

Como filtrar spam ao máximo com Evolution e SpamAssassin

Author: André F Machado<andremachado@techforce.com.br>

Veja como capturar quase 100% dos spam e vírus com Evolution e SpamAssassin no Debian GNU / Linux de maneira fácil, com mínimo trabalho e ainda melhorar a captura desse lixo cada vez mais. Artigo atualizado em julho 2009 para maior velocidade de filtragem e versão mais nova do Evolution.

Escopo

Nas versões 2.x não há possibilidade de integração com o [SpamAssassin](#) a partir da versão 3.x ficou bem preciso na captura de spam.

E ainda ficaram bem mais fáceis de usar para o usuário final.

No Debian GNU / Linux essa tarefa fica bem fácil e simples.

Também poderá filtrar as mensagens contendo vírus. Embora não afetem os sistemas linux, entopem as caixas postais.

Este artigo explicará como fazer a instalação, configuração e uso da maneira mais fácil no [Debian GNU / Linux](#) para alcançar quase 100% de sucesso na filtragem de spam. Outras distribuições podem aproveitar os conceitos deste artigo, mas diferem na forma de instalar e configurar.

O aumento da eficácia vai tornar a filtragem mais lenta (praticamente não aumenta mais o uso de cpu), mas vale a pena.

Visão geral de como melhorar a filtragem antispam

O SpamAssassin 3.x vem com recursos de filtragem por regras e auto aprendizado dos filtros por algoritmos de Bayes.

Quanto mais você filtrar e indicar manualmente o lixo que escapou, mais preciso fica o filtro de Bayes.

Mesmo assim, a filtragem ainda deixa passar uma quantidade de spam e lixo.

Os spammers inventam truques novos todos os dias.

Para melhorar, é preciso usar os recursos adicionais de filtragem.

O SpamAssassin pode incorporar a filtragem por critério [Sender Policy Framework](#)

A consulta (com pontuação ponderada) a bases de dados de amostras de spam, mantidas colaborativamente pelos internautas. São algumas listas negras de bloqueio (RBL Real time Blocking List) que aceitam submissões de spam.

Por serem consultas com pontuação ponderada, e a bases mantidas colaborativamente pelos internautas eleva MUITO a taxa de captura e reduz os falsos positivos (emails legítimos confundidos com spam).

As listas negras de bloqueio (RBL)

O SpamAssassin consulta **várias** listas de bloqueio, mas algumas são bem especiais e merecem uma apresentação de suas vantagens e efeitos na taxa de captura e porque valerá a pena esperar pela filtragem através de suas pontuações.

[SpamCop](#) é uma base colaborativa até com outras listas de bloqueio. Ela mesmo faz consultas a terceiras e as informa de novos spammers. Se encontra um novo open relay, por exemplo, envia para verificação final [Accept Open Relay Data Site](#) internautas para pontuação dos spam. É uma lista bem agressiva nas pontuações. Mas muito ágil em identificar spam em poucos minutos que começa a ser disparado contra os internautas. Você precisará se inscrever gratuitamente no site deles para abrir uma conta de submissão de spam.

[SpamHaus Project](#) mantém famosas e das mais precisas listas de bloqueio. Muito baixa taxa de erros.

[Spam List Realtime Blocklists e URIBL](#)

as mais interessante listas negras no momento. O SpamAssassin consegue extrair os endereços dos sites beneficiados com o spam (que hoje em dia vêm principalmente de milhões de máquinas caseiras windows contaminadas e sempre têm endereços dinâmicos) e consultar essa base de sites beneficiados por spam.

Esse recurso é importante. Pois o spam está vindo de dezenas de milhões (literalmente) de endereços IP diferentes (as máquinas windows contaminadas), mas beneficiam uma lista definida de sites 'espertos' (os spamvertized sites). As consultas ficarão **bem** mais rápidas se instalar um cache de DNS na sua máquina, como o Bind configurado para isso.

[Vipul's Razor](#), [Pyzor](#), [Distributed Checksums](#) e [SpamAssassin](#) são bases de dados mantidas colaborativamente e de forma séria.

Os milhões de internautas ajudam a torná-las rápidas e precisas em identificar novos ataques de spam.

O que você vai precisar

Para consultar e colaborar com as RBL via SpamAssassin, você precisará de alguns módulos perl:

Mail::SPF::Query

Net::DNS

Net::SMTP

Mais alguns programas:

Razor

Pyzor

Bind9 (para cache de consultas DNS) OU djbdns configurado e usado como dns cache

DCC client

SpamAssassin

Evolution

Todos possuem pacotes prontos para o Debian GNU / Linux.

Instalando no Debian GNU / Linux

No Debian GNU / Linux é **muito** fácil instalar tudo.

