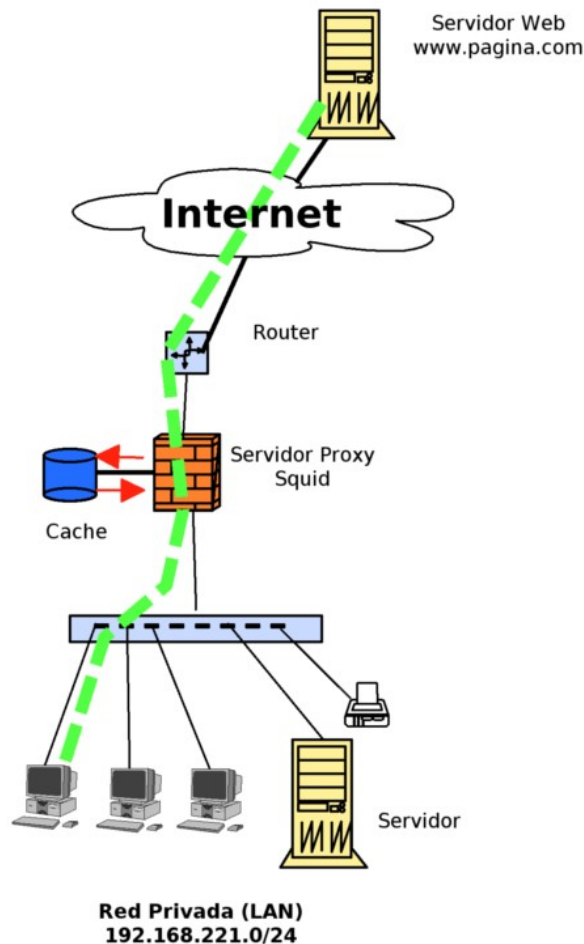


Servicio de Proxy HTTP

Conceptos

Proxy web



Un proxy es un punto intermedio entre un ordenador conectado a Internet y el servidor al que está accediendo. De ellos, el más famoso es el **servidor proxy web** (comúnmente conocido solamente como «**proxy**»). Intercepta la navegación de los clientes por páginas web

Cuando navegamos a través de un proxy, nosotros en realidad no estamos accediendo directamente al servidor, sino que realizamos una solicitud sobre el proxy y es éste quien se conecta con el servidor que queremos acceder y nos devuelve el resultado de la solicitud.

El servidor al que estamos accediendo, por tanto, no sabrá que la petición procedió originalmente del cliente, sino que para el el proxy es el origen de la petición.

Ventajas

En general los proxies hacen posible:

- **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.
- **Modificación.** Como intermediario que es, un proxy puede falsificar información, o modificarla siguiendo un algoritmo.
- **Anonimato.** Si todos los usuarios se identifican como uno sólo, es difícil que el recurso accedido pueda diferenciarlos. Pero esto puede ser malo, por ejemplo cuando hay que hacer necesariamente la identificación.

Desventajas

En general el uso de un intermediario puede provocar:

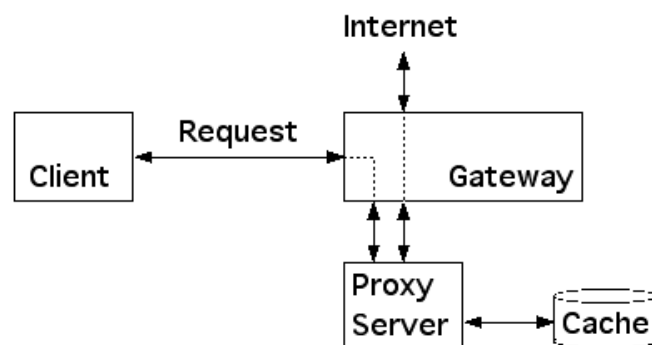
- **Abuso.** Al estar dispuesto a recibir peticiones de muchos usuarios y responderlas, es posible que haga algún trabajo que no toque. Por tanto, ha de controlar quién tiene acceso y quién no a sus servicios, cosa que normalmente es muy difícil.
- **Carga.** Un proxy ha de hacer el trabajo de *muchos* 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. En realidad este problema no existe con los servidores proxy actuales, ya que se conectan con el servidor remoto para comprobar que la versión que tiene en cache sigue siendo la misma que la existente en el servidor remoto.

Proxy cache

En informática se utilizan sistemas caché que almacenan temporalmente la información utilizada, porque está comprobado que esa misma información suele volver a necesitarse. Este truco tiene cierto sentido cuando el sistema de almacenamiento temporal es mucho más ágil (rápido de consultar) que la fuente original que contenía la información.

