

Redes de Computadores II

Aula 04 - Proxy e Servidor Squid - Parte I

Apresentação

Veja aqui uma introdução a esta aula.



Vídeo 01 - Introdução

Conforme vimos nas aulas anteriores, o serviço web teve um imenso crescimento desde sua criação, na década de 1990. Esse crescimento fomentou a adição de uma série de novas funcionalidades ao serviço, dentre as quais destacamos o serviço de *proxy*.

Servidores *proxy* atuam como um intermediário entre o navegador e o servidor *web* requisitado. Normalmente, eles implementam funcionalidades para controlar e, quando possível, acelerar o acesso dos usuários ao serviço. Nesta aula, estudaremos conceitos relacionados ao uso de servidores *proxy*, bem como suas principais funcionalidades.

Na próxima aula, trabalharemos com a implementação do serviço através do software *Squid*.



Vídeo 02 - Apresentação

Objetivos

Após estudar o conteúdo desta aula, você será capaz de:

- Conceituar os servidores Proxy;
- Analisar os benefícios obtidos quando servidores Proxy são utilizados.

Servidores Proxy

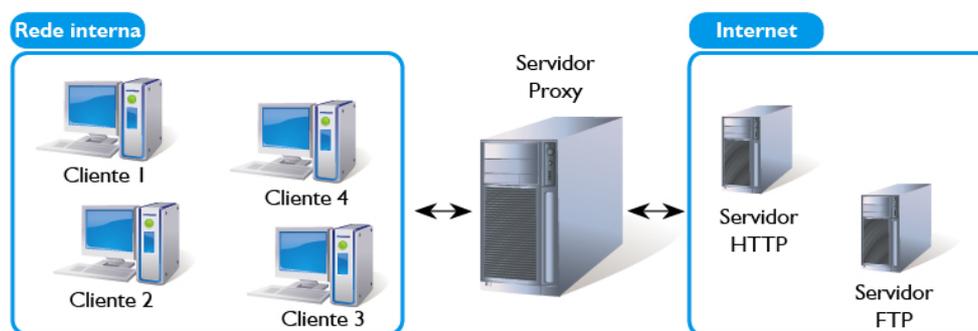
Quando seu Navegador (Firefox, Internet Explorer,...) acessa uma página na Internet, ele guarda uma cópia da página (arquivo HTML, imagens,...) no disco, de modo que se você acessar a página novamente depois, ele não precise obrigatoriamente buscá-la novamente na Internet. Isso torna sua navegação na web mais rápida.

A ideia de um proxy é semelhante, com a diferença que as páginas guardadas são compartilhadas entre todas as máquinas da rede. Desse modo, é mais fácil já existir uma cópia local da página quando alguém tenta acessá-la.

Para isso, servidores proxy atuam como intermediários na comunicação entre clientes e servidores de um serviço. Eles recebem as requisições de um ou mais clientes e as repassam para os servidores de destino. Opcionalmente, um servidor proxy pode implementar uma série de outras funcionalidades, tais como a filtragem das requisições realizadas pelos clientes e o armazenamento das respostas recebidas dos servidores em uma cache.

A Figura 1 mostra um modelo genérico de uso dos servidores proxy. Vale salientar que os servidores proxy não definem novos protocolos de aplicação, mas sim utilizam as funcionalidades dos já existentes, como HTTP e FTP.

Figura 01 - Uso de servidores Proxy



Quando utilizados como visto na Figura 1, os servidores proxy podem ser responsáveis por permitir que os clientes da rede interna acessem um serviço ou um conjunto deles na Internet. Dessa forma, se o servidor proxy dessa figura

implementa os protocolos HTTP e FTP, estes seriam os únicos serviços da Internet acessíveis para os clientes da rede interna.

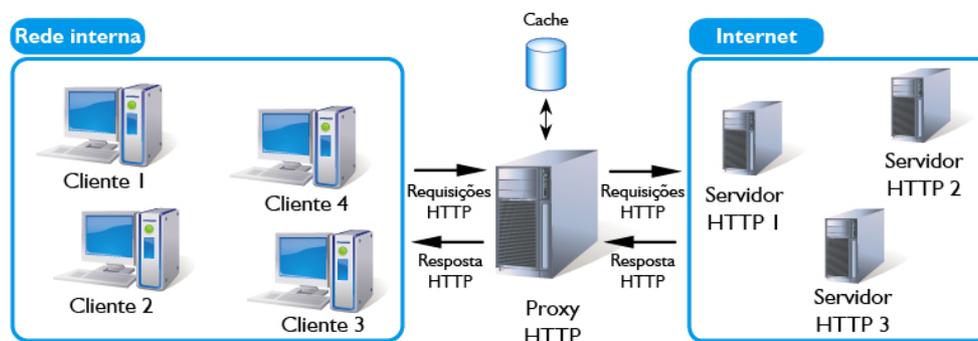
FTP (*File Transfer Protocol*) é um protocolo para transferência de arquivos.

