REVISTA CIENTÍFICA ELETRÔNICA DE SISTEMAS DE INFORMAÇÃO - ISSN 1807-1872
PUBLICAÇÃO CIENTÍFICA DA FACULDADE DE CIÊNCIAS JURÍDICAS E GERENCIAIS DE GARÇA/FAEG
ANO II, NÚMERO, 03, AGOSTO DE 2005. PERIODICIDADE: SEMESTRAL

A IMPORTÂNCIA DA UTILIZAÇÃO DE CONTROLE DE CONTEÚDO NO ACESSO A WEB

ROSA, Adriano Justino
TAMAE, Rodrigo Yoshio
Docente da Faculdade de Ciências Jurídicas e Gerenciais - FAEG/Garça-SP
adriano@faef.br; rytamae@yahoo.com.br

SPINOLLA, Diego de Castro
Acadêmico da Faculdade de Ciências Jurídicas e Gerenciais - FAEG/Garça-SP

RESUMO

O objetivo deste trabalho é demonstrar os resultados obtidos a partir da implementação de uma solução de filtragem de conteúdo Web utilizando os recursos oferecidos para a plataforma Linux (distribuição Slackware), como por exemplo, Squid e Dansguardian, bem como demonstrar a importância do seu uso em ambientes empresariais.
Palavras-chave: Controle de conteúdo, Filtragem, Squid e Dansguardian.

ABSTRACT

The objective of this work is demonstrate initial results obtained from Web filtered content implemented using resources offered to Linux platform (Slackware distribution), by example, Squid and Dansguardian, thus, demonstrate the importance of this use em business environments.
Keywords: Mannager content, Filter, Squid and Dansguardian.

1. INTRODUÇÃO

Haverá um dia que teremos mecanismos de buscas de informações da internet rápidos, precisos, eficientes e inteligentes, talvez utilizando-se de técnicas de IHM (Interface Homem-Máquina), Web Semântica, Ontologias e Realidade virtual, como retratado no filme "*The Time Machine*" de S. Wells, onde a máquina reconhece o usuário e avalia, inclusive, se tal informação é ou não apropriada ou relevante para o usuário, diferentemente do que ocorre

atualmente, pois apesar de rápidos e relativamente precisos, os sistemas de buscas não são inteligentes o suficiente para essa tarefa. Primeiro simplesmente o dado, depois a informação e, finalmente, depois de transformado, o conhecimento é a principal ferramenta de gestão para as empresas. A integração de todo esse universo tornou-se possível principalmente após a popularização da Internet e da viabilização dos sistemas e operadoras de comunicação de dados. Com isso qualquer empresa ou pessoa pode conectar-se com outra em qualquer parte do mundo a custos praticamente irrisórios. No entanto, juntamente com toda essa imensa massa de informação chegam outras que deveriam ser descartadas antes de serem encaminhadas ao seu destino.

Para impedir que informações inerentes aos negócios de determinada organização sejam trafegados do mundo externo para o interno, ocupando largura de banda, tempo produtivo da organização, bem como, influenciando na formação educacional e moral de nossos filhos, devem ser implementados mecanismos capazes de analisar tal conteúdo, através de técnicas que possibilitem avaliar palavras e montando frases ligadas às regras de negócios ou simplesmente realizar, por exemplo, filtragem de conteúdo pornográfico.

2. ASPECTOS RELEVANTES PARA IMPLEMENTAÇÃO DE UMA SOLUÇÃO PARA FILTRAGEM DE CONTEÚDO WEB

Simplesmente limitar o acesso a informações pode reduzir a produtividade, pois dependendo o método de bloqueio será necessário a intervenção do Administrador de Rede para autorizar o acesso a determinados sites. Os principais métodos de bloqueio a sites e serviços de internet são através de filtragem de pacotes, segundo (KUROSE 2006) o datagrama será aceito ou descartado levando-se em consideração aspectos como endereço IP de origem ou destino, porta TCP ou UDP, tipo de mensagem ICMP e datagrama de inicialização de conexão SYN ou ACK.

A filtragem é obtida depois de analisados os datagramas dos pacotes e os critérios estão descritas em regras que são executadas sequencialmente, constituindo assim um Firewall (KUROSE, 2006).

É importante observar a localização física do Firewall na rede, pois as regras nele implementadas sofrerão alteração dependendo desta, ressaltando

que todo o fluxo de dados entre a rede pública e as diversas redes internas deverão passar por um único ponto de distribuição (DIMARZIO, 2001).

Como pode ser observado na Tabela 1 é possível visualizar e imaginar a complexidade da tarefa de gerenciamento de acesso a Web, principalmente quando se tem um número elevado de pontos de acesso com Internet, pois para todo nova regra torna-se necessário a intervenção do Administrador de rede, mesmo com a utilização de proxy-cache que também possuem mecanismos de filtragem de pacotes.

Tabela 1 – Regras de filtragem de pacotes implementas no Firewall

Regra	Endereço Origem	Endereço Destino	Porta	Ação	Comentário
R1	10.66.25/24	0.0.0.0/0	all	accept	Libera trafego
R2	10.66.25/24	Playboy.com.br	all	deny	Bloqueia o acesso ao site playboy.com.br
R3	10.66.25/24	0.0.0.0	all	deny	Bloqueia todo acesso
R4	10.66.25.10	bb.com.br	all	accept	Permite acesso ao BB
R5	0.0.0.0/0	0.0.0.0/0	23	deny	Fecha porta telnet

3. MATERIAIS E MÉTODOS

Com o objetivo de tornar acessível à implementação do sistema em pequenas, médias e grandes empresas, e inclusive em residências, escolheu-se a plataforma Linux Slackware 10.2 com Kernel 2.4.32, pois o objetivo do sistema seria atender aproximadamente 80 usuários em um servidor Intel Pentium IV de 2,8 Ghz, HD IDE de 80 Gigabytes e 1 Gigabyte de memória RAM.

