ANÁLISE DE FERRAMENTAS PARA O CONTROLE DE SPAM

Paulo Manoel Mafra

Departamento de Automação e Sistemas

Universidade Federal de Santa Catarina

88040-900 Florianópolis - SC

mafra@das.ufsc.br

Tópicos

- Algumas Considerações sobre SPAM
- Abordagens Possíveis para Combater SPAM
- Algumas Técnicas Utilizadas
  - Tipos de Bloqueio
- Experiências e Testes
- Problemas Enfrentados e Soluções Adotadas
- Conclusões

Algumas Considerações sobre SPAM

- O que é um SPAM ?
  - E-mail indesejado, contendo propaganda não solicitada
- Por que pessoas odeiam SPAM ?
  - Porque eles enchem a caixa postal dos usuários com informações inúteis
  - Porque seu conteúdo é de péssima qualidade
  - Porque eles consomem banda e sobrecarregam servidores de e-mail

Abordagens Possíveis para Combater SPAM

- Legislação apropriada
  - Problema: Cada país tem a sua legislação
- Taxação para o envio de e-mails
  - Problema: Em qual moeda seria pago, para quem, isso resolveria o problema ?
- Uso de diversas técnicas nos servidores de e-mail

Dificuldades Encontradas para Combater o SPAM

- Os cabeçalhos das mensagens geralmente são falsos
- Uso de endereços IP dinâmicos (xDSL, dial-up)
- Muitas redes não tomam atitudes para prevenir e combater o SPAM gerado pelos seus usuários. Essas redes possuem usuários que utilizam o serviço de e-mail corretamente e usuários que utilizam para enviar SPAM

Algumas Técnicas Utilizadas

- Técnicas de bloqueio:
  - Técnicas de bloqueio por cabeçalho da mensagem
  - Uso de greylisting
  - Técnicas de bloqueio por conteúdo da mensagem

Técnicas de Bloqueio por Cabeçalho

- São técnicas implementadas geralmente no MTA (Mail Transport Agent)
- Algumas dessas técnicas podem ser implementadas individualmente no cliente de e-mail
- São compostas por:
  - Verificação da existência do domínio informado
  - Verificação da existência de um nome para o endereço IP do emissor
  - Listas externas - RBLs
  - Listas internas de acesso

Verificação do Domínio Informado

- Verifica se o domínio do remetente existe
- Pode rejeitar e-mails de servidores mal configurados
- Bloqueia mensagens com cabeçalho falso (que tenham um domínio inexistente)

Verificação de Nome para o Endereço IP do Emissor

- Verifica se existe um nome para o endereço IP que está enviando a mensagem
- Verifica se a partir do nome chega-se ao endereço IP
- Bloqueia muito SPAM
- Bloqueia e-mails legítimos

Listas Externas - RBLs

- Serviço mantido por terceiros
- Não permite um bloqueio de endereços individuais, bloqueia por endereço IP do servidor
- Não garante que todas as mensagens bloqueadas sejam SPAM
- Poucos provedores respondem por queixas de abusos

RBL - Realtime Blackhole List

`http://www.mail-abuse.com/`

Listas Internas de Acesso

- São listas compostas por endereços IP, por CIDRs, por domínios ou por endereços específicos de e-mail
- Não utilizam nenhuma heurística ou método de análise estatística
- Não existe o problema do falso positivo (pelo menos em teoria)
- Essas listas são classificadas em:
  - Whitelist: mensagens oriundas de endereços contidos em uma whitelist são aceitas
  - Blacklist: mensagens oriundas de endereços contidos em uma blacklist são rejeitadas

Listas Internas de Acesso

- Possuem um gerenciamento manual ou automático, a partir de serviços externos
- Pode-se criar listas individuais onde o usuário somente receberá e-mails de pessoas cadastradas nessas listas. Por exemplo, se João envia uma mensagem para Pedro e o endereço de João não está cadastrado, é necessário acessar um *link* e se cadastrar para ter sua mensagem aceita. Isso pode gerar problemas (nem todos os usuários aceitam)

Listas Internas de Acesso

- Problema: estas listas não são dinâmicas, precisam ser gerenciadas
  - O tamanho dessas listas sempre cresce ?
  - Quem irá gerenciar estas listas ?
- Possível solução: uso de *greylisting*

