Métodos Avançados para
Controle de Spam

*Recímero César Fabr*e

Trabalho Final de Mestrado Profissional

Instituto de Computação
Universidade Estadual de Campinas

Métodos Avançados para Controle de Spam

Recímero César Fabre

Fevereiro de 2005

Banca Examinadora:

- Prof. Dr. Paulo Lício de Geus (Orientador)
- Prof. Dr. Pelópidas Cypriano de Oliveira
  Instituto de Artes, UNESP
- Prof. Dr. Siome Klein Goldenstein
  Instituto de Computação, UNICAMP
- Prof. Dr. Rodolfo Jardim de Azevedo (Suplente)
  Instituto de Computação, UNICAMP

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP

Fabre, Recímero César

F114m Métodos avançados para controle de Spam / Recímero César Fabre -- Campinas, [S.P. : s.n.], 2005.

Orientador : Paulo Lício de Geus

Trabalho final (mestrado profissional) - Universidade Estadual de Campinas, Instituto de Computação.

1.Sistemas de segurança. 2. Redes de computadores. 3. Mensagens eletrônicas não solicitadas. I. Geus, Paulo Lício de. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Métodos Avançados para Controle de Spam

Este exemplar corresponde à redação final do Trabalho Final devidamente corrigida e defendida por Recímero César Fabre e aprovada pela Banca Examinadora.

Campinas, fevereiro de 2005

Prof. Dr. Paulo Lício de Geus (Orientador)

Trabalho Final apresentado ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Computação na Área de Redes de Computadores.

Resumo

As tecnologias tradicionais para filtragem de *spam* apresentam algumas limitações que dificultam a precisão na classificação das mensagens, que neste trabalho são denominadas de “falso-negativos” e “falso-positivos”. Em geral, as taxas de falso-positivos são mais graves do que as taxas de falso-negativos, ou seja, visualizar um *spam* é melhor do que não ver uma mensagem importante.

Portanto, um grande número de ferramentas *anti-spam* têm surgido rapidamente para minimizar a ocorrência de falso-positivos e os problemas ocasionados pelo recebimento de mensagens não solicitadas. Este trabalho tem o objetivo de estudar técnicas variadas no combate ao *spam*, em especial, os filtros *Bayesianos anti-spam.* Os resultados obtidos apontam os filtros *Bayesianos anti-spam* como uma excelente solução para controlar o recebimento de *spam*.

Abstract

Traditional technologies to filter out spam present some limitations that impact the accuracy in the message classification process, that is, the rates of false-negatives and false-positives. As a rule, false-positives are much worse than false-negatives; in other words, visualizing a spam is better than missing an important message.

Therefore, a large number of anti-spam tools have arisen lately to try and minimize false-positives rates while reducing the high unsolicited mail volume. The goal of this work is to study a variety of anti-spam techniques, especially the Bayesian filters. The results obtained indicate that Bayesian anti-spam filters are an excellent solution to control spam reception.

Dedico aos meus pais, aos meus grandes amigos e aos professores que me incentivaram, sem os quais este trabalho não teria sido realizado.

Agradecimentos

A Deus, pelo dom precioso da vida.

Aos meus pais, que me deram amor, carinho e dedicação. Abriram as portas do meu futuro, iluminando o meu caminho com a luz mais brilhante que puderam encontrar: o Estudo. Trabalharam dobrado, sacrificando seus sonhos em favor dos meus. Não foram apenas pais, mas amigos e companheiros, principalmente nas horas em que meus ideais pareciam distantes e inatingíveis e o estudo um fardo pesado demais. Mesmo nos momentos de cansaço e de preocupações, sempre estiveram presentes me incentivando a prosseguir.

Ao professor Paulo Lício de Geus, por ter aceito orientar esta monografia, pelo apoio dado e por tudo que me ensinou.

Aos professores Eder Ricardo Biasoli, Jânio Akamatsu, Paulo Yamamura, Pelópidas Cypriano e Vagner José Oliva que me incentivaram e estiveram comigo durante esta caminhada.

Agradecimento especial à professora Mônica de Moraes Oliveira pela sua preciosa colaboração durante a redação final do trabalho.

A todos os professores e amigos, que contribuíram para eu alcançar o meu objetivo, muito agradeço.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Na década de 60, no auge da guerra fria, os Estados Unidos, atendendo a uma necessidade de seu Departamento de Defesa, começaram a investir no desenvolvimento de uma rede de dados que fosse descentralizada, permitindo o fluxo de dados em qualquer sentido e que mantivesse comunicáveis os pontos estratégicos caso houvesse destruição de algumas máquinas.

Em outubro de 1969, uma mensagem foi enviada de um computador no laboratório da UCLA (Universidade da Califórnia em Los Angeles) para outro, localizado no SRI (Instituto de Pesquisas Stanford). Naquele momento, a rede física era composta por cinco máquinas localizadas em lugares distintos: Bolt, Beranek and Newman Inc. (BBN), Instituto de Pesquisas Stanford, Universidade da Califórnia em Los Angeles, Universidade da Califórnia em Santa Bárbara e Universidade de Utah. A essa rede foi atribuído o nome de ARPANET (*Advanced Research Projects Agency Network*, Rede de Pesquisa Avançada do Departamento de Defesa dos Estados Unidos) **[1]**.

Nos anos 80, a ARPANET, que tinha fortes vínculos com Universidades e Institutos de Pesquisa, foi desmembrada em duas, resultando a MILNET, destinada a uso Militar, e a própria ARPANET, reservada à comunidade acadêmica. Essas duas redes, porém, precisavam se comunicar e, interligadas por um protocolo denominado TCP/IP (**TCP:** *Transmission Control Protocol*, Protocolo para Controle de Transmissão **/ IP:** *Internet Protocol*, Protocolo da Internet), formaram um conjunto que recebeu o nome de INTERNET (rede mundial de computadores interconectados) **[2]**.

Os benefícios que poderiam advir dessa nova forma de comunicação, aliados aos avanços tecnológicos que impulsionaram a produção de computadores pessoais, despertaram o interesse comercial pela Internet, que foi tornando-se cada vez mais popular após o surgimento dos primeiros provedores comerciais nos Estados Unidos, na década de 1990.

No Brasil, os primeiros embriões de rede surgiram em 1988, ligando universidades, institutos de pesquisa do Rio de Janeiro, São Paulo e Porto Alegre a instituições nos Estados Unidos. A RNP (Rede Nacional de Ensino e Pesquisa)[1] surgiu em 1989 para unir essas redes embrionárias e formar um *backbone* (espinha dorsal de uma rede) de alcance nacional. Em abril de 1995, o governo resolveu abri-lo, fornecendo conectividade a provedores de acesso comerciais. A partir de 1997, iniciou-se uma nova fase com o investimento em tecnologias de redes avançadas.

Hoje, mais de 800 milhões de pessoas utilizam a Internet em todo o mundo. No Brasil, o número de usuários com acesso à rede é de cerca de 19 milhões de pessoas[2].

A enorme utilização da Internet nas últimas décadas trouxe consigo benefícios provocados pela popularização das novas tecnologias de informática, mas surgiram novos problemas motivados por esses serviços.

A evolução dos microcomputadores e das redes de computadores criou uma nova forma de comunicação entre as pessoas: o serviço de correio eletrônico, também conhecido como e-mail. Ele permite a troca de mensagens entre os usuários, enviando texto, imagens e outros tipos de arquivos que podem ser anexados à mensagem. Entretanto, todas essas vantagens do e-mail têm um preço: estamos sujeitos a receber qualquer tipo de correspondência eletrônica em nossas caixas postais, particularmente mensagens publicitárias enviadas sistematicamente para uma lista de destinatários que não as solicitaram. É isso que chamamos de s*pam*.

Diversas fontes atribuem a origem do termo *spam* a um quadro do grupo humorístico britânico *Monty Python* em que um casal vai a um restaurante onde todos os pratos vêm com *spam* (uma marca americana de carne enlatada da empresa Hormel Foods). A mulher não gosta do alimento, mas não consegue nenhuma opção sem ele. Ao longo do diálogo, a palavra é repetida insistentemente pelos protagonistas, principalmente por um grupo de vikings presentes no local, que começa a cantar: "*Spam spam spam spam. Lovely spam! Wonderful spam!*" **[3, 4]**.

---

1. A internet brasileira teve início por iniciativa da RNP (Rede Nacional de Ensino e Pesquisa), conforme a seguinte referência: http://www.rnp.br/noticias/imprensa/2002/not-imp-marco2002.html
2. Os dados encontrados foram retirados do site Internet World Stats: http://www.internetworldstats.com/

Neste trabalho, considera-se *spam* como o recebimento de mensagens não solicitadas, enviadas através de ferramentas automatizadas ou não para uma ou mais pessoas sem que estas tenham solicitado as informações contidas no mesmo.

A primeira mensagem não solicitada enviada por e-mail de que se tem notícia foi um anúncio da DEC (*Digital Equipment Corporation*), fabricante de computadores, que falava sobre a nova máquina DECSYSTEM-20, em 1978. A mensagem enviada na ARPANET apresentava detalhes sobre o novo produto e convidava pessoas para apresentações na Califórnia **[5]**.

No entanto, o termo *spam* começou a ser realmente difundido a partir de março de 1994, quando Laurence Canter e Martha Siegel, dois advogados da cidade norte-americana de Phoenix, que trabalhavam em casos de imigração, enviaram uma mensagem anunciando serviços que teoricamente ajudavam as pessoas a ganharem vistos de permanência (*Green Card*) nos Estados Unidos. Por causa disso, a mensagem é conhecida como *Green Card Spam* e, já na época, imediatamente gerou as mesmas reações do *spam* atual, com questionamentos sobre ética e legalidade da prática. Não era uma mensagem nova, mas no dia 12 de abril eles usaram uma tática inovadora: contrataram um programador para criar um *script* simples e enviar o anúncio da dupla para todos os milhares de grupos de notícias da USENET (grupos de discussão), o esquema deu certo e todos receberam o primeiro *spam* em larga escala da história, o que contribuiu para difundir o termo **[6]**.

Dentre os tipos mais comuns de *spam* podem ser destacados: propaganda de produtos e serviços, pedido de doações para obras assistenciais, correntes da sorte, propostas de ganho de dinheiro fácil e boatos desacreditando o serviço prestado por determinada empresa.

Em abril de 2004, o tráfego das mensagens não solicitadas chegou a 64% do total de mensagens que circulam na Internet[1], ou seja, todo este lixo é acumulado nas caixas postais, comprometendo o desempenho dos servidores de correio eletrônico e da rede, além de fazer com que boa parte do nosso horário de trabalho seja destinado a limpar nossa caixa postal.

Estima-se que, no ano de 2002, o custo do *spam* foi aproximadamente de $8,9 bilhões de dólares para as empresas nos Estados Unidos e $2,5 bilhões na Europa. No ano

---

1. Conforme dados da empresa Brightmail Anti-Spam as mensagens não solicitadas estão crescendo cada vez mais na Internet. Maiores detalhes no endereço: http://www.brightmail.com/

seguinte, essa proporção ultrapassou os $10 bilhões de dólares para as empresas nos Estados Unidos **[7]**.

