UTILIZAÇÃO DO *SQUID* E DANSGUARDIAN COMO SERVIÇOS DE *PROXY* E FILTRO DE CONTEÚDO EM UM SERVIDOR DE SAÍDA

Cláudio Leones Bazzi[1]; Rafael Laufer[2]; Juliano Rodrigo Lamb[3]; Nelson M. Betzek[4]

RESUMO. Nos dias de hoje, com a disseminação da informação devido ao uso constante da Internet, há a necessidade de aplicação de filtros de conteúdo nas corporações para garantir que os clientes (colaboradores) não utilizem o meio (Internet) de forma indiscriminada. Os filtros atualmente em uso preocupam-se em bloquear ou não determinado conteúdo que posa conter uma palavra não interessante a empresa. Neste trabalho, avaliou-se a utilização de um servidor *squid* e *dansguardian* preocupados com a estatística da palavra em um site, liberando o conteúdo ou não através de níveis de tolerância. Sendo assim, busca-se apresentar uma solução ao problema de uso indevido de conteúdos disponíveis na rede mundial de computadores, de fácil aplicação a empresas de pequeno e médio porte.

PALAVRAS-CHAVE: servidor *proxy*, linux, controle de conteúdo, servidor *cache*.

1 INTRODUÇÃO

A Internet é considerada uma ferramenta essencial para desenvolvimento das atividades referentes ao trabalho, bem como um mecanismo que facilita a comunicação entre as pessoas, tornando toda informação disponível logo após esta ter ocorrido (HAHN, 1995). Se por um lado, a Internet é uma ferramenta essencial, por outro lado, traz um sério problema, no sentido de ser detentora de tanta informação que não se consegue avaliar.

O uso excessivo desta tecnologia pode se tornar uma aliada na falta de produtividade, uma vez que são verificados vários casos onde usuários deixam de exercer suas atividades designadas para realizar a troca de informações com amigos, colegas e afins, ou até mesmo se distraindo com conteúdos que não dizem respeito ao seu trabalho, acessando até mesmo *sites* pornográficos ou não aconselháveis. Outros ainda utilizam grande parte do link de dados da empresa, realizando downloads de vídeos, músicas, bem como a utilização de rádios on-line que simplesmente comprometem o desempenho do link afetando os que realmente desejem utilizar a ferramenta para trabalho.

A utilização de firewall e servidores *proxy* pode vir a limitar o acesso a determinados conteúdos presentes na *web*, geralmente criando-se uma tabela de palavras ou expressões proibidas. Essa abordagem pode trazer alguns problemas, pois pode bloquear *sites* de conteúdo que façam apenas uma citação à palavra bloqueada. O

[1] Mestre em Engenharia Agrícola. Professor da Universidade Tecnológica Federal do Paraná – UTFPR (Campus de Medianeira) bazzi@utfpr.edu.br
[2] Acadêmico do curso de Tecnologia em Gerenciamento de Redes – Cesufoz, Foz do Iguaçu – PR.
[3] Mestre em Engenharia Agrícola. Professor da Universidade Tecnológica Federal do Paraná – UTFPR (Campus de Medianeira), juliano@x87.eti.br
[4] Especialista em Redes de Computadores. Universidade Tecnológica Federal do Paraná – UTFPR (Campus de Medianeira) nmbetzek@utfpr.edu.br

objetivo deste trabalho foi implementar uma solução usando os servidores *Squid* e *Dansguardian* de modo que seja criada uma estatística do numero de vezes que a palavra apareça, para então se decidir sobre o acesso ou não ao *site*.

2 MATERIAIS E MÉTODOS

O estudo foi realizado em uma empresa na cidade de Foz do Iguaçu – PR, que conta com cerca de 40 estações de trabalho, possuindo um link de dados de 2 Mbps. A empresa conta também com o servidor de *firewall*, onde foram aplicadas as configurações do presente experimento.