Técnicas de Bloqueio por Cabeçalho

- Bloqueio em nível de servidor
  - É uma implementação que bloqueia a maior parte dos SPAMs, porém não pode bloquear e-mails legí timos
  - Deve ser genérico, não pode seguir características de apenas um grupo de usuários
- Bloqueio em nível de usuário
  - É uma implementação voltada às características do usuário

Greylisting

- Uso de whitelist e blacklist com manutenção automática
- Nem o administrador da rede nem os usuários precisam inserir os endereços IPs nas listas
- Reduz a carga no MTA
- Ajuda na filtragem de vírus

Greylisting

- O filtro baseia-se no comportamento diferente dos servidores emissores de SPAM e servidores de e-mails tradicionais
  - Servidores de SPAM enviam somente uma vez a mensagem, se ocorreu um erro descartam. Se não descartassem, formariam filas gigantescas nesses servidores
  - Servidores tradicionais seguem a RFC 821 e reenviam a mensagem em caso de falha transiente (código de resposta 45x)

Greylisting - Spamd

- Sistema simples, baseado em triplas contendo:
  - Endereço IP do servidor de e-mail que está enviando a mensagem
  - Endereço do emissor do e-mail (*from*)
  - Endereço do receptor do e-mail (*to*)
  - Tempo *tc* de criação do registro
  - Tempo *tr* para começar a receber o e-mail (30 min após o tempo de criação do registro)
  - Tempo *te* de expiração da tripla (4 horas após o tempo de criação do registro)