Com relação à prática abusiva do *spam*, muito tem se falado sobre soluções de combate a ele e muita controvérsia tem aparecido em eventos técnicos. No entanto, uma conclusão ponderada é a de que não existe uma receita capaz de eliminar o *spam* da Internet, muito menos uma solução única que resolva o problema dos administradores e usuários. Mas, existem algumas alternativas capazes de reduzir o impacto causado pelo *spam*. A mais usada atualmente é a filtragem e, neste caso, já se tem uma variedade de filtros disponíveis em diversos níveis.

Os filtros, como o próprio nome diz, permitem fazer uma triagem nos e-mails recebidos, separando os *spams* dos e-mails válidos. As duas principais técnicas de filtragem de *spam* são: listas de bloqueio / permissão e classificação de conteúdo. A lista de bloqueio, também conhecida como lista-negra (*blacklist*), e a lista de permissão lista-branca (*whitelist*), analisam o cabeçalho das mensagens recebidas, endereços IP, domínios ou endereços de e-mail que devem ser bloqueados ou permitidos. Já a classificação de conteúdo usa uma abordagem diferente. Ao invés de analisar apenas o cabeçalho procurando identificar remetentes suspeitos, ela analisa todo o conteúdo da mensagem (isto é, o texto completo) em busca de padrões suspeitos e, com base na identificação de determinados padrões, utiliza estatística e probabilidade para fazer uma classificação do que é ou não *spam* **[8]**.

1.1 Objetivos da dissertação

Este trabalho tem como objetivos apresentar os problemas ocasionados pelo recebimento de mensagens não solicitadas e pesquisar técnicas variadas no combate ao *spam*, em especial, os filtros *Bayesianos anti-spam* atuando tanto no MTA (*Mail Transfer Agent*, Agente de Transferência de Mensagens) local quanto no MUA (*Mail User Agent*, Agente de Mensagens do Usuário).

1.2 Trabalhos Correlatos

Existem vários trabalhos sobre técnicas variadas no combate ao *spam*, devido aos sérios problemas causados para os usuários de e-mail. Para reduzir ao máximo o recebimento de e-mails não solicitados, pesquisadores têm trabalhado bastante em soluções *anti-spam* nos últimos anos **[9, 10, 11, 12]**.

Entretanto, um trabalho correlato que merece destaque é o projeto de Jinpeng Wei, do Instituto de Tecnologia da Geórgia, que pesquisa técnicas[1] *anti-spam* e aproximações entre filtros de *spam* e *Honeypots* de *spam* (sistema específico para capturar ações de *spammers*). Segundo Wei, os usuários de correio eletrônico têm usado cada vez mais filtros de *spam* para classificar as mensagens como *spam* ou *não-spam*.

Com relação às técnicas de classificação, Wei observa uma infinidade de ferramentas *anti-spam* que estão sendo criadas, mas em geral elas podem ser divididas em três categorias: filtros *Bayesianos*, filtros baseados em Heurística e um Sistema Distribuído para Calcular *Checksum* (DCC).

Os filtros *Bayesianos* são ferramentas baseadas no método *Naive Bayes*. Basicamente, a idéia é colecionar um grande número de palavras consideradas *spam* e uma coleção de palavras *não-spam*. Portanto, quando uma mensagem é classificada, o filtro utiliza estatística e probabilidade para classificá-la em *spam* ou *não-spam*. Os filtros B*ayesianos* têm alguns méritos: primeiro, acumulam um grande número de palavras para tentar identificar *spams* em outras mensagens e são mais flexíveis para detectar *spams* do futuro. Segundo, trabalham para usuários específicos, ou seja, cada usuário poderá construir o seu próprio filtro de mensagens. Uma desvantagem dos filtros B*ayesianos* é que eles exigem do usuário um pouco mais de conhecimento para poder treinar o filtro adequadamente.

Outra possível solução são os filtros baseados em Heurística, os quais olham para padrões de *spam* no cabeçalho e no corpo do e-mail em busca de padrões suspeitos e calculam uma porcentagem de acordo com os resultados obtidos na varredura. Caso a porcentagem exceda o limite pré-definido, o e-mail é considerado *spam*, caso contrário, ele é classificado como *não-spam*. Os filtros baseados em Heurística apresentam um problema:

---

1. A pesquisa sobre técnicas *anti-spam* realizada por Jinpeng Wei encontra-se disponível no endereço: http://www.cc.gatech.edu/people/home/weijp/anti-spam.pdf

para funcionar corretamente devem receber freqüentemente atualizações para reconhecer novos *spams*. Como o *spam* desenvolve-se, as regras baseadas em Heurística também devem desenvolver-se. Se as regras permanecerem estáveis, os *spammers* podem gerar e-mails que contornam facilmente os filtros existentes. Com essa falta de escalabilidade, esse método não é aconselhável, pois poderá resultar em taxa de falso-positivos, ou seja, mensagens legítimas serem incorretamente discriminadas como *spam*.

O último filtro pesquisado por Wei foi o sistema Distribuído para Calcular *Checksum* (DCC), o qual consiste em clientes e servidores de e-mail. No primeiro instante o cliente de e-mail calcula o *checksum* (checar erros) do e-mail recebido e informa o *checksum* para um servidor de e-mail mais próximo. Com isso, o servidor DCC acumula um grande número de *checksums* de mensagens e responde aos clientes. Um servidor DCC poderá também reportar freqüentemente *checksums* para outros servidores vizinhos. A eficácia dos filtros distribuídos conta com o *checksum* para identificar os e-mails semelhantes. Mas, às vezes, o *checksum* ignora alguns aspectos em mensagens consideradas idênticas, dificultando a classificação das mensagens. Os *checksums* são projetados para ignorar somente mensagens desconhecidas. O difícil para os usuários é ter confiança em servidores de e-mail, pois estes encontram problemas para projetar um algoritmo que faça o *checksum* corretamente.

Pelo fato de existirem várias tecnologias *anti-spam*, o presente trabalho faz um estudo das técnicas mais usadas e apresenta os resultados com seus respectivos méritos e problemas.

1.3 Organização do trabalho

A organização deste trabalho traduz as etapas de desenvolvimento da pesquisa aqui apresentada. Desse modo, o Capítulo 2 faz uma apresentação dos problemas ocasionados pela prática de *spam*, ou seja, define o termo *spam* de forma clara, faz um levantamento dos tipos mais comuns de *spam*, apresenta um estudo sobre a evolução do *spam* e estima os custos ocasionados pelo recebimento de mensagens não solicitadas.

A seguir, no Capítulo 3, são explicadas as técnicas utilizadas para combater o recebimento de *spam* e formas de prevenção para reduzir a quantidade de mensagens não solicitadas recebidas diariamente pelos usuários nas suas caixas postais.

Logo após, no Capítulo 4, são analisadas algumas ferramentas *anti-spam* usando filtros *Bayesianos* e, em seguida, são comparadas as ocorrências de falso-positivos (mensagens legítimas serem classificadas como *spam*) e falso-negativos (mensagens consideradas *spam* serem classificadas como *não-spam*) entre as ferramentas avaliadas.

No Capítulo 5, são feitas algumas considerações finais sobre o trabalho desenvolvido e possíveis extensões identificadas para a continuação desta pesquisa.

O Apêndice A tem uma lista das ferramentas *anti-spam* que utilizam a técnica *Bayesiana* para efetuar a classificação das mensagens. O Apêndice B corresponde às configurações realizadas junto ao arquivo do *Procmail* "*.procmailrc*" para processar as mensagens. O Apêndice C apresenta o *Script* utilizado para executar os parâmetros existentes no arquivo "*.procmailrc*".

Capítulo 2

Spam no ambiente Internet

Este capítulo tem por finalidade definir o termo *spam* de forma clara, descrever os tipos mais comuns de *spams* e apresentar a evolução do *spam*. Além disso, analisa os problemas ocasionados pelo recebimento de mensagens não solicitadas.

2.1 Definição do termo *spam*

Na Internet, são encontradas diversas definições de *spam*, mas a maioria delas trazendo o conceito de mensagem não solicitada. Para o âmbito deste trabalho, considera-se *spam* como o recebimento de mensagens não solicitadas, enviadas através de ferramentas automatizadas ou não para uma ou mais pessoas sem que estas tenham solicitado as informações contidas no mesmo.

Teixeira **[13]** define *spam* como

> *"...um abuso e se refere ao envio de um grande volume de mensagens não solicitadas, ou seja, o envio de mensagens indiscriminadamente a vários usuários, sem que estes tenham requisitado tal informação. O conteúdo do spam pode ser: propaganda de produtos e serviços, pedido de doações para obras assistenciais, correntes da sorte, propostas de dinheiro fácil e boatos desacreditando o serviço prestado por determinada empresa."*

Graham **[14]** considera uma tarefa difícil definir *spam*, mas seria conveniente ter uma definição explícita. *"Em primeiro lugar spam não é um e-mail comercial não solicitado. O que define característica de spam não é a razão de ser um e-mail comercial não solicitado, mas sim, e-mails automatizados não solicitados."*

Na verdade é praticamente impossível evitar-se o recebimento dessas mensagens, porém, existem mecanismos que reduzem o volume de mensagens consideradas lixo "*junk e-mail*" em nossas caixas postais.

2.2 Tipos de *spam*

Segundo Teixeira **[13]**, os tipos mais comuns de *spam*, considerando conteúdo e propósito, podem ser classificados da seguinte forma:

Boatos: os boatos, também conhecidos como *hoaxes*, são e-mails que contam histórias, geralmente falsas, inventadas ou distorcidas. Eles abusam da boa fé de seus destinatários, fazendo-os acreditar em uma mentira e incentivando-os a passá-la adiante: ***"envie este e-mail a todos os seus amigos..."***, o que gera um aumento no tráfego de dados na Internet. Algumas classes comuns de boatos são os que apelam para a necessidade que o ser humano possui de ajudar o próximo. Como por exemplo, os casos de crianças com doenças graves, aqueles que difamam empresas ou produtos e até mesmo os que prometem brindes ou ganho de dinheiro fácil. Ainda, dentre os boatos mais comuns na rede, pode-se citar aqueles que tratam de código malicioso, como vírus ou cavalo de tróia. Neste caso, a mensagem sempre fala de vírus poderosíssimos, capazes de destruir seu computador e assim por diante.

Correntes: as correntes, *chain letters*, por sua vez, podem ser de dois tipos: abaixo-assinados contra ou a favor de uma determinada causa (antes de repassar para outras pessoas você deve adicionar seu endereço eletrônico à lista que já existe na mensagem, o que é uma ótima fonte de destinatários para *spammers* – autor dos *spams*) ou simplesmente correntes da sorte, dizendo que você sofrerá alguma conseqüência indesejável se não repassar a mensagem para um determinado número de pessoas. Um exemplo são as correntes de vários santos.

Propagandas: são *spams* com o intuito de divulgar produtos, serviços, novos sites, enfim, propaganda em geral, e têm ganhado cada vez mais espaço nas caixas postais dos internautas. Assim, muitas empresas têm usado este recurso para atingir os consumidores. Isto sem contar a propaganda política que inundou as caixas postais nas últimas eleições. Vale ressaltar que, seguindo o próprio conceito de *spam*, se recebemos um e-mail que não solicitamos, estamos sim sendo vítimas de *spam*, mesmo que seja um e-mail de uma promoção que muito nos interessa.

