Captura e Bloqueio de E-mails com URLs Maliciosas

Eduardo Horowitz, Leandro Fortes Rey

Centro de Processamento de Dados– Universidade Federal do Rio Grande do Sul (UFRGS)
Rua Ramiro Barcelos 2574, 90035-003 – Porto Alegre – RS – Brazil

{eduardoh, leandro}@cpd.ufrgs.br

Abstract. *Nowadays, e-mails with malicious URLs are widespread and are the cause of many misfortunes to users. In this paper, the solution found and implemented by CPD of UFRGS is described. Initially, we expose the idea of using a malicious links black list. After this, we describe the malware capture system that was implemented. Following, we show the original solution that consists on the utilization of the ClamAV antivirus for the detection of malicious links on e-mails. Next, we discuss the results obtained with the proposed solution. Lastly, we conclude that the results were good, but that there are still challenges to be tackled in this area.*

Resumo. *Hoje em dia, e-mails com URLs maliciosas são bastante comuns e causam muitas vítimas. Neste artigo, a solução encontrada e implementada pelo CPD da UFRGS para combater tal problema é descrita. Inicialmente, expomos a idéia de utilização de uma lista negra de links maliciosos. A seguir, descrevemos o sistema de captura de malwares implementado. Mais adiante, mostramos a solução original relativa a utilização do anti-vírus ClamAV para detecção de links maliciosos em e-mails. Depois, discutimos os resultados conseguidos com a solução proposta. Por fim, concluímos que os resultados foram positivos, mas que ainda há desafios nesta área.*

1. Introdução

No final de 2005, houve um grande aumento na quantidade de e-mails contendo links maliciosos. Isto se deveu a vários fatores, especialmente à popularização das botnets [10], redes de computadores contaminados que se tornaram robôs (*bots*) controlados remotamente. As botnets são utilizadas para as mais diversas atividades maliciosas, como ataques de negação de serviço e envio de spam. Para assegurar sua consolidação e expansão, as botnets também precisam contaminar novos computadores e uma das mais eficientes formas para tal é o envio de e-mails não solicitados com URLs maliciosas. Na Figura 1 encontra-se um exemplo deste tipo de e-mail, que parece ser da Receita Federal, mas no qual todos os links apontam para um arquivo .exe, que é um malware (*malicious software*) de um domínio fora do Brasil.

Devido ao fato de muitos e-mails falsos com URLs maliciosas serem difíceis de identificar, especialmente por usuários leigos, que formam boa parte dos usuários de uma Universidade, tornou-se cada vez mais urgente uma forma de minimizar o impacto negativo de tais mensagens. Na pesquisa que foi realizada com esse fim, a utilização de RBLs (*Real-time Blackhole Lists*) e de outras técnicas *antispam* não se mostrou suficientemente satisfatória. Seria necessário encontrar um novo tipo de solução.