Você poderá usar a interface gráfica Synaptic, ou a interface com menus em texto Aitude ou ainda a linha de comando no apt-get.

Loggado
como root faça os comandos:

```
apt-get update
apt-get install libnet-perl libnet-dns-perl libmail-spf-query-perl \
  dcc-client razor pyzor dnscache-run spamassassin evolution \
  perl clamav unrar lha arj unzoo gzip bzip2 unzip \

zip ligsys-hostname-long-perl libnet-ident-perl \

libio-socket-ssl-perl libdbi-perl libnet-ph-perl \

libnet-snpp-perl libnet-telnet-perl
```

O Apt calculará as dependências e instalará o que for necessário, verificando o que já está na sua máquina.

O SpamAssassin também estará funcional. Mas precisaremos adequar e otimizar a configuração mais adiante.

Na instalação do clamav, será perguntado quando atualizar a base de assinaturas de vírus. Em sistemas domésticos que são ligados quando se vai utilizar, você pode escolher atualizar na hora do boot.

O bind9 (servidor DNS) no Debian GNU / Linux é instalado pré-configurado como cache de requisições DNS. Portanto, não se preocupe com a configuração deste serviço servidor. Estará feita para você nos aspectos essenciais.

Atualização de julho 2009 sobre dnscache:

O bind9 é um serviço de dns completo. Como apenas o cache de DNS é necessário, uma solução melhor é instalar o pacote dnscache-run que baixará outros programas (daemontools, daemontools-run, djbdns, ucspi-tcp) e configurará tudo como um dns cache para a máquina toda. O /etc/resolv.conf será alterado para apontar o nameserver como o localhost (127.0.0.1).

As consultas às URI BL e algumas RBL ficarão mais rápidas à medida que fores baixando os emails. Geralmente, num mesmo dia, um certo IP ou uma URI aparece listada em URIBL e repetida em muitos spam.

Configurando no Debian GNU / Linux

Estes passos podem ser semelhantes para sua distribuição diferente. Confira a sua documentação específica.

Para uma explanação detalhada dos comandos abaixo, consulte a documentação dos programas nos respectivos sites ou nas man pages em seu computador.

Loggado

como usuário normal e no seu home directory execute os comandos num terminal e fique bem atento aos resultados dos comandos:

```
cd
pyzor discover
razor-admin -create
razor-admin -register
```

Otimizando a configuração

O sistema já poderia filtrar e reportar, mas ainda faltam otimizações e configurar melhor o SpamAssassin.

Na [página de documentação sobre o arquivo de configuração](#) do SpamAssassin, você encontrará todas as opções de configuração detalhadas.

O SpamAssassin procura um arquivo em sua área de usuário primeiro. Se não encontrar, utilizará as configurações gerais de sistema.

Para simplificar e reduzir riscos, vamos configurar apenas para SUA área de usuário.

Você encontrará um diretório invisível spamassassin na sua área de usuário e nele o

arquivo user_prefs .

O SpamAssassin ainda está fazendo consultas às RBL serialmente, e isso atrasa bastante o processamento das mensagens. Em conexões de alta velocidade, você poderá reduzir o tempo de espera máxima para respostas dos servidores RBL. Dos 10 segundos tempo padrão para algo como 3 segundos. Você terá de experimentar e avaliar seus resultados.

Também terá de configurar seus endereços de inscrição no SpamCop.

A partir da versão 3.1, o SpamAssassin adota como default desabilitar as consultas ao Razor e DCC (motivos de licenciamento de uso). Portanto será preciso habilitar na configuração.

Edite o arquivo ~/.spamassassin/user_prefs:

```
use_dcc      1
use_razor2   1

use_pyzor    1

#SpamAssassin > 3.1.x  need this line poiting to where is dccifd
dcc_home     /var/lib/dcc

spamcop_to_address
submit.SEU_CODIGO_SECRETO_NO_SPAMCOP@spam.spamcop.net

spamcop_from_address SEU_ENDERECO_EMAIL_INSCRITO_NO_SPAMCOP
ok_languages  en es pt

# distributed clearing house rbl test timeout in seconds
dcc_timeout  3
# pyzor rbl test timeout in seconds
pyzor_timeout 3
# razor rbl test timeout in seconds
razor_timeout 3
```

Ainda resta testar as configurações já feitas e configurar e testar o Evolution.

Testando as configurações já feitas

Use seu programa de email e salve um identificado spam num diretório temporário na sua área de usuário.

Por exemplo, ~/temp/spam1.txt.

Abra-o com algum editor (sem salvar) ou liste-o para conferir que é mesmo um spam contendo os cabeçalhos de email.

Execute o comando seguinte e fique bem atento ao resultado

```
spamassassin -D -t ~/temp/spam1.txt
```

Deverão aparecer os resultados das ponderações feitas pelo SpamAssassin.