Fraudes e Golpes (Scams): infelizmente, o mundo virtual, na maioria das vezes, imita o mundo real e não demorou muito para que aparecessem as fraudes e golpes na

Internet. Com o advento do comércio eletrônico e a realidade do Internet Banking, tem-se notado a proliferação e a diversificação das fraudes e golpes via rede, muitas vezes usando o e-mail como ferramenta. Observa-se um certo padrão nos e-mails fraudulentos, já que a maioria é de mensagens com conteúdo falso, usando o nome de empresas ou instituições idôneas e visando induzir o usuário a instalar aplicativos que são na verdade códigos maliciosos. Para não levantar suspeita sobre a legitimidade do e-mail, é comum usar contas de e-mail forjadas, com nomes ou domínios muito parecidos aos da empresa real. Por exemplo: suporte@empresa.nom.br, enquanto que o domínio verdadeiro seria empresa.com.br. Os textos dos e-mails fraudulentos geralmente não são bem escritos, têm erros gramaticais e de ortografia, mas tentam convencer o usuário de que se trata de um comunicado oficial, usando recursos de engenharia social (é um método de ataque, onde alguém faz uso indevido, muitas vezes abusando da ingenuidade ou confiança do usuário, para obter informações que podem ser utilizadas para ter acesso não autorizado a computadores ou informações).

Pornografia: um dos conteúdos freqüentes em e-mails não solicitados é a pornografia. Na maioria das vezes, o usuário ao abrir a caixa de entrada se depara com fotos e cenas degradantes, incluindo pedofilia. Vale ressaltar que os casos de pedofilia têm sido investigados com rigor e, portanto, recomenda-se denunciar eventuais e-mails com conteúdo desse gênero aos órgãos competentes.

Ameaças, brincadeiras e afins: alguns *spams* são enviados com o intuito de fazer ameaças, brincadeiras de mau gosto ou apenas por diversão. Ainda assim são considerados *spam*. Casos de ex-namorados difamando ex-namoradas, e-mails forjados assumindo identidade alheia e aqueles que dizem: ***"olá, estou tentando uma nova ferramenta spammer e por isto, você está recebendo este e-mail"***, constituem alguns exemplos. Vale lembrar que não há legislação específica para casos de *spam*. No entanto, pode-se enquadrar certos casos nas leis vigentes no atual Código Penal Brasileiro, tais como: calúnia e difamação, falsidade ideológica ou estelionato.

Uma pesquisa realizada pela companhia de software a*nti-spam* SurfControl[1] apresenta de forma estatística os tipos de *spams* mais comuns que ocorrem diariamente nas organizações (ver Figura 2.1):

---

1. Dados obtidos do centro de pesquisa SurfControl soluções *anti-spam*. Maiores detalhes no endereço: http://www.surfcontrol.com/resources/Anti-Spam_Study_v2.pdf

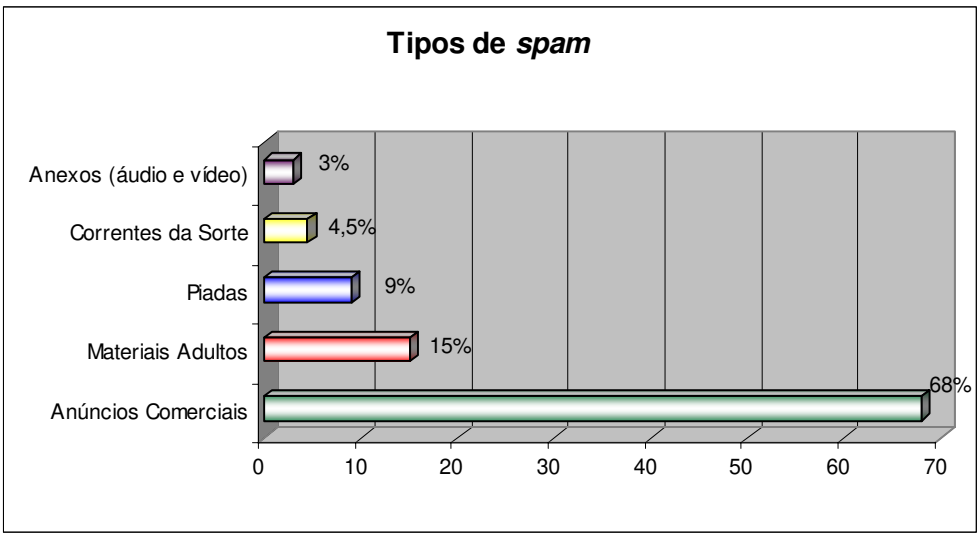

Figura 2.1: Tipos de *spam*

Na Figura 2.1, nota-se que a maior parte das mensagens não solicitadas se referem a anúncios comerciais (produtos e serviços em geral), totalizando 68% das mensagens não solicitadas. O restante é categorizado como: 15% materiais adultos (pornografia), 9% piadas, 4,5% correntes da sorte e 3% arquivos anexos contendo áudio e vídeo. Essa pesquisa foi realizada em setembro de 2002. No item 2.8 é possível verificar, através da Figura 2.4, uma pesquisa mais recente dos tipos de *spam*.

2.3 Os Protagonistas (*spammers*)

De acordo com Vianna **[15],**

> *"... o spammer é um oportunista, deseja ganhar dinheiro facilmente ou promover rapidamente seu produto. Não tem conhecimentos de publicidade, nem de marketing. Compra listas de e-mails, envia mensagens sem sequer conhecer o perfil dos destinatários e espera obter algum lucro. Dificilmente se pode crer que um spammer seja um empresário sério, com um produto sério, ou tenha aplicado um mínimo de marketing na elaboração de sua mensagem. Em muitos casos, o e-mail disponibilizado para contato é falso, destruindo a confiança do destinatário e aumentando as suspeitas de estelionato."*

Segundo o Centro de Estudos sobre Segurança e Vulnerabilidade na Internet do Rio Grande do Sul (CERT-RS)[1], os *spammers* na maioria das vezes "seqüestram" servidores de

1. Para consultar a lista de documentos sobre segurança produzidos pelo (CERT-RS) acesse o endereço: http://www.cert-rs.tche.br/servicos/infosec.html

e-mail para enviar suas mensagens não desejadas por toda Internet. Os *spammers* usam a retransmissão para aumentar o número de mensagens que eles podem enviar. Um simples computador que está na ponta de uma linha telefônica só pode enviar um número limitado de mensagens. Se, entretanto, o *spammer* tomar controle de um servidor (HOST) de e-mail com uma conexão de rede rápida, então ele pode enviar centenas de vezes mais "lixo". Além disso, se ele puder retransmitir através de vários servidores de e-mail em paralelo, pode inundar a rede com quantidades extraordinárias de "lixo". A crença *do spammer* é: "*Por que pagar por recursos de redes e computação caras quando nós podemos roubar os seus*?".

De acordo com as definições acima, pode-se concluir que s*pammers* são falsificadores de cabeçalhos (*headers*), que usam indevidamente os *relays* abertos (servidor de e-mails de terceiros mal configurados), na maioria das vezes utilizando declarações falsas nas mensagens e através de ferramentas do tipo (*harvesting*) com o propósito de capturar endereços de e-mail válidos na Internet para compor a sua base de dados.

2.4 Problemas causados pelo *spam*

Segundo o NBSO (Grupo de Resposta a Incidentes para a Internet Brasileira)[1], os usuários do serviço de correio eletrônico podem ser afetados de diversas formas. Alguns exemplos são:

Não recebimento de e-mails: boa parte dos provedores de Internet limita o tamanho da caixa postal do usuário no seu servidor de e-mails. Caso o número de *spams* recebidos seja muito grande, o usuário corre o risco de ter sua caixa postal lotada com mensagens não solicitadas. Se isso ocorrer, todas as mensagens enviadas a partir de então serão devolvidas ao remetente e o usuário não conseguirá mais receber e-mails até que possa liberar espaço em sua caixa postal;

Gasto desnecessário de tempo: para cada *spam* recebido, o usuário necessita gastar um determinado tempo para ler, identificar o e-mail como *spam* e removê-lo da caixa postal;

Aumento de custos: independentemente do tipo de acesso à Internet utilizado, quem paga a conta pelo envio do *spam* é quem o recebe. Por exemplo, para um usuário que utiliza

---

1. Para consultar a lista de documentos produzidos pelo (NBSO) acesse o endereço: http://www.nbso.nic.br/docs/

acesso discado à Internet, cada *spam* representa alguns segundos a mais de ligação que ele estará pagando;

Perda de produtividade: para quem utiliza o e-mail como uma ferramenta de trabalho, o recebimento de *spams* aumenta o tempo dedicado à tarefa de leitura de e-mails, além de existir a chance de mensagens importantes não serem lidas, serem lidas com atraso ou apagadas por engano.

Além de causar esses problemas ao usuário, o *spam* aumenta os custos para os provedores de acesso e empresas. São eles:

Impacto na banda: para as empresas e provedores, o volume de tráfego gerado por causa de *spams* os obriga a aumentar a capacidade de seus *links* de conexão com a Internet. Como o custo dos *links* é alto, isso diminui os lucros do provedor, muitas vezes podendo refletir no aumento dos custos para o usuário;

Má utilização dos servidores: os servidores de e-mail dedicam boa parte do seu tempo de processamento para tratar das mensagens não solicitadas. Além disso, o espaço em disco ocupado por mensagens não solicitadas enviadas para um grande número de usuários é considerável;

Perda de Clientes: os provedores muitas vezes perdem clientes que se sentem afetados pelos *spams* que recebem ou pelo fato de terem seus e-mails filtrados por causa de outros clientes que estão enviando *spam*;

Investimento em pessoal e equipamentos: para lidar com todos os problemas causados pelo *spam,* os provedores necessitam contratar mais técnicos especializados e acrescentar sistemas de filtragem de *spam*, que implicam na compra de novos equipamentos. Como conseqüência, os custos do provedor aumentam.

2.5 Mutações do *spam*

Em outubro de 2003, Singel publicou uma matéria onde se refere às mensagens de *spam* que mudam a cada momento **[16]**. Os *spammers* anunciam outro tipo de medicamento quando o mercado de "*Viagra"* começa a murchar. Eles escrevem o nome desse medicamento de todas as formas possíveis e inserem *tags* de HTML (*HyperText Markup*

Language, Linguagem de Formatação de Hipertexto) invisíveis entre as letras de outras palavras-chave para que não sejam detectadas por programas *anti-spam*.

Os desenvolvedores de software *anti-spam* afirmam que os *spammers* estão criando e adaptando novas estratégias para vencer os software cada vez mais sofisticados usados pelos internautas para proteger suas caixas de entrada. Mas eles esperam que a ampla adoção de software de filtragem acabe desencorajando estratégias como as que os *spammers* vêm tomando até agora, forçando-os a adotar meios mais inócuos de anunciar seus produtos. "*Existe uma pressão evolutiva predatória no mundo. É o spam procurando forçar a passagem*", disse Roger Matus **[16]**, cuja empresa comercializa o *InBoxer*, um avançado programa de bloqueio de *spam*. "*O sucesso tem sido limitado, pois, até mesmo os modernos filtros Bayesianos têm dificuldades em bloquear spams que não parecem spam*". Segundo Matus, as mensagens que conseguem ultrapassar esses filtros não passam de um simples e-mail convidando o destinatário a visitar um determinado s*ite* e isso pode convencer os *spammers* de que jogar limpo pode ser uma boa idéia.