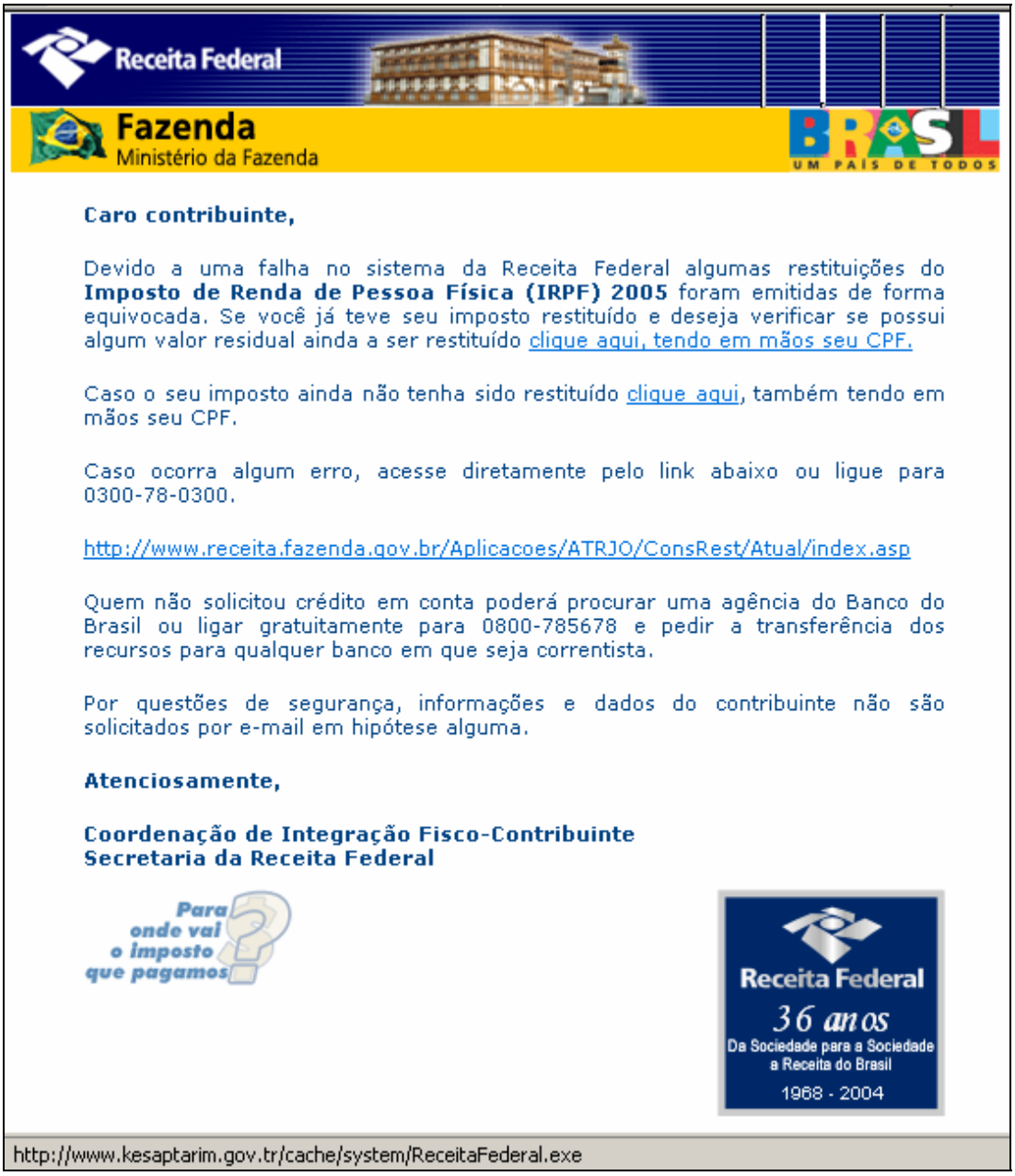

Figura 1: Exemplo de e-mail com URL maliciosa

2. Lista Negra de Links Maliciosos

A nova solução idealizada seria uma lista negra de links maliciosos contidos em e-mails, de forma que estes e-mails pudessem ser bloqueados antes de chegar a suas vítimas e de forma que o acesso a essas URLs fosse negado. Inicialmente, pensamos em utilizar a SURBL [1], mas esta fugia um pouco do propósito pretendido. A idéia de criar uma lista negra das URLs encontradas em e-mails que passavam pelos servidores da UFRGS era outra possibilidade, mas poderia levar algum tempo para ser implementada e uma solução imediata era necessária.

Foi então que, em dezembro de 2005, durante as pesquisas em busca de uma solução, encontramos a *Malware Block List* (MBL). Na definição contida no próprio site da lista na época: "*The Malware Block List is a free, automated and user contributed system for checking URLs for the presence of Viruses, Trojans, Worms, or any other software considered Malware. The list is available in some formats*." [2]. Desta forma, a MBL se mostrou uma solução adequada, bastante próxima do que desejávamos, inclusive com a possibilidade de podermos contribuir com URLs capturadas em nossos sistemas de e-mail.

3. Sistema de Captura

A fim de capturar URLs que apontavam para malwares contidas em e-mails, para que fossem analisadas, bloqueadas e alimentadas na Malware Block List, decidiu-se pelo desenvolvimento de um pequeno sistema. Este sistema consiste basicamente da utilização de programas em linguagem C e de shell scripts. O requisito inicial do sistema era o de capturar pacotes que continham as URLs maliciosas. Estas são identificadas como links que possuem alguma extensão potencialmente maliciosa, como exe e scr. É importante ressaltar que o objetivo de captura do sistema era de links para malwares que tinham como alvo sistemas nos quais a extensão em arquivos é importante, mais especificamente sistemas Microsoft Windows, que são as principais vítimas de tais ataques.