Quando falamos de *Squid* e Dansguardian, não estamos falando de um *Firewall* mas sim de filtros de conteúdo, ou seja, todas as requisições feitas por aplicações, sejam elas *Hyper Text Transfer Protocol - HTTP, File Transfer Protocol - FTP*, entre outras, podem ou não passar por eles (MARCELO, 2006).

Um *Firewall* vem ainda antes do próprio *Proxy*, ele é responsável por fazer o compartilhamento, segurança e redirecionamento de portas do servidor. No caso, um único equipamento foi-se utilizado para fazer o papel tanto do *Firewall* como do supervisor de conteúdo, conforme pode ser visto na Figura 1.

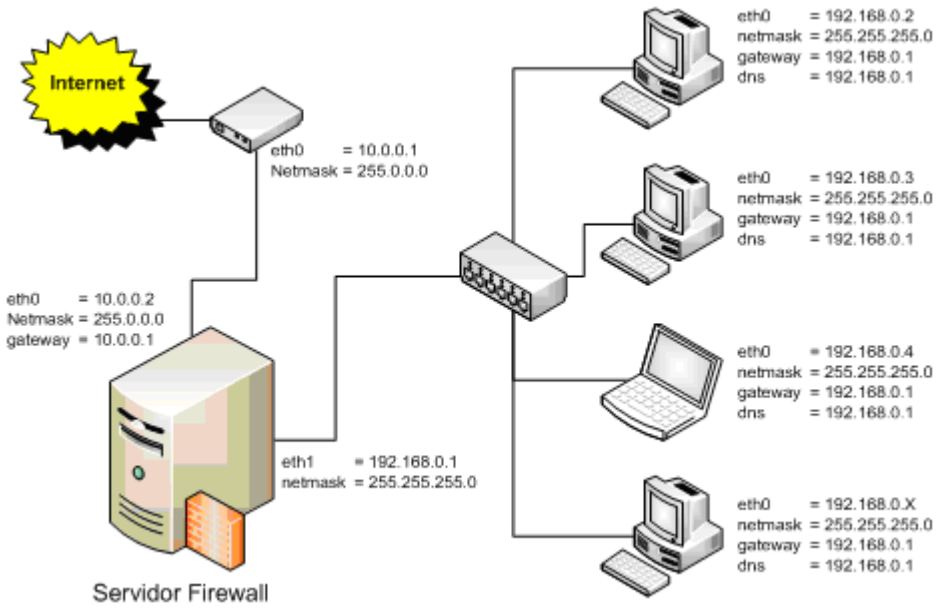

Figura1. Estrutura de Rede

Verificou-se que somente com a utilização de um servidor *Proxy* e *Cache* (*Squid*), pode-se resolver os problemas de acesso indevido, ou seja, permitindo a opção de simplesmente criar uma lista de *sites* bloqueados (ou o inverso, bloqueando tudo, exceto alguns *sites*). Na intenção de realizar um melhor controle e utilização da banda de acesso (link) da empresa, buscou-se a utilização do *software* Dansguardian, responsável por dar "inteligência" ao *Squid*.

O *Squid Cache* é uma poderosa ferramenta para implementação de *Proxy*. Um sistema aberto entende-se como gratuito, rodando em ambientes Linux principalmente. Mantido por uma grande gama de voluntários. Suportando as tecnologias de HTTP, FTP, Certificados SSL, entre outros.

O *Proxy* trabalha como *cache*, armazenando no servidor cópias dos *sites* recentemente abertos, permitindo que um segundo usuário ao requisitar o *site* que esteja

armazenado previamente no *cache*, este por sua vez verifica primeiramente se o usuário tem ou não permissão para o mesmo (trabalhando como um filtro de conteúdo), caso não tenha exibirá uma mensagem de erro padrão (que pode ser personalizada) caso tenha permissão, o *Squid* irá verificar se existem arquivos alterados no *site* requisitado, se houver, ele baixa apenas estes arquivos alterado permitindo assim um maior desempenho na navegação.

