sobre servidores Proxy

https://stackoverflow.com/ – Foro de preguntas y respuestas para programadores

http://www.akamai.com/html/support/esi.html – Información oficial de ESI.