Confira se ele está consultando as RBLs e SPF desejadas

Repare que se DCC

não tiver registrado o spam ainda ou não estiver respondendo, vai retornar o header **X-DCC em branco**, o que será visto na verborrágica saída do comando assim:

```
dbg: dcc: dccifd check failed - no X-DCC returned:
```

Para você testar se o firewall não está bloqueando as consultas e respostas do cliente dcc aos servidores, [faça conforme as instruções aqui](#)

```
cdcc info
```

Esse comando deverá retornar algo similar a:

```
andremachado@debian:~$ cdcc info
# 03/02/07 19:34:57 BRT /var/lib/dcc/map
# Re-resolve names after 21:21:26 Check RTTs after 19:37:09
# 455.30 ms threshold, 485.59 ms average 12 total, 10 working servers
IPv6 ondcl.dcc-servers.net,- RTT+0 ms anon
# ::ffff:137.208.8.26,- wuwien
ID 1290
# 100% of 4 requests ok 455.30+0 ms RTT 115 ms queue wait
# ::ffff:142.27.70.214,-
# not answering
# ::ffff:194.228.41.13,- CTc-dcc2
ID 1031
# 94% of 32 requests ok 617.23+0 ms RTT 133 ms queue
waitdcc2.dcc-servers.net,- RTT+0 ms anon
# ::ffff:80.69.8.186,- MC
ID 1128
# 50% of 10 requests ok 2136.86+0 ms RTT 100 ms queue wait
# ::ffff:136.199.199.102,- URT
ID 1060
# 50% of 2 requests ok 1722.02+0 ms RTT 102 ms queue wait
# ::ffff:192.84.137.21,- INFN-TO
ID 1233
# 17% of 6 requests ok 3367.83+0 ms RTT 100 ms queue
waitdcc3.dcc-servers.net,- RTT+0 ms anon
# ::ffff:216.134.200.215,-
```

```

ID 1113
#      50% of  2 requests ok 1587.84+0 ms RTT      101 ms queue
waitdcc4.dcc-servers.net,-      RTT+0 ms      anon
#      ::ffff:194.228.41.73,-      CTc-dcc1
ID 1030
#      60% of  5 requests ok 1010.91+0 ms RTT      102 ms queue wait
#      ::ffff:209.169.14.30,-
ID 104
#      50% of  4 requests ok 2073.80+0 ms RTT      101 ms queue
waitdcc5.dcc-servers.net,-      RTT+0 ms      anon
# *  ::ffff:67.66.138.141,-
ID 1356
#      90% of 10 requests ok  361.09+0 ms RTT      101 ms queue wait
#      ::ffff:71.246.8.99,-      Misty
ID 1170
#      100% of 3 requests ok  656.80+0 ms RTT      304 ms queue
wait127.0.0.1,-      RTT-1000 ms  32768
#      ::ffff:127.0.0.1,-
#      not answering#####
# 03/02/07 19:34:57 BRT GreyList /var/lib/dcc/map
# Re-resolve names after 21:34:57
# 1 total, 0 working servers
# skipping asking Greylist server 16 seconds more127.0.0.1,-
#      Greylist 32768
#      ::ffff:127.0.0.1,-
#      not answering

```

Repare que alguns servidores respondem e outros não, eventualmente.

Mas se

TODOS não estiverem respondendo, é uma forte indicação de que o firewall está bloqueando a porta e ou protocolo necessários.

Analisando a saída de depuração do comando do SpamAssassin, também pode ser que o servidor do pyzor esteja sobrecarregado ou fora do ar.

```
dbg: pyzor: got response: 66.250.40.33:24441 TimeoutError:
```

Você poderá incluir manualmente, EM PRIMEIRO LUGAR (o SpamAssassin só usa o primeiro listado), no arquivo ~/.pyzor/servers um servidor não oficial alternativo como solução temporária (o servidor central tem um banco de spams muito maior)

```
82.94.255.100:24441
```

Teste se recebe resposta:

```
pyzor ping
```

Só que na próxima vez que rodar pyzor discover, irá sobrescrever a informação.

Você também pode simplesmente deixar como está e aguardar que o servidor central volte a responder normalmente.

Agora verifique se poderá reportar spam às bases colaborativas.

Execute o comando seguinte e fique bem atento ao resultado e eventuais indicações de erros

```
spamassassin -r --debug ~/temp/spam1.txt
```

Configurando o Evolution



O Evolution 2.x já identificará a instalação do SpamAssassin.

Os spam que escaparem da filtragem poderão ser marcados como "Indesejada" usando o botão na barra.

Os emails legítimos que acabarem marcados como indesejados poderão ser remarcados usando o botão "Não é indesejada".