O *Squid Cache* corresponde a um servidor de *cache* de páginas da internet. Em suas configurações existem formas de se restringir *sites*, porém isso não é o suficiente para que ele seja considerado um verdadeiro filtro de conteúdo. Aí que surge o *software* Dansguardian, que trabalha junto do *Squid* agregando certa inteligência ao mesmo.

Caso fosse utilizada uma base muito extensa, o *Squid* acaba por ocupar muita memória do equipamento, ao contrário do Dansguardian. Um exemplo que pode ser citado é que o *Squid* trabalha com termos exatos, por exemplo: caso fosse solicitado o bloqueio de download de arquivos ".zip" e o *site* acessado estiver disponibilizando arquivo "_zip", os mesmos serão baixados sem problema. Um outro exemplo, seria a tentativa de burlar o *Squid* utilizando caracteres de escape ASCII como tentar baixar um arquivo ".zi%112", onde %112 corresponde a letra "p".

Tendo em vista que o *Squid Cache* é um servidor de *cache* (simplesmente *cache*) o Dansguardian passa a ter um papel importante no que se diz respeito a controle de conteúdo.

Seu funcionamento é bastante simples: o cliente faz uma requisição de uma URL no browser, este por sua vez solicita ao Dansguardian, que por sua vez irá comparar a URL com sua base de bloqueio. Caso seja encontrada, o processo para neste ponto onde o Dansguardian irá redirecionar a página para uma página interna de negação do acesso, evitando assim um consumo maior de banda e processamento com esta solicitação inapropriada. Caso a URL solicitada passe pelo Dansguardian, este envia a requisição ao *Squid*, que busca em sua base *cache*, e na internet. Após encontrado, o *Squid* entrega a página ao Dansguardian que irá "ler" todo o conteúdo da mesma na busca de palavras pré-estabelecidas para fazer a pontuação da "leitura".

Somando a pontuação, das vezes na em que determinados termos estão escritos no *site*, obtém-se um resultado da pontuação, este resultado é comparado com a pontuação máxima de tolerância definida. Se o valor ultrapassar o mínimo a página de bloqueio será exibida, caso contrário o conteúdo poderá ser visualizado.

Como o foco principal foi a implementação do *Squid* e Dansguardian, partiu-se do princípio que já existia um servidor de *Firewall* devidamente implementado, assim, pode-se concentrar apenas na implementação dos dois serviços. O *Proxy Squid*, por padrão opera na porta 3128, enquanto o Dansguardian na porta 8080, estes números de portas podem ser alterados, mas há a necessidade de se ter muito cuidado na alteração de modo que não comprometa o funcionamento de nenhum outro serviço.

Existem duas formas principais de operar-se o sistema, no modo transparente (1), o usuário final (cliente) não necessitaria efetuar nenhuma configuração em suas aplicações, ele sequer iria saber (a não ser quando uma mensagem de bloqueio aparecer) que está passando por um filtro de conteúdo. A forma não transparente (2), todas as requisições para porta 80, 443, 21 (as portas ficaram a critério de cada sistema) seriam bloqueadas e haveria a necessidade de se configurar um *Proxy* na aplicação do cliente.

Optou-se por fazer o modo transparente, assim apenas as requisições para a porta 80 seriam internamente direcionadas para o Dansguardian e armazenadas no *cache* do *Squid*. Para isso utilizou-se o seguinte comando que foi adicionado no *firewall*:

```
gandalf:~#   iptables -t nat -A PREROUTING -i eth1 -p tcp --dport 80 \
      -j REDIRECT --to-port 8080
```

No experimento, utilizou-se um servidor Pentium 4 HT de 3.0 Ghz, com 512 de memória e HD Sata de 80 GB para a instalação do Servidor. Foi instalado o GNU Linux Debian Etch 4.0. Após a instalação e configuração do Servidor como *Firewall*, houve a necessidade da instalação manualmente dos pacotes do *Squid* e Dansguardian.