Para a captura de pacotes, o programa ngrep (*network grep*) [3] mostrou-se eficiente e simples de usar, sendo o escolhido. No entanto, os pacotes capturados contém dados binários e uma série de outros dados supérfluos, sendo necessária, desta forma, a utilização de algum procedimento para se extrair apenas a URL que fora detectada. Devido ao *modus operandi* dos malfeitores que enviam e-mails com URLs maliciosas (isto é, o envio de milhares de mensagens idênticas mas para contas diferentes), não seria eficiente em relação a espaço armazenar a mesma URL milhares de vezes. Desta forma, um programa em linguagem C realiza tanto a extração da URL quanto a eliminação de redundâncias. A fim de não gerar arquivos com capturas de pacotes muito grandes, um shell script coordena a captura de pacotes e sua extração, realizando tal tarefa a cada 30 minutos. No final do dia, outro shell script envia para os responsáveis da rede da UFRGS a lista de URLs novas apreendidas nas últimas 24 horas. Os responsáveis, então, fazem uma análise manual das URLs, removendo as que podem, de alguma forma, comprometer a privacidade dos usuários do sistema de e-mail. Além disso, URLs comprovadamente não maliciosas também são removidas, a fim de evitar falsos positivos. A lista final é então processada e enviada para a MBL. Este procedimento de limpeza realizado manualmente por um humano necessita, em média, de 15 minutos, diariamente.

Na Figura 2 encontra-se um breve esquema em alto-nível do sistema de captura de URLs.

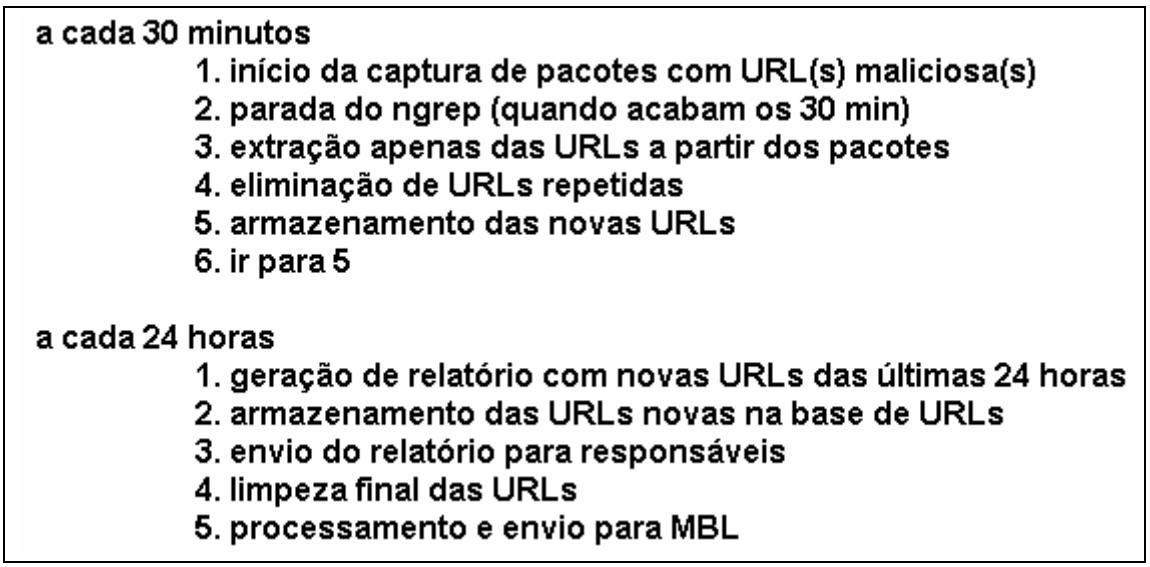

Figura 2: Esquema do sistema de captura de URLs maliciosas

O CPD da UFRGS começou a contribuir com a MBL quando haviam sido enviadas aproximadamente 7 mil URLs para a mesma. Contribuímos com mais de 12 mil URLs, sendo que

a MBL já recebeu mais de 50 mil URLs.