O *spam* está evoluindo, reagindo às medidas preventivas que vão desde as mais simples, como regras de bloqueio de mensagens baseadas em palavras-chave e banco de dados contendo endereços, até as mais tecnicamente inteligentes, como os filtros *Bayesianos*.

Os filtros *Bayesianos* fazem uso de estatística e probabilidade para classificar mensagens através da análise das palavras e cabeçalhos de e-mails passados. Eles aprendem com o tempo. Por exemplo, assim que as mensagens de *spam* começam a anunciar "*Viagr@*", isso ensina ao filtro que, a partir de então, as mensagens em que o famoso remédio para impotência for escrito com o sinal "@" serão mensagens de *spam*.

2.6 Custo do s*pam*

Uma pesquisa realizada pela Ferris Research estima que no ano de 2002, o custo do *spam* foi de aproximadamente $8,9 bilhões de dólares para as empresas nos Estados Unidos e $2,5 bilhões na Europa **[7]**. No ano seguinte, essa proporção ultrapassou os $10 bilhões de dólares para as empresas nos Estados Unidos. Desse valor, 44% são consumidos pela infra-estrutura de TI (Tecnologia da Informação) absorvida na tentativa de filtrar e apagar os

spams, 39% vêm da perda de produtividade dos usuários finais e 17% se devem ao aumento do uso do suporte técnico para a solução de problemas originados pelos e-mails (ver Figura 2.2):

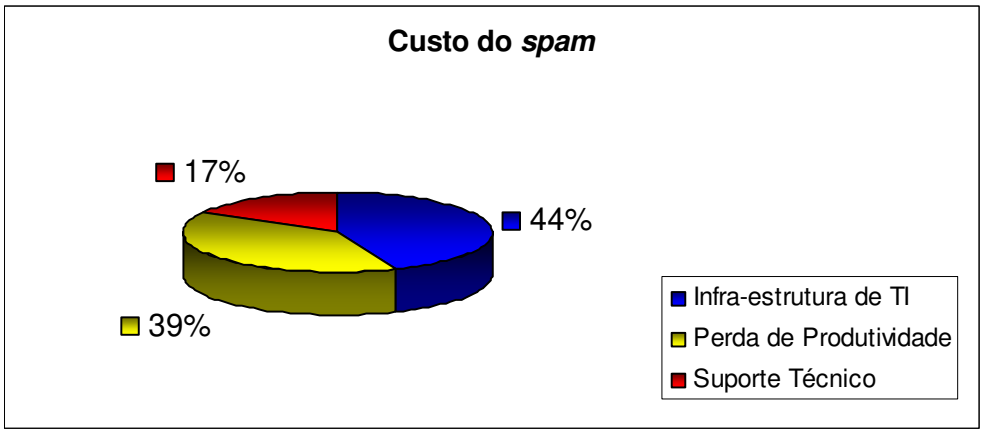

Figura 2.2: Custo do *spam* para as corporações

O maior problema com a propaganda por *spam* é que a Internet se mostra como meio fértil para divulgação de produtos, atinge um grande número de pessoas e a custo baixo ou praticamente nulo. Usando técnicas profissionais, os *spammers* muitas vezes, mandam mensagens aproveitando brechas em servidores de terceiros mal configurados.

No caso de recebimento de *spam*, quem paga a conta é o destinatário, pois acaba gastando mais tempo conectado ao seu provedor para selecionar mensagens válidas, em meio aos tantos *spams* recebidos diariamente. Sem contar que a banda do provedor e seu servidor de correio eletrônico ficam sobrecarregados com o grande volume de mensagens desnecessárias todos os dias.

A companhia de software Computer Mail Services desenvolveu uma ferramenta[1] *online* que calcula o custo do *spam*. Através dessa ferramenta, os internautas podem inserir os dados de uma determinada empresa e calcular os custos. Como por exemplo, ao calcular os custos de uma empresa que possui 500 (quinhentos) empregados, cada um deles recebendo apenas 5 (cinco) *spams* por dia e gastando somente 10 (dez) segundos do seu tempo para confirmar que se trata de um *spam* e deletá-lo, essa ferramenta gera como resultado que se perde cerca de R$ 35.000,00 por ano (ver figura 2.3):

---

1. Esta ferramenta encontra-se disponível no seguinte endereço: http://www.cmsconnect.com/Marketing/spamcalc.htm

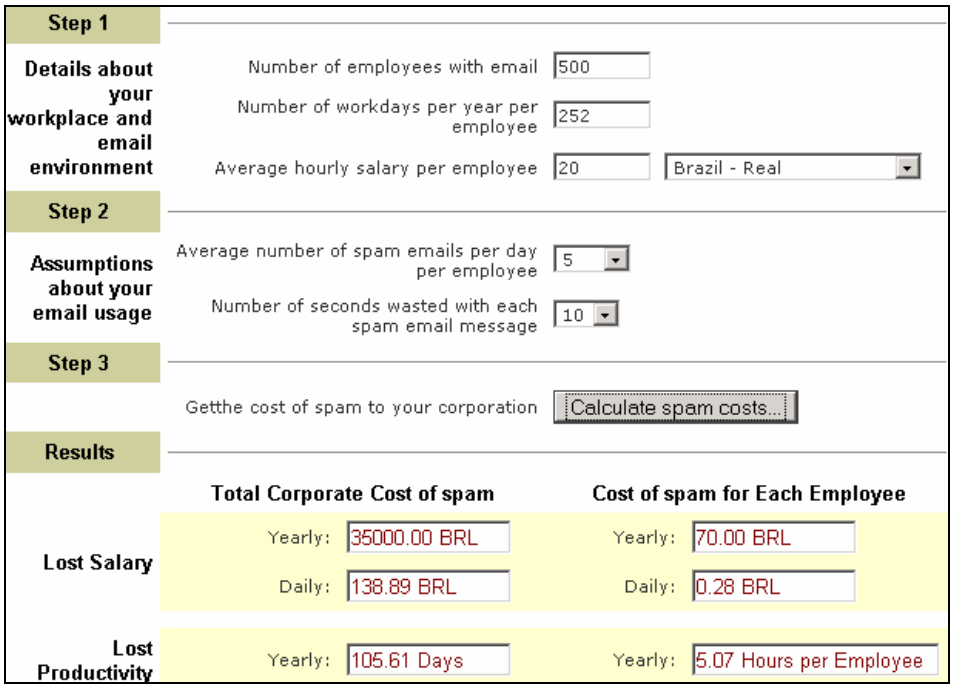

Figura 2.3: Ferramenta para calcular o custo do *spam*

A suposta facilidade e o falso baixo custo de deletar um *spam* transformam-se em números preocupantes quando agregados em volumes significativos. Trinta e cinco mil reais é o valor calculado pela Computer Mail, mas deve-se levar em conta que 5 (cinco) *spams* é um número pequeno e o tempo de 10 segundos nem sempre é o suficiente para realizar essa desagradável tarefa.

2.7 Reações contra o *spam*

"Combater o spam é uma questão complexa que envolve diversos aspectos como tecnologia, auto-regulamentação, legislação e punição aos infratores", enumera Emilio Umeoka, presidente da Microsoft Corporation Brasil **[17]**.

A polêmica em torno do *spam* é grande e já existem diversos grupos na Internet que lutam contra esse gênero publicitário. Eis algumas referências mais importantes:

- ***Movimento Brasileiro de Combate ao Spam:*** página com técnicas para evitar o *spam* e notícias sobre ele no Brasil. Veja em http://www.antispam.org.br;
- ***Spam Laws:*** página com leis mundiais *anti-spam* no endereço http://www.spamlaws.com;
- ***SpamCon Foundation:*** fornece mecanismos para reduzir o volume de *spam*. Maiores detalhes acesse http://spamcon.org;
- ***CAUCE – Coalition Against Unsolicited Commercial Email:*** é um grupo que luta contra o e-mail comercial não solicitado no endereço http://www.cauce.org;
- ***Fight Spam on the Internet:*** página que prega o boicote ao *spam* e ensina como se livrar dele. Detalhes podem ser obtidos em http://spam.abuse.net/;
- ***SpamAbuse:*** página que fornece aos usuários ajuda de como combater a prática abusiva do *spam* com o uso de filtros *anti-spam*. Acesse http://www.spamabuse.org/.

Além dessas referências acima citadas, podem ser encontradas na Internet várias outras entidades que também lutam contra o *spam*, procurando minimizar o recebimento dessas mensagens não solicitadas.

2.8 Estatísticas de *spam* no Mundo

Segundo a Brightmail, especializada em soluções *anti-spam*, dos mais de 96 bilhões de e-mails filtrados pela companhia em abril de 2004, 64% eram *spam.* Desses 64%, a Brightmail[1] fez um levantamento dos tipos mais comuns de *spams* ocorridos diariamente na Internet (ver Figura 2.4):

1. Com base nos dados obtidos pela companhia Brightmail Anti-Spam é possível observar os tipos mais comuns de *spams* que ocorrem com maior freqüência na Internet. A companhia disponibiliza esses dados através do endereço: http://www.brightmail.com

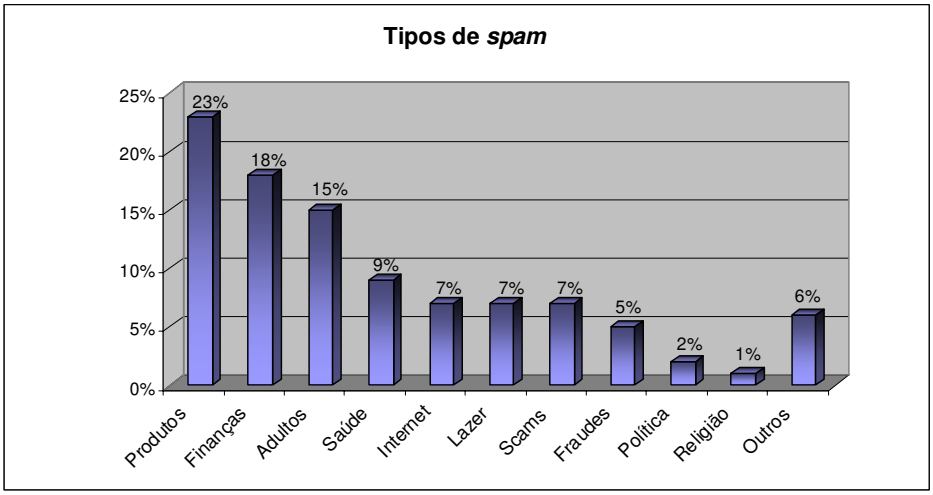

Figura 2.4: Porcentagens de *spam*

Pode-se verificar, a partir da análise da Figura 2.4, que os tipos de *spam* mais comuns são:

- ***Produtos (23%)*** – e-mails oferecendo produtos e serviços em geral. Ex.: serviços de investigação e produtos de maquiagem;
- ***Finanças (18%)*** – e-mails contendo oportunidades financeiras. Ex.: empréstimos e investimentos;
- ***Adultos (15%)*** – e-mails contendo produtos e serviços referentes às pessoas adultas. Ex.: pornografia;
- ***Saúde (9%)*** – e-mails oferecendo produtos e serviços pertinentes à área da saúde. Ex.: produtos farmacêuticos e tratamento médico;
- ***Internet (7%)*** – e-mails oferecendo serviços ligados à Internet. Ex.: hospedagem de sites;
- ***Lazer (7%)*** – e-mails oferecendo prêmios e atividades de lazer. Ex.: ofertas de férias, cassinos *online* e jogos;
- ***Scams (7%) –*** e-mails categorizados como fraudulentos. Ex.: um e-mail chega à caixa de entrada do usuário oferecendo promoções e vantagens ou solicitando algum tipo de recadastramento;

- ***Fraudes (5%)*** – e-mails fraudulentos agindo de má fé, tentando obter os dados pessoais como informações financeiras. Ex.: atualização de conta bancária e verificação de cartão de crédito;
- ***Política (2%)*** – e-mails notificando campanhas políticas em geral. Ex.: Partido Político e Eleições;
- ***Religião (1%)*** – e-mails contendo informações religiosas ou espirituais. Ex.: Astrologia e religião em geral;
- ***Outros (6%)*** – e-mails pertinentes às outras categorias.