Servidores Proxy HTTP

Dadas as suas especificidades, normalmente servidores proxy irão trabalhar com uma aplicação em particular (ou um pequeno conjunto delas).

Atualmente, o serviço de web os utiliza de forma bastante disseminada, através dos proxies HTTP. Dessa forma, um proxy HTTP irá receber e repassar as mensagens de requisição e resposta enviadas por clientes e servidores, conforme mostrado na Figura 2.

Figura 02 - Rede com proxy HTTP



Nessa figura, vemos o proxy em seu papel de intermediar a comunicação entre clientes e servidores. Nela também aparece um novo elemento de grande importância para os proxies HTTP, chamada de cache.

A cache nada mais é do que um espaço em disco onde são armazenadas as páginas já acessadas. Uma parte dessas páginas é também armazenada na memória principal (RAM), para agilizar o acesso a elas.

De modo resumido, podemos dizer que a cache funciona armazenando uma “cópia” de todas as respostas recebidas para as requisições dos clientes. Assim sendo, podemos ter o seguinte cenário:

1. Cliente 2 tenta acessar a página <http://www.imd.ufrn.br/>;
2. Proxy HTTP recebe a requisição e verifica se já possui em sua cache uma cópia da página solicitada;
3. Assumindo ser a primeira vez que um cliente acessa a página, nenhuma referência a ela será encontrada na cache. O Proxy HTTP irá então encaminhar a requisição do cliente ao servidor de destino;
4. Quando receber a resposta, o Proxy HTTP irá encaminhá-la para o cliente 2, guardando também uma cópia em sua cache.

Após algum tempo, outro cliente pode acessar a mesma página, situação em que teríamos o seguinte cenário:

1. Cliente 4 tenta acessar a página <http://www.imd.ufrn.br/>;
2. Proxy HTTP recebe a requisição e verifica se já possui em sua cache uma cópia da página solicitada;
3. Como esse é o segundo acesso à página, o proxy HTTP encontra uma cópia da página armazenada em sua cache;
4. Proxy HTTP utiliza o mecanismo de GET-Condiciona do protocolo HTTP, para se comunicar com o servidor web e verificar se a cópia da página encontrada em seu cache está atualizada.
 - Se a cópia existente na cache estiver atualizada, ela será entregue ao cliente, sem que o servidor de destino precise enviá-la novamente ao Proxy.
 - Se a cópia existente na cache não estiver atualizada, o servidor de destino enviará a página novamente ao Proxy, que a encaminhará ao cliente, e armazenará uma nova cópia na cache.



Vídeo 03 - Servidores Proxy

Atividade 01

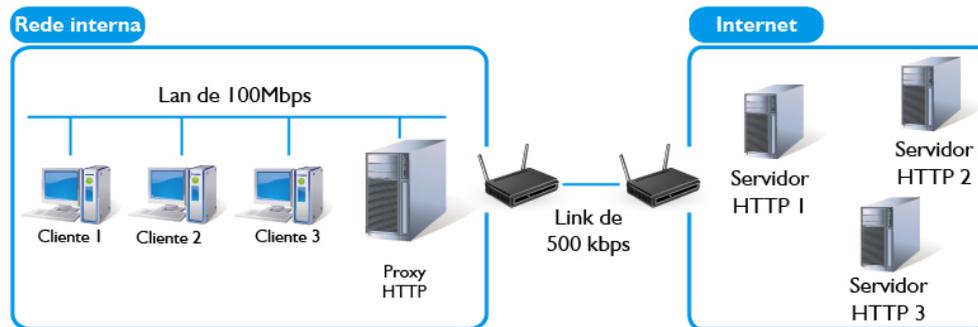
1. Pesquise sobre o funcionamento do mecanismo de GET-Condicional do protocolo HTTP. Entenda como um servidor Proxy pode certificar-se de que uma cópia de uma página presente em sua cache está atualizada.

Benefícios da Utilização da Cache

Os proxies HTTP com mecanismo de cache trazem, de imediato, duas grandes vantagens para as redes que os utilizam, conforme descrito a seguir, usando como apoio a rede da Figura 3:

- **Ganho de velocidade para páginas já acessadas:** O proxy normalmente se encontra na rede interna de cada empresa ou instituição que os utiliza. Dessa forma, o acesso a uma página que já esteja em seu cache será bem mais rápido que o acesso ao servidor de destino original.
- **Diminuição de tráfego no link com a Internet:** Em um servidor proxy bem configurado, boa parte das páginas mais acessadas pelos clientes da rede que o utilizam será encontrada diretamente em sua cache. O fato de essas páginas não precisarem estar sendo recuperadas constantemente da Internet, mas sim da cache do proxy, leva a uma diminuição importante no uso do link dessa rede com a Internet.