Uma vez possuindo um modo de contribuir com a MBL, era necessário um mecanismo para bloquear e-mails que continham URLs maliciosas da lista negra. É interessante notar que, neste ponto, a utilização da MBL com um proxy Web (como Squid Web Proxy Cache [4]) já era possível e foi utilizada para bloquear o acesso às URLs maliciosas. No entanto, o proxy na UFRGS era de uso opcional e, além disso, nada impediria que, recebendo e-mails com estas URLs, os usuários as abrissem em algum outro lugar.

4. Detecção de URLs Maliciosas em E-mail: ClamAV

Alguns pré-requisitos para que se bloqueiem e-mails que contenham URLs maliciosas são:

1. Eficiência - é necessário que a análise extra para verificar a existência das URLs não atrase a entrega dos e-mails;
2. Tratamento de Base64 - boa parte dos e-mails maliciosos são enviados não em texto plano, mas em HTML com codificação MIME (Multi-purpose Internet Mail Extension), que utiliza Base64, o que fazia a verificação de URLs neste tipo de e-mail fundamental;
3. Simplicidade na implementação, utilização e atualização.

Para tentar atingir esses pré-requisitos, foram estudadas algumas alternativas. Inicialmente, se pensou em utilizar filtros nativos do Postfix (*body-checks*), que é o servidor de SMTP utilizado pelo Chasque Mail [5], o serviço de e-mail administrado pelo CPD da UFRGS. No entanto, encontrou-se o seguinte em [6]: "*Header/body checks do not decode message headers or message body content. For example, if text in the message body is Base64 encoded (RFC 2045) then your regular expressions will have to match the Base64 encoded form.*" e "*Despite warnings, some people try to use the built-in filter feature for general junk e-mail and/or virus blocking, using hundreds or even thousands of regular expressions. This can result in catastrophic performance failure.*". Desta forma, os pré-requisitos 1 e 2 não seriam atingidos. O primeiro pois o tamanho da lista da MBL já ultrapassava mil URLs e o segundo pelo fato de Base64 não ser tratada.

Outra possível abordagem era a criação de um programa que casasse padrões (no caso URLs) em e-mails. Esta era uma alternativa viável para atingir os pré-requisitos 1 e 2. No entanto, o pré-requisito 3 possivelmente seria comprometido pelo tempo necessário para desenvolver e testar tal programa antes de colocá-lo em produção.

A solução original a que se chegou foi a utilização de um programa já consolidado e especializado no casamento de padrões de forma eficiente (pré-requisito 1), no tratamento também eficiente de e-mails em Base64 e que se mostrou bastante simples de implementar, utilizar e atualizar: o anti-vírus Clam (ClamAV) [7].

O ClamAV é um anti-virus para gateway de e-mail gratuito e com código livre bastante utilizado no mundo inteiro. O Chasque Mail nos forneceu uma boa experiência com o mesmo e, uma vez determinado que ele atendia nossos requisitos, pesquisou-se em como integrá-lo com a MBL. Encontrou-se, pela documentação sobre como fazer assinaturas de vírus para ClamAV [8], que seria necessário, apenas, criar uma assinatura para cada URL que se desejava bloquear. A criação de tais assinaturas se mostrou bastante simples, consistindo na conversão da URL, caractere a caractere, para hexadecimal. Devido a simplicidade natural de uma URL (isto é, o fato de ela ser formada por uma seqüência contínua de caracteres), a utilização de assinaturas do tipo básico do ClamAV é estritamente suficiente para cumprir com o objetivo em questão. A utilização de assinaturas do tipo estendido apenas consumiria mais recursos.

Apesar de o ClamAV se mostrar bastante apto a realizar a tarefa em questão, é importante notar que este uso é, de certa forma, um "abuso" do mesmo. O ClamAV foi idealizado para ser uma solução completa de anti-vírus de gateway de e-mail e não um simples casador de padrões para corpos de e-mails. Desta forma, existe a possibilidade de surgirem alguns efeitos colaterais na utilização do ClamAV para detectar URLs maliciosas em corpos de e-mail. Um destes é o fato de, por suas opções padrões, o ClamAV varrer não apenas o corpo da mensagem, mas todos seus anexos, incluindo arquivos compactados. Assim, se, por exemplo, uma URL maliciosa encontra-se dentro de um documento no formato Microsoft Word compactado no formato PKZIP, o ClamAV a encontrará.