Segundo levantamento do renomeado Projeto *Anti-spam Spamhaus*, entidade que publica regularmente endereços IP denunciados por envio de *spams*, o Brasil aparece em quarto lugar durante pesquisa realizada em maio de 2004 **[18]**.

Os endereços coletados são armazenados em uma das bases de dados do projeto chamada *Spamhaus Block List* (SBL), lista de endereços IP bloqueados da entidade *Spamhaus*. A partir dessa lista, é possível informar aos usuários em tempo real as fontes reais de *spam*.

Através do *site* da companhia pode-se verificar os 10 (dez) países com maior índice de *spam* no mundo em maio de 2004 (ver Tabela 2.1):

Os dez países com maior índice de *spam* em maio de 2004	
1	Estados Unidos
2	China
3	Coréia do Sul
4	Brasil
5	Canadá
6	Taiwan
7	Argentina
8	Rússia
9	Itália
10	Reino Unido

Tabela 2.1: Países com maior índice de *spam*

De acordo com a Tabela 2.1, o Brasil ocupa a 4ª (quarta) posição no *ranking* das nações mais afetadas. O primeiro lugar cabe aos Estados Unidos e em seguida vem a China

e a Coréia do Sul.

2.9 Estatísticas de *spam* no Brasil

A fama mundo afora não é a única conseqüência grave da ausência de medidas mais enérgicas em relação ao *spam*. No Brasil, usuários domésticos e empresas vêm sofrendo com o aumento da disseminação dessas desagradáveis mensagens. Basta observar o volume de e-mails impessoais recebidos nos últimos tempos, onde a maioria deles são escritos em inglês, o que mostra que há *spammers* internacionais que se aproveitam da legislação do Brasil com relação ao tema para continuar agindo com suas mensagens não solicitadas.

Portanto, alguns grupos foram criados com o objetivo de receber, analisar e responder a incidentes de segurança em computadores como, por exemplo, o CERT Coordination Center (CERT/CC), que realiza um levantamento sobre os incidentes reportados pelo mundo e o grupo de respostas a incidentes de segurança NIC BR Security Office (NBSO), ligado ao Comitê Gestor da Internet no Brasil, responsável por esse tipo de monitoração no país.

Segundo o NBSO, um fator preocupante observado pelos especialistas foi o aumento do tráfego de *spam* nos últimos anos[1]. Em julho de 2004, o grupo publicou em seu *site* as estatísticas de notificações de *spam* referentes ao segundo trimestre de 2004 (ver Tabela 2.2):

Mês	Total	*SpamCop*								Outras Fontes (%)	
		Spamvertised (%)		*Proxy* (%)		*Relay* (%)		Outras (%)			
abr	**322.614**	63.484	19,68	135.758	42,08	117	0,04	80.452	24,94	42.803	13,27
mai	**385.401**	118.622	30,78	135.570	35,18	133	0,04	86.545	22,46	44.531	11,55
jun	**450.560**	165.664	36,77	136.777	30,36	135	0,03	104.969	23,30	43.015	9,55
total	**1.158.575**	347.770	30,02	408.105	35,22	385	0,03	271.966	23,47	130.349	11,25

Tabela 2.2: Totais mensais classificados por tipo de reclamação

Considerando as características citadas acima, pode-se observar que a entidade *SpamCop*, por exemplo, possui um serviço que processa reclamações de *spam* e as envia

1. *Spams* reportados ao NBSO (abril a junho de 2004). Maiores detalhes podem ser obtidos no endereço: http://www.nbso.nic.br/stats/spam/

para os responsáveis pelas redes envolvidas com o abuso. Essas reclamações envolvem máquinas com *proxies* ou *relays* abertos, possivelmente sendo abusados, ou máquinas hospedando páginas com informações de produtos e serviços sendo oferecidos no *spam* (*spamversited website*). A Tabela 2.2, na coluna Outras, refere-se às reclamações enviadas pelo *SpamCop*, mas não se enquadram nos tipos anteriores e a coluna Outras Fontes, são reclamações enviadas por fontes diferentes do *SpamCop*.

Um *proxy* é um serviço que atua como intermediário entre um cliente e um servidor **[19]**, e um *relay* é uma funcionalidade do serviço SMTP (*Simple Mail Transfer Protocol*, Protocolo Simples para a Transferência de Correio Eletrônico) que permite receber e-mails de clientes e retransmiti-los, sem modificações, para outro servidor SMTP **[20, 21].** Máquinas com *proxies* e *relays* abertos podem ser utilizadas indiscriminadamente por terceiros para enviar *spam*, dificultando a identificação da real origem. *Proxies* abertos podem também ser usados como pontes para realização de invasões e desfigurações de páginas *Web* ou como meio de obter anonimato ao cometer crimes como estelionato ou pornografia envolvendo crianças **[22]**.

O grupo de resposta a incidentes NBSO, ao receber as informações da entidade *SpamCop* sobre quais máquinas estão mal configuradas e possivelmente sendo abusadas na rede brasileira, informa rapidamente o administrador ou analista de segurança para solucionar o problema e evitar que sua rede seja utilizada indevidamente por terceiros.

2.10 *Opt-in versus Opt-out*

A empresa de consultoria *Peppers & Rogers Group* do Brasil enfatiza uma campanha de marketing por e-mail livre de *spam* **[23]**. As informações contidas no site da consultoria têm por objetivo esclarecer as empresas e as pessoas, com idéias, informações e conteúdo para que possam estabelecer práticas de marketing por e-mail de forma ética e aceitável. Segue abaixo uma breve descrição dos termos *Opt-in* e *Opt-out*.

2.10.1 *Opt-in* (Permissão)

A boa notícia é que campanhas de marketing podem ser conduzidas por e-mail sem causar um volume significativo de reclamações por parte dos clientes e sem desperdiçar largura de banda. Basta adotar a política correta, conhecida pelo nome de *opt-in*.

A idéia básica do *opt-in* é que o cliente "pede para entrar". Ou seja, quem recebe e-mails da empresa são apenas aquelas pessoas que já indicaram explicitamente seu interesse nos produtos e serviços oferecidos (ver Figura 2.5):

Figura 2.5: *Opt-in* (Permissão)

Na Figura 2.5, o cliente acessa o *site* de uma determinada empresa e começa a preencher os formulários *online* para realizar futuras compras. Ao se cadastrar, o cliente fornece alguns dados, como nome completo e endereço de e-mail. O formulário inclui um campo especial que diz algo assim: "Quero receber informações por e-mail". O usuário só clica acionando esse campo se concordar em receber informações e ofertas dessa empresa.

Portanto, o método *opt-in* permite ao cliente escolher se quer ou não receber e-mails da empresa. Mais importante, o cliente sabe que pode optar pela inclusão nesse momento.

2.10.2 *Opt-out* (Privacidade)

O termo *opt-out* significa "pedir para sair": nesse esquema, os usuários recebem mensagens por e-mail que explicam claramente o método a ser utilizado para se excluírem da lista. Na maioria das vezes, esses usuários nunca pediram explicitamente para aderir à lista e, portanto, continuarão a receber e-mails até que peçam para sair. Para isto, utilizam um

hyperlink embutido no e-mail que leva para um *site*, um mecanismo simples de resposta por e-mail, ou algum outro meio (ver Figura 2.6):

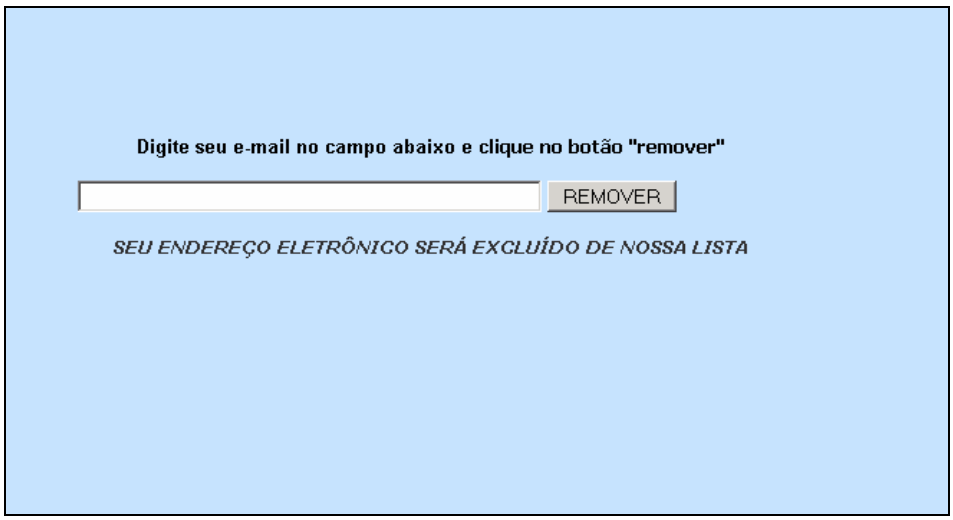

Figura 2.6: *Opt-out* (Privacidade)

A chave para o sucesso de uma campanha de *opt-out* reside em como fazer a coleta e selecionar endereços de e-mail para criar o banco de dados. Por exemplo, se uma determinada empresa tem um banco de dados sobre os clientes que já compraram seus produtos, envia uma mensagem apenas de agradecimento em primeiro lugar, não esquecendo de explicar como podem sair da lista como mostrado na Figura 2.6.

Ainda assim, é certeza que uma campanha do tipo *opt-out* irá gerar reclamações por parte dos clientes e que não é a melhor abordagem de marketing por e-mail.

Com relação às definições acima pode-se concluir que *opt-in* e *opt-out* são propagandas via e-mail. Alguns *sites* oferecem aos clientes a opção “quero receber informações por e-mail”. Quando não solicitados denominam-se *spam*.

2.11 Conclusão

Este capítulo apresentou os principais conceitos do termo *spam* e problemas ocasionados pelo recebimento de mensagens não solicitadas, buscando fornecer aos usuários de correio eletrônico o embasamento teórico das graves conseqüências proporcionadas pela prática abusiva de *spam*.

Capítulo 3

Combate ao *Spam*

Este capítulo tem por finalidade apresentar técnicas variadas no combate ao *spam* e formas de prevenção para reduzir a quantidade de mensagens não solicitadas recebidas diariamente pelos usuários nas suas caixas postais.