Figura 03 - Vantagens do Proxy com Cache.



Apenas essas duas vantagens já podem representar uma economia financeira importante, de forma a se justificar a instalação de um Proxy HTTP na rede interna de uma empresa ou instituição. Vale salientar também que, apesar das vantagens, o mecanismo de cache é de uso opcional, podendo ser desabilitado pelo administrador do servidor proxy.

Veja aqui a explicação em vídeo sobre o funcionamento de um proxy.



Vídeo 04 - Proxy Cache

Autenticação e Controle de Acesso

Historicamente, além da cache, os proxies HTTP implementam mecanismos de **autenticação** e **controle de acesso**. Os mecanismos de autenticação servem para determinar quais usuários de uma rede (ou da Internet) estão habilitados a fazer uso do proxy. Os mecanismos de controle de acesso determinam que páginas da internet podem ou não ser acessadas pelos usuários de um proxy.

Nos servidores proxy atuais, os mecanismos de autenticação podem ser bastantes simples, baseando-se, por exemplo, no endereço IP da máquina cliente que tenta utilizar o proxy. Contudo, existe a possibilidade de se configurar mecanismos mais elaborados, em que o cliente deverá fornecer um usuário e senha

para poder utilizar o proxy. Nesse caso, a lista dos usuários autorizados pode estar cadastrada localmente, no próprio proxy, ou remotamente em um servidor Radius, LDAP etc.

Administradores devem tomar um cuidado especial no momento de configurar os mecanismos de autenticação de um servidor proxy. Quando mal configurados, pode-se acabar por permitir o acesso irrestrito de qualquer máquina da Internet, ao servidor proxy de sua rede interna.

Atividade 02

1. Cite quais as vantagens de um servidor proxy com cache para uma rede.
2. Pesquise sobre os problemas de segurança relacionados aos servidores conhecidos como "open proxy".

Autenticação e Controle de Acesso

O controle de acesso é uma funcionalidade bastante utilizada nos servidores proxy. Ela permite a definição de um conjunto de ACL's (*Access Control Lists*) que, com base em uma série de parâmetros, irão permitir ou não o acesso de um cliente a uma página web.

Dentre os parâmetros que podemos utilizar para a definição das ACL's, temos:

- Endereço IP de origem;
- Endereço IP de destino;
- Domínio (site) de destino;
- Tipo de arquivo solicitado;

- Palavras chave (conteúdo da URL);
- Protocolo (HTTP, HTTPS, FTP);
- Tipo de navegador utilizado pelo cliente;
- Quantidade de conexões abertas pelo cliente;
- Data e hora do acesso à página e etc.

A definição do conjunto de ACL's a ser implementada em um servidor proxy é uma tarefa administrativa. Tomando-se como exemplo uma empresa qualquer, pode-se solicitar da direção geral um documento, como mostrado a seguir, informando os tipos de restrições a serem implementadas.

Com base nele pode-se criar as ACL's para um servidor proxy particular. Como as ACL's são bastante flexíveis, futuras mudanças nesse documento também serão facilmente refletidas no conjunto de ACL's já existentes.

Listas de controle de acesso - empresa "Fulana LTDA"

Versão: XX

Data: DD/MM/YYYY

Máquinas da diretoria:

- Acesso irrestrito à Internet.

Máquinas dos gerentes:

- Bloquear acesso ao "facebook" e "youtube" no horário de expediente.

Máquinas dos demais funcionários:

- Bloquear acesso ao "facebook", "youtube", sites do "terra" e "uol".

Máquinas dos estagiários:

- Bloquear acesso ao "orkut", "youtube", sites do "terra" e "uol".

- Bloquear todas as páginas que contenham a palavra "jogos" no endereço.

O uso adequado dos mecanismos de autenticação e controle de acesso permite controlar, de forma bastante granular, o uso da Internet em uma empresa ou instituição. Esse controle, juntamente com a cache, se constituem como as grandes vantagens do uso de servidores proxy.

Veja aqui a explicação em vídeo sobre o funcionamento dos mecanismos de controle de acesso um proxy.