GREY:10.0.0.1:<spam@spam.com>:<usuario@dominio>:tc:tr:te

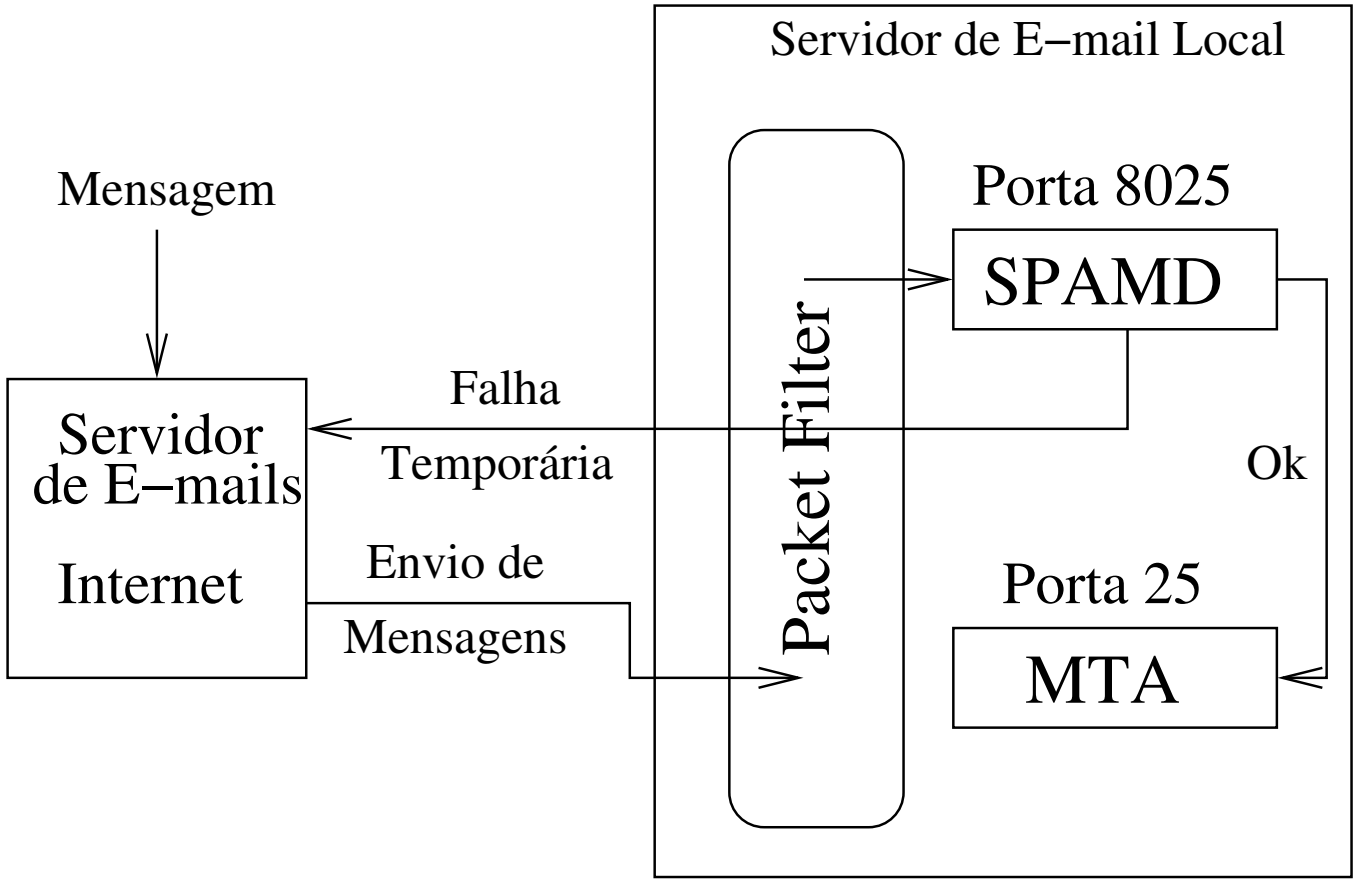
Greylisting - Spamd

- Faz-se necessário o uso do filtro de pacotes (*Packet Filter*) para redirecionar os pacotes enviados à porta 25 para uma outra porta (8025) onde roda o filtro
- Ao receber o e-mail o filtro verifica se o endereço IP do servidor emissor está na *whitelist*, se está, o e-mail é repassado para o MTA que está rodando na porta 25
- Verifica se o endereço IP do servidor emissor está em uma *blacklist*, se está, rejeita o e-mail

Greylisting - Spamd

- Verifica se a tripla existe
  1. Se a tripla não existe, cria um novo registro e retorna uma falha temporária ao servidor emissor
  2. Se a tripla existe e o seu tempo *tr* para começar a receber não expirou, retorna uma falha temporária ao servidor emissor
  3. Se a tripla existe e o tempo *tr* expirou, então adiciona o endereço IP do servidor na *whitelist* e atribui um tempo de 36 dias para expiração. A tripla fica assim: *WHITE:10.0.0.1:::tc:tr:te*

Greylisting - Spamd



Greylisting - Problemas

- Usuários impacientes podem ficar insatisfeitos
- As mensagens enviadas por servidores que não estão na *whitelist* levam algum tempo para chegar
- Alguns portais tipo *hotmail* por exemplo nem sempre reenviam a mensagem pelo mesmo servidor, fazendo com que a mensagem demore muito para chegar
- O uso de múltiplos MXs pode causar problema se cada servidor tiver as suas próprias listas. Corre-se o risco do mail ser barrado temporariamente múltiplas vezes, uma para cada MX