3.1 Filtros *Anti-spam*

Muito tem se falado sobre soluções de combate ao s*pam* e muita controvérsia tem aparecido em eventos técnicos. No entanto, uma conclusão ponderada é a de que não existe uma receita capaz de eliminar o *spam* da Internet, muito menos uma solução única que resolva o problema dos administradores e usuários, mas existem algumas alternativas capazes de reduzir o impacto causado pelo s*pam*. A mais usada atualmente é a filtragem e, neste caso, já se tem uma variedade de filtros disponíveis em diversos níveis.

As opiniões conservadoras defendem que filtrar e-mail na Internet vai prejudicar um dos serviços básicos da rede que é o correio eletrônico. Para justificar o uso de filtros, basta computar o prejuízo referente à banda consumida pelo *spam* e o tempo gasto pelo usuário para limpar sua caixa postal todos os dias ou várias vezes ao dia.

Os filtros, como o próprio nome diz, permitem fazer uma triagem nos e-mails recebidos, separando os *spams* dos e-mails válidos. É importante ressaltar que, quando filtros são usados, é recomendada a configuração de uma mensagem padrão que será enviada à origem do e-mail, explicitando que este não pode ser entregue, pois o (domínio / usuário / rede) do remetente está listado como *spammer* na referida lista de filtragem, ou ainda, por ter conteúdo considerado suspeito de *spam* **[24]**.

Segundo o grupo de pesquisa Info-Tech **[7]**, a técnica de filtragem de *spam* mais antiga ficou conhecida como *pattern matching* (casamento de padrões). Esse método faz uma varredura no cabeçalho e no corpo do e-mail em busca de expressões regulares pré-definidas (palavras ou frases). Caso o e-mail possua alguma dessas expressões, o mesmo

será descartado. Esse método não é aconselhável pois resulta em muito falso-positivos, ou seja, algumas expressões pré-definidas podem incriminar uma mensagem legítima.

A partir da expressão *pattern matching* surgiram várias técnicas de filtragem de *spam* sendo que as mais importantes são listas de bloqueio / permissão, classificação de conteúdo e autenticidade do remetente.

3.1.1 Listas de Bloqueio / Permissão

Ferramentas baseadas em lista de bloqueio são conhecidas como lista-negra (*blacklist*) e, lista de permissão, como lista-branca (*whitelist*). Ambas analisam o cabeçalho das mensagens recebidas e identificam endereços IP, domínios ou endereços de e-mail que devem ser bloqueados ou permitidos, respectivamente.

As listas-negras são cadastradas de duas formas: ***automática***, através de bancos de dados *online* mantidos por entidades na Internet, baseando-se em denúncias comprovadas; ***manual***, através de um meio para o usuário indicar que uma mensagem é *spam* e, assim, o remetente ser adicionado à lista-negra.

Devido ao grande número de *spams* enviados atualmente na Internet, várias ONGs - Organizações Não Governamentais - têm sido criadas com a finalidade de conscientizar e mobilizar a comunidade Internet contra essas mensagens indesejáveis. Exemplos claros destas ONGs são a ORDB e a MAPS, dentre outras:

ORDB (Open Relay Database, Banco de Dados de Relays Abertos) **[25]:** é uma organização sem fins lucrativos que armazena uma lista de endereços IP de *relays* abertos verificados. Esses *relays* são, ou provavelmente serão, utilizados para o envio de mensagens não solicitadas. Acessando essa lista, os administradores de sistemas podem escolher entre aceitar ou rejeitar mensagens vindas de servidores nesses endereços. A intenção é que as empresas, principalmente os provedores de acesso, passem a utilizar esse cadastro como referência para o bloqueio das mensagens enviadas por endereços ali listados e, dessa forma, forçar o responsável pelo referido servidor a corrigir a configuração do mesmo.

***MAPS (Mail Abuse Prevention System, Sistema de Prevenção de Abuso de E-mails)* [26]:** tem uma missão um pouco mais abrangente, visa, além de manter um cadastro *de relays* abertos *MAPS RBL (Realtime Blackhole List)*, a educação dos provedores de acesso à Internet, para que estes não permitam o envio de *spams* por seus clientes e/ou servidores. Essa organização também disponibiliza outras formas de bloqueio de *spams* como *MAPS RSS (Relay Spam Stopper)*, uma lista com a mesma finalidade da ORDB, porém, orientada ao servidor DNS – *Domain Name System*, *MAPS DUL (Dynamic User List)*, refere-se às redes Dial-up e DSL reconhecidamente como fonte de spam e *MAPS OPS (Open Proxy Stopper)*, lista de endereços IP que possuem *proxies* mal configurados e indevidamente usados para envio de *spam*.

Em função de toda essa mobilização, com a devida complacência dos provedores, a prática disso acaba estabelecendo uma regra, onde os provedores que permitem aos seus clientes o envio de tais mensagens acabam sendo listados nesses órgãos e impedidos de enviar mensagens a terceiros até a regularização da situação.

Para forçar a aceitação incondicional de determinados remetentes conhecidos, mesmo quando a origem estiver cadastrada em lista-negra, as ferramentas oferecem ao usuário o recurso de cadastrar também uma lista-branca com esses endereços permitidos. A lista-branca tem, em geral, precedência sobre a lista-negra.

A filtragem de *spam* baseada em listas é bastante precisa e seletiva, desde que as listas de bloqueio e permissão sejam sempre atualizadas. Mas elas perdem a eficácia na medida em que os *spammers* (remetentes de *spam*) se protegem dessa filtragem sendo "nômades" mudando freqüentemente de endereço eletrônico, provedor e utilizando endereços de e-mails falsos.

3.1.2 Classificação de Conteúdo

A classificação de conteúdo usa uma abordagem diferente. Ao invés de analisar apenas o cabeçalho procurando identificar remetentes suspeitos, ela analisa todo o conteúdo da mensagem, isto é, o texto completo em busca de padrões suspeitos e, com base na identificação de determinados padrões, utiliza estatística e probabilidade para fazer uma

classificação do que é ou não *spam* **[8]**. Essa técnica *anti-spam* é baseada na filtragem *Bayesiana*.

O filtro *Bayesiano* é mais flexível do que a filtragem por listas de bloqueio, pois não depende nem da identificação de remetentes nem da manutenção de listas destes. Mesmo que o *spammer* seja nômade e consiga disfarçar sua origem, o texto da mensagem pode denunciá-lo. Mas essa filtragem apresenta alguns problemas, pois depende de aprendizagem e não é exata, é estatística, e por isso terá sempre um risco, mesmo que gradativamente menor, de classificação incorreta, isto é, da ocorrência de eventuais falso-positivos (mensagens legítimas indevidamente descartadas) e falso-negativos (deixar passar mensagens que na verdade são *spam*).

É importante também compreender o que é a "aprendizagem" necessária para o filtro *Bayesiano*. Inicialmente é difícil definir parâmetros e padrões de texto fixos para se determinar o que é ou não *spam*. As ferramentas baseadas em filtro provêm ao usuário uma forma de classificar manualmente cada mensagem como *spam* ou não. À medida que as classificações são feitas, a ferramenta mapeia o conteúdo – palavras, padrões de texto – das mensagens já classificadas, formando uma base estatística para classificar automaticamente mensagens futuras. Quanto mais tempo e mensagens classificadas, mais o filtro terá amostragem maior, mais diversificada e detalhada para tornar sua classificação automática cada vez mais precisa.

Os *spammers* atentos à existência de filtros *Bayesianos* tentam driblar a classificação de conteúdo, introduzindo no meio das palavras do texto caracteres ou letras a mais, visando dificultar a determinação de padrões automáticos, sem prejudicar muito a legibilidade humana do texto. Assim, ao invés de escrever simplesmente a palavra "sexo", podem escrever "se-xo" ou "sexxo", por exemplo. Outro despiste é o uso de imagens (*banners*) ao invés de texto, nas mensagens.

3.1.3 Autenticidade do Remetente

É um filtro de *spam* que verifica a identidade do remetente[1] em vez de filtrar o conteúdo da mensagem. Ao contrário dos filtros *Bayesianos* essa técnica não necessita de treinamento inicial, ou seja, a criação de uma base de dados de *spam* e não *spam* para a realização da

1. Disponível no endereço: http://www.alphaworks.ibm.com/tech/fairuce

filtragem das mensagens. Segundo especialistas os filtros de conteúdo exigem manutenção freqüentemente para bloquear os *spams* do futuro e, além disso, exige uma grande quantidade de processamento para técnicas complexas tal como os filtros *Bayesianos* e regras baseadas em Heurísticas.

O filtro Autenticidade do Remetente, praticamente elimina endereços IP forjados e muitos vírus que procuram por DNS *look-ups*. Essa técnica foi postada em Novembro de 2004 e, com isso, espera-se que a mesma seja testada o suficiente para posterior implantação junto ao servidor de e-mail (SMTP) de uma organização. O autor desta tecnologia aponta que essa ferramenta *anti-spam* seja no futuro o filtro mais adequado para combater o recebimento de mensagens não solicitadas.

Devido ao fato que essa ferramenta esteja em fase inicial de testes e em avançados estudos, o presente trabalho irá focar em especial na classificação de conteúdo (filtros *Bayesianos*).

3.2 Filtros *Bayesianos*

Como um presbiteriano que morreu em 1761, bem antes da era da Internet, pode hoje ajudar a identificar, filtrar e classificar os *spams*? A idéia é simples, apesar de necessitar de algum conhecimento especializado para ser implementada, não é algo tão difícil de ser entendida: armazene um grande banco de dados de e-mails reconhecidamente *spam*, e em seguida, classifique-os. Como se não bastasse o benefício lógico de classificar as mensagens que não se quer ler, o sistema também pode ser realimentado com novas mensagens, aumentando ainda mais a base de dados e a precisão do sistema.

3.2.1 Rev. Thomas Bayes

Nascido em Londres, Thomas Bayes (1702-1761) foi educado por um tutor, algo que parecia necessário para o filho de um reverendo não-conformista naquela época. Nada é conhecido sobre o seu tutor, mas pesquisas admitem a possibilidade de ele ter sido aluno do De Moivre que, certamente, dava aulas particulares (era um tutor) em Londres naquela época. Thomas Bayes foi ordenado reverendo não-conformista como seu pai, e

primeiramente assistiu seu pai em Holborn. Mais tarde em 1720 veio a ser ministro do Presbyterian Chapel in Tunbrige Wells, próximo a Londres. Bayes, aparentemente, tentou retirar-se do ministério em 1749, mas continuou ministro em Tunbridge Wells até 1752, quando se aposentou, mesmo continuando a morar em Tunbridge Wells. Bayes divulgou sua teoria de probabilidade em *Essay towards solving a problem in the doctrine of chances,* publicado no *Philosophical Transactions of the Royal Society of London* em 1764. O trabalho citado foi enviado para Royal Society por Richard Price, um amigo de Bayes, o qual observou sobre o trabalho do reverendo:

> *"...eu envio a vocês um ensaio que eu encontrei entre os papers de nosso amigo Sr. Bayes, e que, na minha opinião, tem grande mérito... Na introdução que ele escreveu para este ensaio, ele disse ser este um projeto de um primeiro pensamento em subjetividade, para encontrar um método pelo qual nós possamos julgar (relativa) a probabilidade de um evento ocorrer, em dadas circunstâncias, sobre suposição que nós não sabemos nada a respeito dele e que, sobre algumas circunstâncias, ele ocorre um número certo de vezes, e falha outro número certo de vezes"*[1].