Vídeo 05 - Proxy Controle de Acesso

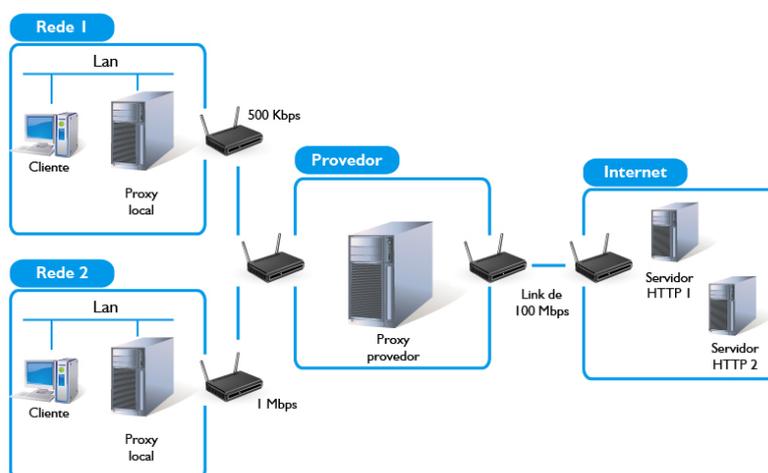
Atividade 03

1. Cite pelo menos 3 parâmetros utilizados em ACL's explicando as suas importâncias.

Hierarquias de Servidores Proxy

Servidores proxy podem ser utilizados de forma isolada ou configurados em uma hierarquia, conforme visto na Figura 4.

Figura 04 - Hierarquias de proxy.



Nesse caso, um proxy de menor capacidade (de uma empresa pequena), pode ser configurado para encaminhar todas as suas solicitações a um outro proxy (de um provedor de médio ou grande porte), e assim sucessivamente, até se chegar nos servidores de destino.

Quando se utiliza uma hierarquia de proxies, deve-se observar que mesmo que o acesso a uma página web esteja liberado no “proxy local” (Figura 4), se a mesma página web estiver bloqueada pelas ACL's do “proxy provedor”, não será possível acessá-la.

Dimensionando Servidores Proxy

A eficiência da operação de um proxy depende diretamente de uma série de fatores, tais como o espaço em disco destinado para a cache e as características de hardware da máquina onde ele está instalado.

Devido às características dos serviços que ofertam, os proxies são grandes consumidores de recursos, como processamento, memória RAM e disco rígido. Idealmente, deve-se utilizar sempre uma máquina com arquitetura de servidor, ficando ela dedicada exclusivamente para esse serviço.

Atividade 04

1. Qual a vantagem de utilizarmos uma hierarquia de proxies em nossa rede?
2. Qual a principal desvantagem que o proxy pode trazer?

Servidor proxy Squid

O Squid é um servidor proxy que suporta os protocolos HTTP, HTTPS e FTP. Seu código fonte pode ser obtido na página <<http://www.squid-cache.org/>>. Contudo, diversas distribuições GNU/Linux, como o Debian, disponibilizam pacotes pré-compilados para a instalação do mesmo. Dentre suas características, vale ressaltar:

- Suporte a *proxying* e *caching* de HTTP e FTP;
- Suporte a *proxying* de HTTPS;
- Suporte a hierarquias de proxy;
- Suporte a listas de controle de acesso (ACL's);
- Estável, seguro, alta performance;
- Disponível para diversos sistemas operacionais.

Em um servidor executando o sistema operacional Debian, para instalar o Squid, basta executar o comando:

```
apt-getinstall squid3
```

Após a instalação, o Squid já estará executando, com um conjunto de configurações básicas, porém funcionais. No entanto, é **altamente** recomendável que se personalize minimamente as configurações relativas ao:

- Uso de memória RAM;
- Uso da cache em disco;
- Listas de controle de acesso.



Vídeo 06 - Squid

Resumo

Nesta aula, você aprendeu sobre um dos serviços mais amplamente utilizados nas redes de empresas e instituições, os servidores proxy HTTP. Vimos que sua utilização trás pelo menos três grandes vantagens: o ganho de velocidade para páginas já acessadas; a diminuição de tráfego no link com a Internet e a possibilidade de definição de regras que controlam o acesso das máquinas clientes às páginas web.

Autoavaliação

Assuma que na rede interna de uma empresa o “Laboratório 1” possua os ips 10.1.1.0/24 e o “Laboratório 2” os ips 10.2.2.0/24. Usando *acl* e *http_access*, crie uma configuração do Squid para:

1. Permitir que apenas as máquinas do Laboratório 1 possam acessar o site <http://www.ufrn.br/>;
2. Fazer com que o site <http://www.xxx.com.br> só possa ser acessado pelas máquinas do Laboratório 2 e, ainda assim, apenas entre as 14h e as 18h.
3. Ninguém acessar URLs que contenham a palavra “downloads”.

Referências

HOW Squid ACLs work. Disponível em: <<http://workaround.org/Squid-acls>>. Acesso em: 4 ago. 2012.

KUROSE, J.; ROSS, K. **Redes de computadores e a internet**. 5. ed. São Paulo: Addison Wesley, 2010.

SARG: Squid Analysis and Report Generator. Disponível em: <<http://sarg.sourceforge.net/>>. Acesso em: 4 ago. 2012.

SQUIDHTTP Proxy. Disponível em: <<http://www.Squid-cache.org>>. Acesso em:4 ago. 2012.