Em março de 2006 começamos a testar esta solução, fazendo download da MBL em formato de texto puro e convertendo-o para assinaturas do ClamAV. Devido ao sucesso obtido, contatamos o criador da MBL, que se interessou em disponibilizar a lista neste novo formato. Em junho de 2006, após esclarecimentos sobre o funcionamento do novo formato, a lista foi posta em produção e o documento (*howto*) básico sobre como configurar o ClamAV para usar a MBL desta nova forma foi disponibilizado em [9]. Na Figura 3 tal documento é mostrado.

```
To use the Malware Block List with Clamav, one has to do the following:

1) Download the MBL in Clamav signatures format. For example:
/usr/bin/wget -O - http://www.malware.com.br/cgi/submit?action=list_clamav > mbl.db

2) Copy the file to Clamav's DatabaseDirectory. This can be found in the
clamd.conf file and usually is where the files main.cvd and daily.cvd are.
It is important that the file copied must have the extension .db and has
the right permission so that Clamav can read from it.
/bin/cp mbl.db /var/clamav

3) Reload/Restart clamav (if you are using the daemon)
/etc/init.d/clamad reload

That is it. From now on, emails that contain URLs from the MBL, when
scanned with Clamav, will be marked as infected by MBL.%NUM%.
```

Figura 3: Passos para se usar a MBL com ClamAV

5. Resultados

Na Figura 4 encontra-se um gráfico que mostra o número de e-mails bloqueados utilizando-se a MBL do período de agosto a dezembro de 2006. Neste período foram bloqueados, ao todo, mais de 93 mil e-mails apenas usando-se a MBL.

Ao nosso ver, este resultado se mostrou bastante positivo. Empiricamente, notamos imediatamente uma diminuição no número de e-mails maliciosos que nossos usuários estavam recebendo. No entanto, apesar de serem recebidos com muito menos freqüência, alguns e-mails com links maliciosos ainda chegavam às caixas postais de nossos usuários. O mais interessante era que muitos dos domínios das URLs nestes e-mails já constavam em algumas das URLs capturadas por nosso sistema. Observando estas reincidências, decidimos agir pro ativamente, bloqueando apenas estes domínios que comprovadamente hospedavam diversos malwares. Esta lista de domínios bloqueados, que é gerada a partir da análise estatística da reincidência de domínios se mostrou também eficiente, como mostra a Figura 5, referente ao período de agosto a dezembro de 2006. Neste período, foram bloqueados mais de 84 mil e-mails através desta técnica.