Por causa desse artigo, o seu nome é hoje sinônimo de toda uma filosofia na estatística e o principal benefício desse trabalho é o famoso Teorema de *Bayes*. O Teorema é um modo de calcular a probabilidade de que um evento irá ocorrer baseando no número de vezes em que o evento ocorreu anteriormente **[27]**.

Mesmo sem ter publicado nenhum trabalho com seu nome, em 1742, Thomas Bayes foi eleito membro da Royal Society of London.

---

1. O texto completo encontra-se na University of St Andrews, Scotland, School of Mathematics and Statistics no endereço: http://www-history.mcs.st-and.ac.uk/history/Mathematicians/Bayes.html

3.2.2 Teorema de *Bayes*

O Teorema de *Bayes* é usado na inferência de estatística para atualizar estimativas da probabilidade de que diferentes hipóteses sejam verdadeiras, baseando-se nas observações e no conhecimento de como essas observações se relacionam com as hipóteses. O Teorema de *Bayes* pode ser representado da seguinte forma:

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

O teorema determina que, para eventos **X** e **Y**, a probabilidade de **X**, dado que **Y** tenha acontecido denotado por **p(X|Y)** é igual à probabilidade de **Y,** dado que **X** denotado por **p(Y|X)** tenha acontecido, vezes a probabilidade de **X** acontecer **p(X)**, dividido pela probabilidade de **Y** acontecer **p(Y) [27]**.

Este teorema é uma das pedras angulares da estatística das probabilidades combinadas e é largamente utilizado em áreas à primeira vista pouco relacionadas, como Medicina e Informática.

Na primeira, o paradigma embasado em evidências é todo construído em cima do Teorema de *Bayes*. Baseado na experiência acumulada de exames e testes para tentar diagnosticar uma doença, o médico enquadra seus pacientes e pode estimar qual a probabilidade de que uma dada doença esteja se manifestando. Ou seja, dada uma probabilidade inicial (por exemplo, o paciente é fumante) e aplicado um exame em que, se sabe, há uma probabilidade de falso-positivos e falso-negativos (por exemplo, uma biópsia de pulmão), o médico sabe qual a probabilidade resultante de aquele paciente ter a doença (por exemplo, câncer de pulmão).

Na informática, muitos dos sistemas de classificação automática são baseados no Teorema de *Bayes*. Inicialmente, o sistema é treinado aceitando entrada de humanos que dizem que uma dada entrada pertence a determinado grupo. Com o tempo o sistema acumula um grande banco dessas informações e, aplicando o Teorema de *Bayes*, consegue estimar a probabilidade de cada novo dado de pertencer a cada grupo já classificado **[28]**.

3.3 Redes *Bayesianas*

As redes *Bayesianas* foram desenvolvidas nos anos 70 com o objetivo de modelar processamento distribuído na compreensão da leitura, onde as expectativas semânticas e evidências perceptivas deveriam ser combinadas para formar uma interpretação coerente. A habilidade para coordenar inferências bidirecionais preencheu uma lacuna na tecnologia de sistemas especialistas no início dos anos 80, e as redes *Bayesianas* têm emergido como um esquema de representação genérico para conhecimento incerto. Uma rede *Bayesiana* é um grafo direcionado acíclico onde os nós representam as variáveis (de interesse) de um domínio e os arcos representam a dependência condicional ou informativa entre as variáveis. A força da dependência é representada por probabilidades condicionais que são associadas a cada grupo de nós pais-filhos na rede **[29]**.

3.3.1 Redes *Bayesianas* Reconhecendo *Spam*

O casamento de padrões tem se mostrado uma alternativa mais promissora. Inicialmente, a filtragem de *spam* era feita por *scripts* para alguns (MTAs); atualmente técnicas têm sido experimentadas com relação ao casamento de padrões para *spam*. Algumas técnicas utilizam uma base de dados única, mantida por um conjunto de pessoas, essas estratégicas têm se mostrado ineficazes devido à rápida adaptação dos s*pammers* aos novos padrões detectados **[30]**.

Embora ainda não se tenha um mecanismo computacional suficientemente eficiente para barrar o *spam*, é sabido que ele pode ser facilmente identificado com sucesso por seres humanos. Portanto, pode-se imaginar que o tipo de problema a ser resolvido é um problema de inteligência artificial. A inteligência artificial é um ramo da Ciência da Computação interessado na automação de comportamento inteligente.

Uma das características que torna difícil a detecção automática de s*pam* é a grande capacidade de adaptação das pessoas que os distribuem na Internet (*spammers*). Na Figura 3.1 é apresentado um caso de s*pam* “adaptado” para escapar dos filtros tradicionais, que não possuem capacidade de aprendizado. Neste exemplo, as palavras são “cortadas” por comentários introduzidos apenas para despistar os mecanismos *anti-spam*.

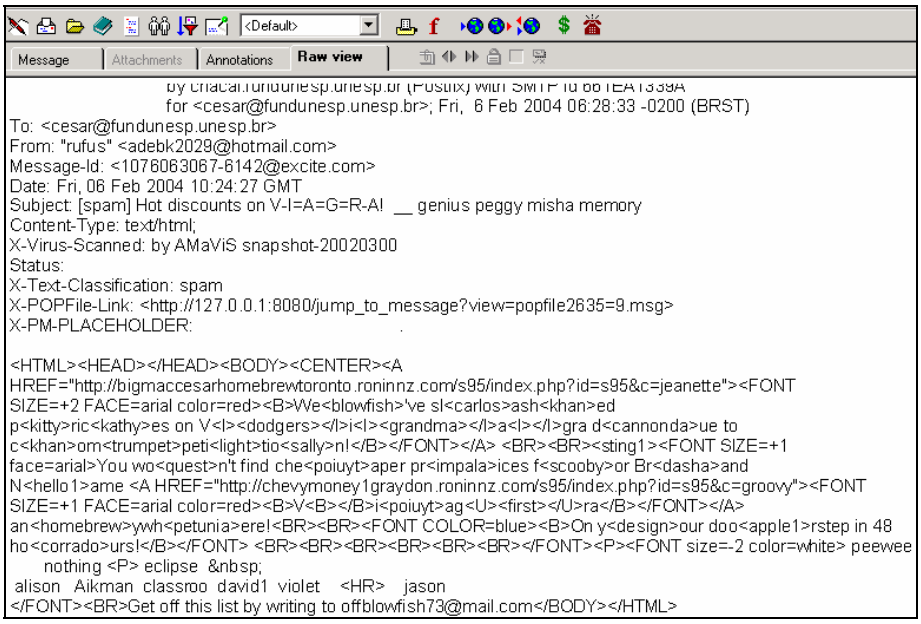

Figura 3.1: *Spam* HTML adaptado

A Figura 3.2 mostra o exemplo apresentado sem as *tags* HTML, da forma como ela é mostrada para o usuário. Nesse caso, um mecanismo *anti-spam,* que trabalhe sobre a mensagem em HTML, deve ter a capacidade de "aprender" as *tags* introduzidas nos comentários pelos *spammers*.

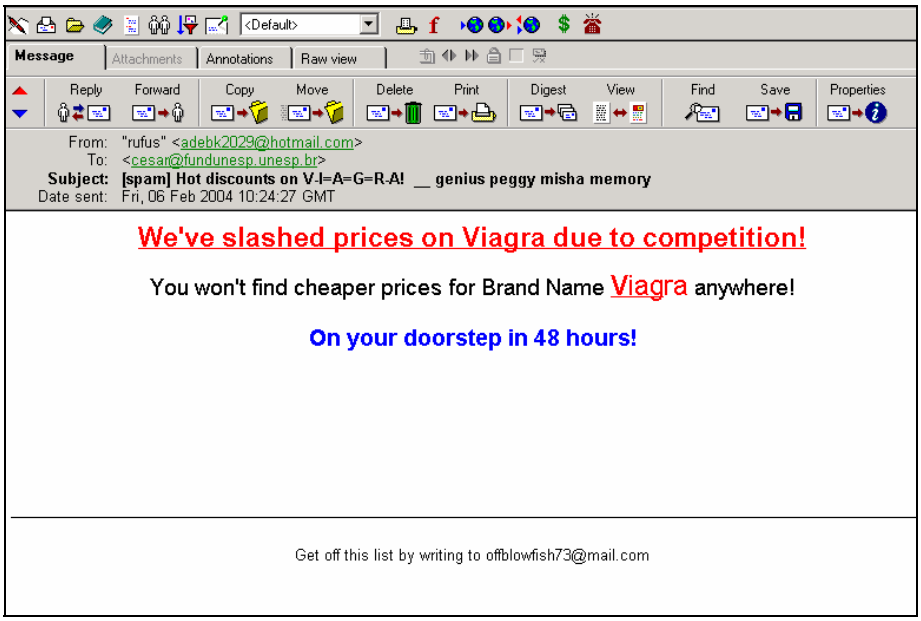

Figura 3.2: *Spam* HTML adaptado transformado em texto

Um problema também comum aos programas de detecção de *spam* é o alto índice de falso-positivos muitas vezes apresentado. A ocorrência de falso-positivos pode acabar se

tornando um problema maior que o próprio s*pam*, pois detectar como *spam* um e-mail importante pode causar grandes transtornos ao destinatário da mensagem.

A detecção de s*pam*, portanto, passa a ser um caso tipo de casamento de padrões, um problema que pode ser atacado por diversas técnicas de Inteligência Artificial. Uma dessas formas é através do uso de redes *Bayesianas*, aplicadas à classificação de dados. Uma forma de redes *Bayesianas* comumente aplicada a esse problema são as redes *Bayesianas* do tipo *Naive* ou ingênuo.

A seguir, serão apresentadas sucintamente as redes *Bayesianas* do tipo *Naive*, assim como a forma de aplicação desse tipo de rede ao problema de detecção de s*pam*.

3.3.2 Método *Naive Bayes*

Classificação de textos é a tarefa de atribuir automaticamente um texto a uma ou mais categorias pré-definidas. Enquanto mais e mais informações textuais estão disponíveis *online*, a busca efetiva de informação relevante é difícil sem a devida indexação e sumarização do conteúdo de documentos. A aplicação de classificação de textos resolve esse problema. Um crescente número de métodos estatísticos de classificação e técnicas de aprendizado de máquinas tem sido aplicado à classificação de textos nos últimos anos. Os métodos de classificação são os procedimentos que efetivamente classificam o documento em determinada classe **[31]**.

A classificação de *Naive Bayes* é feita utilizando dados de treinamento para estimar a probabilidade de o documento pertencer a cada classe. São utilizados os termos do documento com seus respectivos pesos para realizar a classificação. Para cada termo do documento é calculada a probabilidade de o mesmo pertencer à categoria **[32]**. É feita uma combinação das probabilidades levando em consideração o peso dos termos de acordo com a regra de *Bayes*. Se o resultado for maior que determinado coeficiente, o documento é incluído na categoria.