Essas correções serão incluídas na base de treinamento bayesiana.

As mensagens indesejadas vão para uma pasta "Indesejada".

Confira-a regularmente em busca de eventuais emails legítimos incorretamente classificados e remarque-os como descrito acima.

Configurando os filtros no Evolution 2.22.x

O programa já tem uma pasta Spam para armazenar as mensagens indesejadas. Falta você configurar

Editar > Preferências > Preferências de correio > aba Spam

(marcar) Procurar em mensagens recebidas por spam.

(marcar) Verificar cabeçalhos personalizados por spam

(adicionar):

```
X-Spam-Flag      Yes
X-Spam-Level     *****
```

Plugin padrão de spam: SpamAssassin

Opções de SpamAssassin

(Marcar) Incluir testes remotos

Pronto!

Outras opções da janela podem ser convenientes ou não a sua instalação. Avalie.

Configurando os filtros no Evolution 2.6.x

Crie uma pasta "virus" na árvore do computador local.

Configure o filtro antivírus para mover a uma pasta "virus" que você criou antes no Evolution (ou apague a mensagem viral)

```
Ferramentas > Filtros > Adicionar
Nome da regra: antivirus
Se enviar (pipe) para programa : clamscan -
  E não retorna 0 (zero)
Então mover para pasta: virus
```

Configure o relatório de spam às bases de dados colaborativas:

```
Ferramentas > Filtros > Adicionar
Nome da regra: report_spam
Se teste de indesejada: mensagem é indesejada
Então enviar (pipe) para programa : spamassassin -r
```

Pronto.

A filtragem ficará mais lenta, porém **muito** mais eficaz.

Tente ajustar os tempos de time-out para sua conexão.

Espero que no futuro o SpamAssassin faça consultas em paralelo às RBL pyzor, dcc e razor, como já faz com as outras, para evitar esse somatório dos tempos.

Eu mesmo já tenho um perl script para consultas em paralelo a RBLs aqui neste site.

Portanto, é possível fazer.

Otimizando desempenho

Você pode escolher filtrar spam ANTES de filtrar vírus, movendo a regra para cima na lista de filtros. O impacto no desempenho depende da proporção da quantidade de mensagens de vírus que recebes contra a quantidade de spam.

Também pode ajustar os tempos de time out dos filtros de consulta a bases de dados colaborativas.

Outra medida é mudar os seguintes parâmetros de ~/.spamassassin/user_prefs para

```
bayes_learn_to_journal 1
```

```
bayes_learn_during_reporting 1
```

Confirmando no SpamCop

Você precisa ~~que você confirme seus relatórios~~ de spam no site [SpamCop](#) seu login de conta registrada previamente.

Você pode se interessar pelo programa SpamCup.

É

importante confirmar os relatórios no SpamCop pois esta RBL **repassa** informações para várias outras RBLs especializadas, que fazem testes adicionais e ele consulta depois. Por exemplo, a citada SURBL só aceita submissões através do repasse pelo SpamCop, para minimizar relatos falsos. Também a Open Relay Data Base recebe dados para verificações e testes extensivos repassados pelo SpamCop.

O SpamCop também envia relatos para o [CERT BR](#)

Se você não se registrar no SpamCop e não configurar o SpamAssassin para enviar para seu endereço de código secreto, ele irá enviar para uma conta "genérica" de muito baixa prioridade para recebimento de envios automatizados do SpamAssassin.

O efeito é que serão necessárias muitas e muitas denúncias de um mesmo spam para ele ser colocado na lista negra por 48 horas. Já se você confirmar manualmente o spam, isso terá um alto peso e bastam algumas denúncias para ele ir parar na lista de bloqueio.

Medidas drásticas

Com algumas semanas de captura de mensagens, e treinamento das que escaparam usando o botão "indesejada", aliados às RBL e URIBL colaborativas, a taxa de captura de spam supera 99% e os falsos positivos serão muito raros.

Porém, dependendo do volume de mensagens que você recebe, mesmo esse 1% pode ser ainda muito spam escapando.

Então é hora de medidas drásticas.

As configurações extras não serão triviais. Mas se você chegou até esse ponto, estará disposto a um certo trabalho e estudo adicional.

Você terá de conhecer o [SpamAssassin Rules Emporium](#) **analisar** quais regras extras podem ser úteis para seu perfil.

Certamente você ~~que está utilizando diariamente suas~~ [Rules du Jour](#) regras novas.

Se ~~ainda~~ não for suficiente, você pode querer conhecer as regras da [SA-Black list](#) [Bill Stearn](#)
Há ~~algumas dicas de configuração aqui.~~

Atenção que estas regras não oficiais podem elevar seus falsos positivos.

As regras oficiais do SpamAssassin são mais conservadoras e minimizam os falsos positivos.