En el acceso web muchas páginas son consultadas repetidamente al cabo del día; podemos configurar el servidor proxy de forma que guarde una copia de las páginas que van visitando los usuarios. Así, cuando otro usuario (o el mismo usuario) quiera volver a consultar esa página, el servidor proxy ya no necesita solicitarla al servidor que la aloja en otro servidor remoto.

Este sistema se le suele denominar PROXY-CACHÉ.

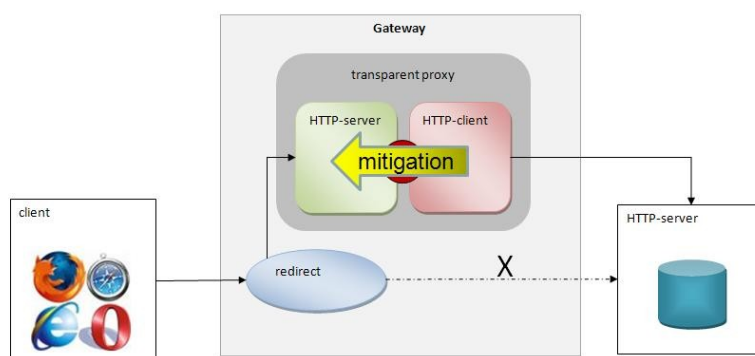


Los servidores proxy-caché no almacena todas las páginas web solo lo hará de las páginas estáticas. Existen otras páginas que el servidor remoto "construye" en el momento de servirlos (webs dinámicas).

Por ejemplo, cuando consultas un buscador, la página de resultados se prepara en el momento, dependiendo de la consulta concreta.

El tiempo que se mantiene en el proxy-caché la copia de una página concreta depende de cómo esté configurado el proxy-caché, de la actividad de los usuarios, y de otros detalles. Lo normal es que el proxy-caché almacene cualquier página que sea solicitada por un internauta; si en ese momento ya estuviera lleno el espacio de almacenamiento temporal (disco-caché) entonces hay que liberar previamente algún espacio: lógicamente se descarta la página "más vieja", la que no ha sido solicitada hace más tiempo.

El beneficio de usar un proxy-caché depende de la agilidad del sistema proxy-caché utilizado, así



como de la cantidad de información almacenada.

Algunas veces ocurre que los sistemas proxy-caché son demasiado lentos y se tarda más en recibir la información del proxy-caché que si se solicitara esa misma información del servidor remoto. Si el sistema proxy-caché no está bien dimensionado y ajustado, esto puede ocasionar una ralentización en la navegación de los usuarios.

Proxy transparente

El servidor proxy escucha por un puerto, por defecto el 3128, por lo tanto, si queremos que nuestros equipos accedan a él debemos configurar los navegadores para que pasen por el mismo.

Si no queremos tener que configurar manualmente cada uno de los navegadores o no queremos que los usuarios de la red se puedan saltar el proxy lo que se hace es configurarlo en modo transparente.

Para conseguirlo se redireccionan todas las peticiones a la puerta de enlace por el puerto 80 al puerto y la dirección del proxy.

Diferencia de filtrado entre proxy y firewall



Un firewall se puede definir como un sistema o grupo de sistemas cuya finalidad es hacer

cumplir una política de control de acceso entre dos o más redes interconectadas. Es un mecanismo que permite proteger a una red confiable de las redes no confiables con las que esta se encuentra conectada, permitiendo o denegando el tráfico entre ambas.

Los firewalls filtran a nivel de IP, por lo que sólo pueden tomar decisiones

inspeccionando la MAC, IP o puerto de origen y destino

Los proxys funcionan a nivel de transporte por lo que pueden inspeccionar los paquetes en lugar de sólo filtrarlos de acuerdo a sus direcciones o números de puerto, es decir, puede considerar el **contenido** de los paquetes.

Por lo tanto con un proxy podemos filtrar por un gran número de criterios.

- Realizar filtros utilizando el sistema de etiquetas PICS (Platform for Internet Content Selection). PICS, además de proporcionar un medio para bloquear contenidos considerados inadecuados para los niños, puede también utilizarse para describir contenidos en términos educativos, para realizar potenciales violaciones de la privacidad o para cualesquiera otros criterios que impliquen clasificación de sitios de Internet.
- Filtrar comprobando que las extensiones de los archivos y los tipos MIME no estén en una lista de extensiones y tipos MIME prohibidos.
- Filtrar de acuerdo con las URLs, incluyendo expresiones regulares.
- 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.