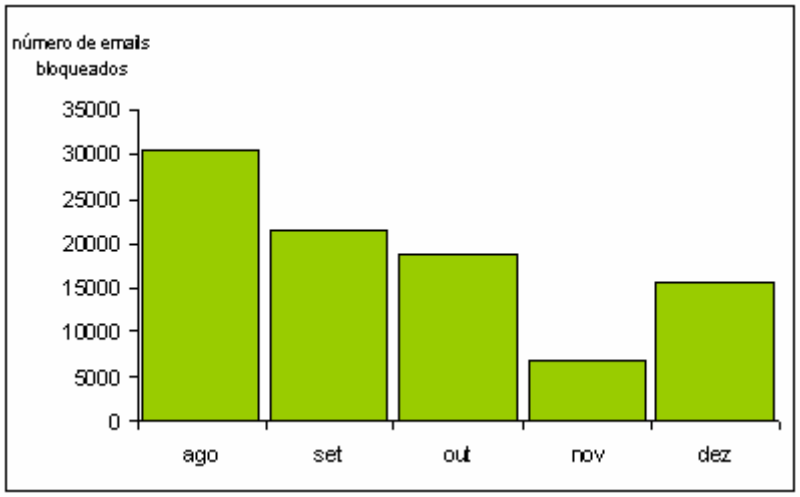

Figura 4: E-mails bloqueados utilizando a lista MBL de agosto à dezembro de 2006

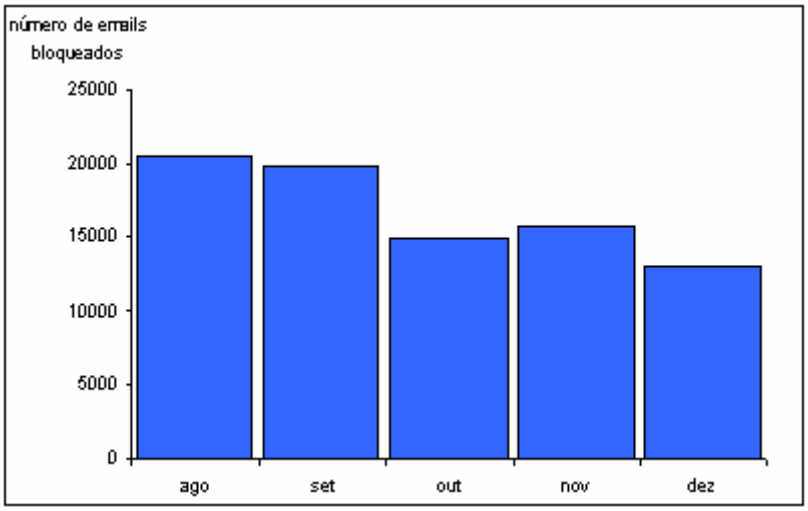

Figura 5: E-mails bloqueados utilizando a lista de DOMÍNIOS de agosto à dezembro de 2006

Ao todo, no período acima mencionado, foram bloqueados mais de 177 mil e-mails usando os mecanismos desenvolvidos e relatados neste artigo. Isto gera uma média de aproximadamente 1157 e-mails bloqueados diariamente.

Quanto a falso positivos, não temos conhecimento da ocorrência dos mesmo na MBL, visto que na versão normal por nós utilizada ela especifica de maneira extensa a URL (incluindo subdiretórios). Como boa parte dos domínios por nós bloqueados são serviços de hospedagem gratuita (que não se preocupam se hospedam ou não malware), soubemos, até hoje, de 2 falso positivos. Em geral, o sistema tem se mostrado bastante preciso.

6. Considerações Finais

A partir dos resultados acima mencionados, é possível concluir que a meta deste projeto foi alcançada: uma solução nova e eficaz para o problema de e-mails contendo links com malwares foi encontrada e implementada de maneira simples e eficiente. Com isso, nossos usuários ficaram mais

bem protegidos de maneira transparente. Pudemos, também, contribuir com a MBL e ajudar outros administradores a proteger melhor suas redes.

No entanto, como pode ser visto na Figura 4, a eficácia da solução vem diminuindo com o tempo. Isto algo esperado, como pode ser visto na batalha constante que se trava contra o spam. A maneira de burlar a MBL que os malfeitores encontraram é enviar e-mails com URLs que não apontem diretamente para os malwares e que, consequentemente, não possuam uma extensão suspeita. Assim, para se contaminar, a vítima deve acessar a URL e será redirecionada para onde o malware se encontra. Combater esta utilização de redirecionamentos diretos e indiretos em links maliciosos é uma tarefa que, em breve, mudará de importante para fundamental.

Referências

[1] SURBL - Spam URI Realtime Blocklists. Disponível em: http://www.surbl.org/. Acesso em jan/2007.
[2] Malware Block List. Disponível em: http://www.malware.com.br/. Acesso em jan/2007.
[3] ngrep - network grep. Disponível em: http://ngrep.sourceforge.net/. Acesso em jan/2007.
[4] Squid Web Proxy Cache. Disponível em: http://www.squid-cache.org/. Acesso em jan/2007.
[5] Marchi, Alexandre. Flores, Bárbara. Horowitz, Eduardo. Rey, Leandro. Tonin, Rafael. Campos, Tallitha. Chasque: o correio eletrônico da UFRGS migrando para software livre. 6° Fórum Internacional Software Livre. Porto Alegre. 2005.
[6] Postfix Built-in Content Inspection. Disponível em: http://www.postfix.org/BUILTIN_FILTER_README.html#limitation. Acesso em jan/2007.
[7] Clam AntiVirus. Disponível em: http://www.clamav.net/. Acesso em jan/2007.
[8] Creating signatures for ClamAV. Disponível em: http://www.clamav.net/doc/latest/signatures.pdf. Acesso em jan/2007.
[9] Malware Block List with ClamAV. Disponível em: http://www.malware.com.br/clamav.txt. Acesso em jan/2007.
[10] Websense Security Labs. Security Trends Report: Second Half 2005. 2005.