3 RESULTADOS E DISCUSSÃO

Com a implantação do sistema apresentando, verificou-se um controle bastante apurado quanto ao conteúdo acessado pelos usuários, bem como o controle da real disputa anteriormente apresentada ao link pelos usuários.

A internet passou a funcionar de forma colaborativa, uma vez que após um determinado usuário requisitar um *site*, o mesmo ficou no *cache*, assim o próximo terá um acesso mais rápido. Fazendo as devidas restrições, apenas os *sites* de interesse da empresa são liberados aos usuários, assim como ilustra a Figura 2, pode-se realmente comprovar a eficácia no bloqueio dos mesmos.

Relatório de Acessos

Periodo: 03May2007-03May2007
Usuario: sonicwall
Ordem: TIME, reverse
Usuario Relatorio

LOCAL ACESSADO	CONEXÃO	BYTES	%BYTES	IN-CACHE-OUT		TEMPO GASTO	%TEMPO	
www.boerse-online.de	3	8.94K	0.03%	100.00%	0.00%	00:00:00	0.00%	NEGADO
www.autogaleria.pl	3	8.88K	0.03%	100.00%	0.00%	00:00:00	0.00%	NEGADO
busca.igbusca.com.br	4	11.62K	0.03%	100.00%	0.00%	00:00:00	0.00%	NEGADO
www.gazetaesportiva.net	3	8.71K	0.02%	100.00%	0.00%	00:00:00	0.00%	NEGADO
tools.hpg.com.br	2	5.84K	0.02%	100.00%	0.00%	00:00:00	0.00%	NEGADO
ads1.feminice.com.br	3	9.40K	0.03%	100.00%	0.00%	00:00:00	0.00%	NEGADO
www.corinthiansfutebol.com.br	3	8.78K	0.03%	100.00%	0.00%	00:00:00	0.00%	NEGADO
www.papajogos.com.br	3	8.68K	0.02%	100.00%	0.00%	00:00:00	0.00%	NEGADO
www.tuningmania.com.br	2	5.95K	0.02%	100.00%	0.00%	00:00:00	0.00%	NEGADO
www.msn_br.br.toing.com.br	2	5.82K	0.02%	100.00%	0.00%	00:00:00	0.00%	NEGADO
www.pepsi.com.br	2	5.78K	0.02%	100.00%	0.00%	00:00:00	0.00%	NEGADO
n.i.bol.com.br	10	29.04K	0.08%	100.00%	0.00%	00:00:00	0.00%	NEGADO
xml11es.farolatino.com	3	8.97K	0.03%	100.00%	0.00%	00:00:00	0.00%	NEGADO

Figura 2. Exemplo de relatório de negação da estação “sonicwall”

4 CONCLUSÃO

A utilização do servidor de *Proxy* trouxe agilidade na utilização de tarefas rotineiras, como a emissão de contratos de financiamento, acesso à informações de seguradoras, prospecção de vendas de veículos, pois estas são apenas de algumas da utilização da internet como ferramenta de trabalho. Pelo fato destes *sites* serem constantemente utilizados, o *Proxy* se tornou uma ferramenta indispensável. Tendo vista que para fins de teste, depois de estar sendo utilizado por 30 dias, desativou-se o *Squid* por um certo período de tempo e deixou-se sem *cache*, tornando os usuários insatisfeitos pela lentidão apresentada.

Os problemas com vírus foram quase extinguidos, uma vez que e-mail’s maliciosos costumam a levar para *sites* externos para serem instalados nas máquinas, e estes *sites* estão bloqueados. Os colaboradores passaram a dedicar mais tempo ao seu trabalho, uma vez que não perdem mais tempo se distraindo com *sites* de conteúdo impróprio à sua função.

BIBLIOGRAFIA

HAHN, Harley; STOUT, Rick. Dominando a Internet. São Paulo: Makron Books, 1995.

MARCELO, Antonio. Guia Rápido do Administrador de Redes. Livro *Squid* - Configurando o *Proxy* para Linux. Brasport, 2006.