Conforme pode-se observar na figura 1, neste servidor foram instalados e configurados três serviços: O IPTables com objetivo de definir as políticas de acesso em cada uma das redes, bem como a filtragem básica dos pacotes que são entregues pelo roteador de borda; O Squid com a função de buscar e manter em cache as páginas mais utilizadas pelos usuários e, finalmente, o Dansguardian que recebe diretamente as solicitações dos usuários, realiza a busca na *Blacklist* e, caso não encontre, solicita ao Squid que efetua, então, a busca e o conteúdo que será novamente entregue ao Dansguardian para que, desta vez, analise todo o conteúdo da página.

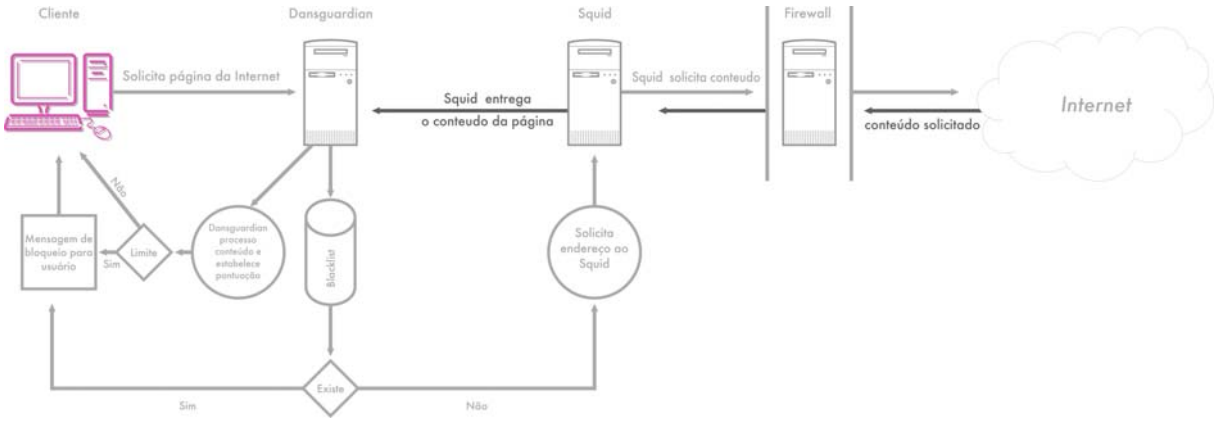

Figura 1 - Fluxo de dados recebidos pelo usuário

Durante a análise deste conteúdo o sistema irá pontuar palavras previamente estabelecidas (como por exemplo: Sexo (30), Anal (30), Teen (30), Adolescentes (20)) e depois o sistema efetua a somatória e estabelece uma pontuação que será comparada com os parâmetros de configuração do sistema.

Caso a pontuação seja maior que o permitido o Dansguardian entrega uma mensagem ao usuário informando que a página em questão está fora dos padrões para ser exibida, o motivo do bloqueio e inclui o referido endereço em sua lista de bloqueio (e com isso o sistema “aprende”) e não devolve o conteúdo ao cache do Squid.

Após a primeira semana foram iniciados os devidos ajustes, parametrizando os arquivos de palavras-chave e seus respectivos pesos. Como o cenário foi uma instituição de ensino superior onde alunos trabalham com projetos de pesquisa nas mais diversas áreas, em algumas situações o usuário ao realizar a pesquisa no Google com a frase: “Sexo com adolescentes” era imediatamente bloqueado pelo Dansguardian, mas ao reformular a frase e utilizar: “Sexo na adolescência” ou “Sexo com adolescentes .pdf” o sistema trará o conteúdo sem maiores problemas.

A solução também foi implementada em uma instituição de ensino, porém atendendo a alunos de faixa etária mais baixa, enfatizando ainda mais a preocupação com conteúdo pornográfico.

Em ambos os casos, foram efetuadas análises dos registros de *logs* e de cache do Squid e do Dansguardian para serem identificadas as principais palavras que poderiam ser consideradas inerentes as regras de negócios das respectivas instituições.

4. CONCLUSÕES

A solução mostrou-se altamente satisfatória, permitindo identificar, logo num primeiro momento, alguns pontos de ajustes técnicos. O cache mantido pelo Squid proporciona economia da utilização do *link* em mais de 50%, o que poderia, inclusive, representar redução de custo, pois permite que a instituição utilize *link* menor.

Em relação a eficiência do controle de conteúdo pode-se verificar que a solução atendeu por completo as expectativas, pois todos os sites relacionados a "sexo explicito", "páginas de bate-papo" e "sites de relacionamento" considerados pela instituição como conteúdo inadequado foram bloqueados.

Operacionalmente, o aspecto que mais se destaca ocorre em relação a manutenção da solução, pois praticamente não a intervenção do Administrador de rede que efetua somente alterações nas listas de exceções e em outros parâmetros de configurações, bem como, *backup* do sistema e de seu cache.

Em relação à produtividade pode-se afirmar que os colaboradores estão acessando somente conteúdo adequado e inerente aos negócios da instituição.

5. REFERÊNCIAS BIBLIOGRÁFICAS

KUROSE, J. F.; ROSS, K.W. **Redes de Computadores e a Internet**. 3.ed. São Paulo: Pearson Addison Wesley, 2006.

DIMARZIO, J. F. **Projeto e Arquitetura de Redes** – Um Guia de Campo para Profissionais de TI. Rio de Janeiro: Campus, 2001.