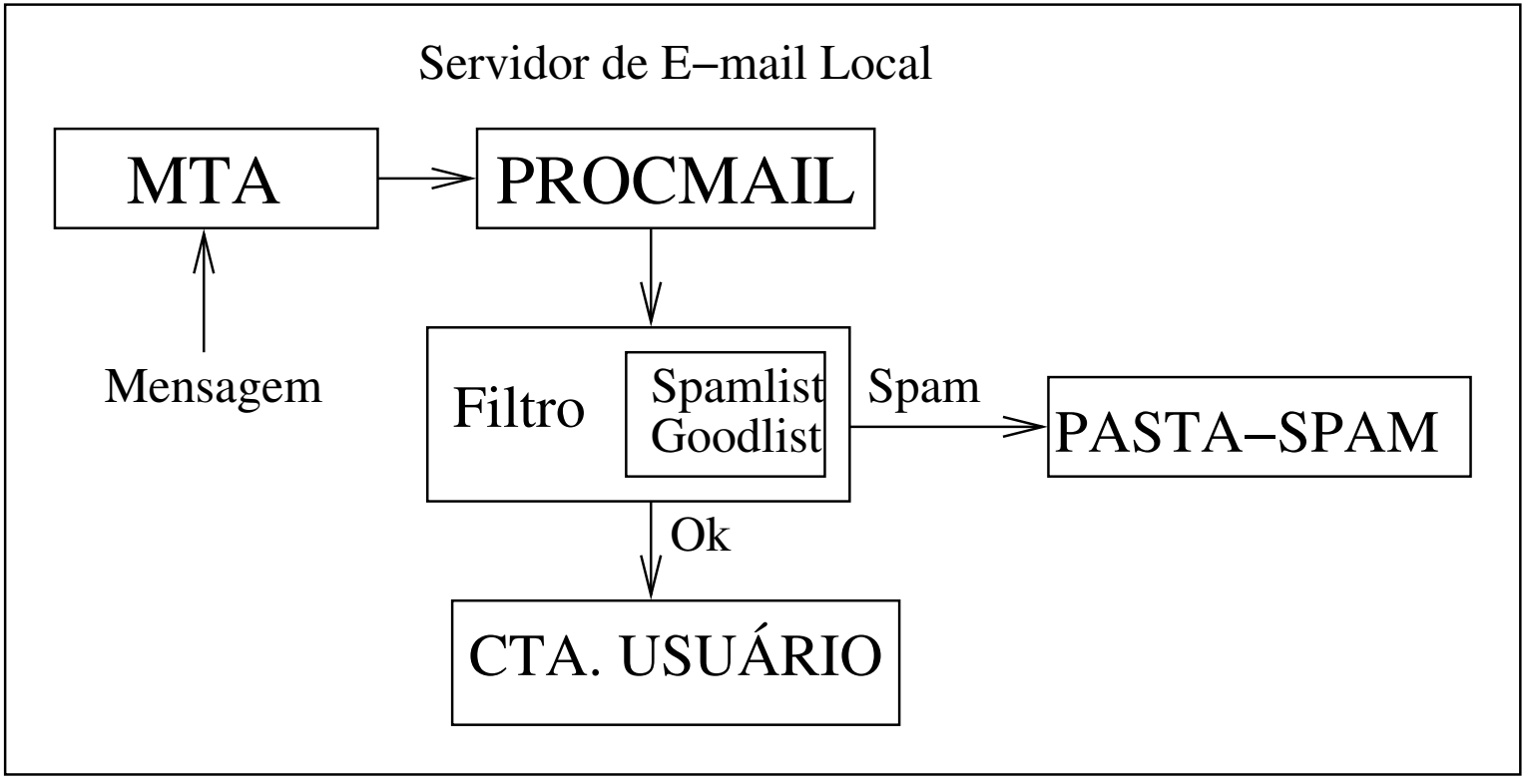
Técnicas de Bloqueio por Conteúdo

- Os filtros por conteúdo classificam as mensagens através de uma análise do seu conteúdo
- O conteúdo da mensagem é comparado com duas bases de dados que armazenam palavras consideradas boas(*goodlist*) e ruíns (*spamlist*)
- As palavras tem um peso. O peso das palavras é atribuído durante o “treinamento” do filtro. Cada vez que o filtro é treinado o peso de cada palavra é modificado considerando o seu peso atual e o número de vezes que a palavra apareceu no treinamento (a frequência da palavra)

Técnicas de Bloqueio por Conteúdo

- Compara-se as palavras contidas na mensagem com essas duas bases (*goodlist e spamlist*) e atribui-se uma pontuação à mensagem
- Mensagens que receberam uma pontuação maior que uma taxa estipulada são classificadas como SPAM

Exemplo de Implementação



Técnicas de Bloqueio por Conteúdo

- Exemplos de filtros:
  - *Bayesian Mail Filter*

    `http://sourceforge.net/projects/bmf`
  - *Bogofilter*

    `http://bogofilter.sourceforge.net/`
  - *Quick Spam Filter*

    `http://www.ivarch.com/programs/qsf.shtml`
  - *SpamAssassin*

    `http://www.spamassassin.org/`

Problemas dos Filtros por Conteúdo

- Barram muitas mensagens legítimas (falso positivo)
- Alguns SPAMs continuam incomodando (falso negativo)
- Os SPAMs utilizam sequências de palavras sem sentido para confundir o filtro
- Precisam ser treinados
- Funcionam melhor individualmente, porém cada usuário precisa treiná-lo

Experiências e Testes

- Foi utilizado o sistema operacional *OpenBSD* versão 3.5
- Como MTA foi utilizado o *Postfix* versão 2.0.19
- Para o filtro Greylisting utilizou-se o *Spamd*, este faz parte do sistema *OpenBSD*
- Esses filtros podem ser implementados em qualquer sistema UNIX
- Cada tipo de filtro foi testado individualmente
- Criou-se uma configuração para utilizar os filtros em conjunto

Testes - Filtros por Conteúdo

- Os filtros foram treinados com 6374 mensagens. Foram aplicadas aos filtros 700 mensagens produzindo os resultados mostrados na tabela abaixo:

tipos de filtros	falso positivo	falso negativo	acertos
BMF	2	35	663
Bogofilter	1	89	610
Quick Spam	21	102	577
SpamAssassin	2	110	588

Testes - Filtros por Conteúdo

- O *BMF* é um filtro rápido e apresentou resultados muito bons
- O *Bogofilter* é semelhante ao *BMF* em termos de velocidade mas não apresentou resultados tão bons
- O *Quick Spam* foi o mais rápido porém errou muito na classificação dos e-mails
- O filtro *Spamassassin* é computacionalmente muito pesado (utiliza uma vasta gama de testes de heurísticas) e não teve um alto índice de acerto
- Optou-se por utilizar o *BMF* como filtro auxiliar individual (opcional)

Testes Individuais

- Com o objetivo de verificar a eficiência dos filtros selecionados, efetuaram-se testes com cada um dos filtros individualmente. A tabela abaixo apresenta os resultados obtidos:

filtro	Mensagens filtradas	Falso pos.	Falso neg.	Índice acerto
Spamd	258531	0	90763	64.89%
Regras MTA	31316	8	2893	90.73%
BMF	700	2	35	94.71%

Testes em Conjunto

- Para os testes em conjunto, além dos filtros anti-spam, utilizou-se o filtro anti-vírus *Clamav*. A tabela abaixo apresenta os resultados obtidos da análise dos logs gerados durante oito dias consecutivos:

filtro	Mensagens recebidas	Mensagens entregues	Mensagens bloqueadas
Spamd	258531	136580	121951
Regras MTA	136580	46428	90152
Clamav	46428	45803	625
BMF	45803	44802	1001

Testes em Conjunto

- O índice de acerto dos filtros em conjunto foi de 99.61%
- Optou-se por uma configuração padrão que visa não perder mensagens autênticas (greylisting + regras de filtragem básicas no MTA + anti-vírus)
- Para usuários que desejam uma filtragem maior, pode-se instalar o *BMF* e aumentar o nível de filtragem para as regras no MTA

Problemas Enfrentados e Soluções Adotadas

- Alguns usuários deixaram de receber e-mails por causa dos filtros mas ninguém ficou sabendo
  - Solução: por causa disso, desenvolveu-se *shell scripts* para geração de relatórios diários dos e-mails que foram rejeitados por cada filtro, por usuário
  - Lição que se tira: filtros silenciosos são um convite a problemas
- Verificação de DNS reverso causa muitos falsos positivos
  - Solução: inclusão de determinados servidores em uma *whitelist*, mediante solicitação do usuário

Problemas Enfrentados e Soluções Adotadas

- Uma *blacklist* externa listou um servidor importante
  - Solução: desativar o uso daquela *blacklist*. Utilizávamos 6 *blacklists* externas e hoje utilizamos somente duas
- O intervalo de tempo (tr e te) do Spamd (30 minutos e 4 horas) não era suficiente para alguns servidores reenviarem a mensagem
  - Solução: ajustou-se o Spamd para utilizar tempos de (tr e te): 18 minutos e 26 horas

Conclusões

- É impossível bloquear 100% dos SPAMs
- O uso dos filtros em conjunto melhora bastante o índice de acerto do sistema
- Podemos dividir os usuários em três grupos: os que não se importam com SPAM, os que odeiam SPAM e os que não gostam mas não se importam em receber alguns SPAMs

Dúvidas?

Obrigado!