No caso de detecção de s*pam*, o classificador deve identificar s*pam* e *não-spam*. Para isto, o treinamento é feito através da apresentação de mensagens consideradas *spam* e mensagens consideradas *não-spam*, a partir dessa etapa de aprendizagem pode ser montada uma base de dados com as palavras (*tokens*) encontradas, e qual a incidência de cada uma. São montadas listas de *tokens* bons (encontrados em *não-spams*) e *tokens* ruins

(encontrados em s*pams*). A partir dessas listas são montadas as distribuições de probabilidades que permitem o cálculo de probabilidades na apresentação de uma mensagem qualquer (ver Figura 3.3):

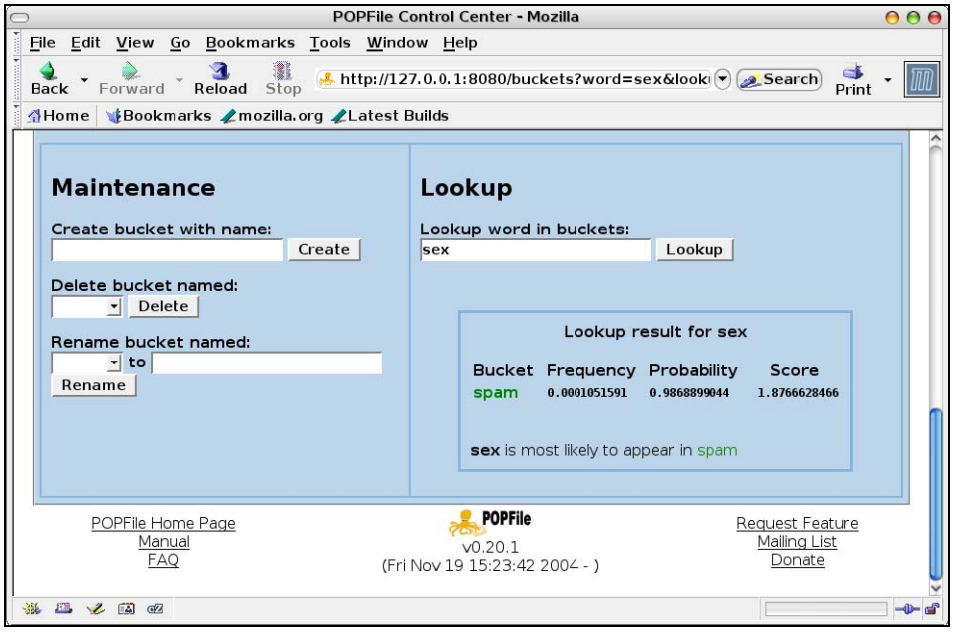

Figura 3.3: Procura por palavras nos baldes (*buckets*)

De acordo com a Figura 3.3, pode-se observar que a palavra "*sex*" é mais provável de aparecer nas mensagens com conteúdo de *spam* e, por conta disso, recebe nota elevada como 0.98% de probabilidade da palavra pertencer à base de dados de *spams*.

Como visto na figura acima o software *POPfile* permite visualizar a probabilidade de cada palavra através de uma interface *Web* e, além de fornecer informações sobre a probabilidade ele também apresenta dados como a freqüência em que a palavra ocorre e a pontuação (*score*) da palavra no presente momento.

3.4 Prevenção contra *spam*

Segundo Teixeira **[13]**, não existe uma receita milagrosa capaz de solucionar todos os problemas relacionados a *spam*. No entanto, alguns cuidados devem ser tomados pelo Administrador da Rede, enquanto outros devem ser tomados pelo usuário final.

3.4.1 Recomendações ao Administrador

Faz parte das atribuições do Administrador da rede tratar os casos de *spam* originados ou destinados à rede sob sua responsabilidade. Assim, algumas recomendações imprescindíveis são: fazer a configuração correta de seus servidores para não ser conivente com o envio de *spam*; cuidar das configurações capazes de reduzir o volume de *spam* recebido e educar os usuários sobre como lidar com *spam* e não ser um *spammer*.

Abaixo estão relacionadas algumas medidas preventivas para minimizar o recebimento de mensagens não solicitadas:

Relay: Um servidor de correio eletrônico atua como *relay* quando ele processa um e-mail, sendo que nem o remetente, nem o destinatário são usuários do servidor em questão. Servidores de correio que permitem *relay* já foram usados na Internet de maneira válida. No entanto, atualmente, constituem uma ameaça na rede, pois são usados pelos *spammers* para disparar seus "*junk e-mail*" indiscriminadamente. O uso de servidores como *relay* permite ao *spammers* aumentar o envio desses e-mails, driblar filtros, despistar sua verdadeira identidade e sem pagar nada por essa atitude. As versões mais recentes dos pacotes para servidor de e-mail são *anti-relay*. O site http://www.abuse.net/relay.html permite testar se o servidor está com *relay* aberto.

Filtros: Dentre as alternativas para filtrar e-mails indesejáveis, tem-se: a definição de *Blackhole Lists* no SMTP *server*, como por exemplo no *Sendmail* e *Postfix*. Com esse recurso o servidor rejeita os e-mails originados de potenciais redes ou usuários *spammers*, pré-definidos numa lista. Pode-se definir filtros no cliente de e-mail também. Utilitários como o Eudora e o Pine, por exemplo, possuem essa funcionalidade. Alguns administradores questionam o uso de filtros na máquina do usuário final, argumentando que o *spam* já atingiu parte de seu objetivo, pois já desperdiçou recursos do servidor e banda do provedor. No entanto, ainda assim é uma alternativa a se considerar. O grande problema com a utilização de filtros é o cuidado em não rejeitar e-mails válidos. Portanto, é recomendável que sejam usados filtros em casos específicos e com a devida autorização da empresa.

Listas da ORDB e MAPS: Existem duas entidades na Internet que mantêm bases de dados de servidores de e-mail que permitem *relay*: ORDB (*Open Relay Database*) e o MAPS (*Mail Abuse Prevention System*). Uma prática recomendada aos administradores é a

configuração de seus servidores de e-mail para rejeitar e-mails originados das redes listadas nessas duas bases de dados. A experiência tem demonstrado que o uso desse recurso reduz significativamente os problemas com *spam* através de *spam relay*.

Educação e conscientização dos usuários: A educação continua sendo a melhor alternativa. Educar e conscientizar os usuários de sua rede sobre como lidar com *spam*, como reclamar, a quem recorrer, como não colaborar com *spam* na rede e por que não enviar *spam* são as principais recomendações ao usuário.

Spam e Políticas de segurança: As políticas de uso aceitável da rede devem ter normas claramente definidas para casos de *spam*, para que se tenha como advertir e até punir o usuário que não seguir as regras estabelecidas. Tais políticas devem prever desde advertências em caso de mau uso da conta de e-mail, até o cancelamento da mesma em casos recorrentes de *spam* enviados pelo mesmo usuário.

RFC 2142: O RFC 2142, *Mailbox Names for Common Services, Roles and Functions* recomenda os *aliases* básicos necessários para garantir a comunicação entre as inúmeras redes na Internet **[33]**. Os principais *aliases* recomendados, relacionados com incidentes de segurança e abusos na rede são: *abuse@dominio*, *postmaster@dominio* e *security@dominio*. Todo bom administrador deve ter implementado os *aliases* mencionados, ler e responder as notificações recebidas através deles.

Envio de reclamações: Enviar reclamações, exigindo providências aos responsáveis pelo *spam* ou por uso de *relay* é, principalmente, tarefa do administrador.

3.4.2 Recomendações ao Usuário

O número de usuários na Internet cresce assustadoramente a cada minuto, sendo que muitos estão aprendendo a viver ou sobreviver nessa "aldeia global". Assim, cabe ao administrador de rede conscientizar seus usuários sobre regras, dicas e cuidados que devem ser seguidos para melhor conviver no mundo virtual. Como agir diante do recebimento de *spam*, como não incentivar o surgimento de *spam*, ou ainda, cuidados para não se tornar um *spammer*. A seguir, são listados alguns conselhos básicos aos usuários:

Siga a Netiqueta: *e*mbora a filosofia da Internet seja um tanto quanto anárquica, existem algumas regras básicas de bom comportamento na rede. Algo como as regras de boa educação para viver em sociedade: "por favor", "obrigado", "com licença", "não

gritar" e assim por diante. O RFC 1855 (*Netiquette Guidelines*), que trata da Netiqueta, pode parecer antigo por ser de 1995, mas ainda é muito adequado, principalmente com relação à comunicação por e-mail e WWW **[34]**.

Não repasse boatos ou correntes: Verifique sempre a veracidade de uma determinada mensagem antes de repassá-la. Na dúvida, não repasse. Existem casos de funcionários demitidos por justa causa e processados por repassarem boatos. Quando decidir repassar mensagens desse tipo, mesmo após certificar-se da veracidade da mesma, restrinja ao máximo os destinatários e pense sempre se seus amigos estariam realmente interessados em receber tal informação: cuidado para não se transformar em um *spammer*. A regra básica é: fuja das correntes e fique atento aos boatos.

Não caia em* contos do vigário *do tipo "remove me": Fique atento ao conteúdo dos *spams* recebidos e não seja ingênuo, não caia nos artifícios usados pelos *spammers*. Esse é um dos artifícios mais freqüentes usados atualmente. São os *spams* do tipo "*remove me*". Não responda! Na verdade, ao responder o usuário estará confirmando a legitimidade de seu e-mail e este possivelmente será inserido em listas de *spammers* pelo mundo afora.

Nunca responda para um spammer, nem se envolva em discussões com o mesmo. Isto gera mais Spam!: A*o* receber um s*pam*, entre em contato com os administradores da rede local para as devidas providências. Não responda ou tente reclamar diretamente ao s*pammer*, caso seja possível identificá-lo no e-mail. Afinal, um s*pammer* convicto poderá gerar algum esquema de retaliação que só fará piorar a situação.

Não tente revidar, atacando o spammer, este tipo de retaliação não funciona: A idéia de "olho por olho, dente por dente" não se aplica nesse contexto. Não tente revidar a perturbação ou até mesmo o ataque recebido de um *spammer*. Esse tipo de atitude não é ética, não é recomendável e não vai resolver o problema. Se o usuário decidir retaliar um s*pam*, usando o mesmo método, lembre-se que estará se tornando um s*pammer*. Além disto, existem várias maneiras de se forjar um e-mail de s*pam* e, portanto, você está arriscando a retaliar o domínio errado. Finalmente, a retaliação estará atraindo mais atenção e publicidade para o *spammer*: tudo que ele mais queria.

Cuidado de higiene com seu(s) e-mail(s): evite se cadastrar em sites que prometem não divulgar seus dados. Evite se cadastrar em vários sites e listas de divulgação de atualizações de informação. Caso sua postura pessoal seja de um internauta ávido por

informações e que gosta de receber malas diretas, divulgação de sites, etc, então, uma prática recomendada e muito utilizada é manter contas de e-mail separadas para seus interesses pessoais, fora do ambiente de trabalho. Isso pode não solucionar o problema, mas ajuda a minimizá-lo.

Filtros: Alguns programas de clientes de e-mail apresentam funcionalidades que permitem filtrar e-mails de *spam*. Novamente, tais funcionalidades não resolvem todos os problemas, mas podem driblar um pouco a questão, diminuindo o volume de "*junk e-mail*" em sua caixa postal. Lembre-se sempre de relatar ao administrador de sua rede o recebimento de *spams*, ele poderá incrementar a política de defesa contra *spam* da rede como um todo.

3.5 Conclusão

Este capítulo apresentou as técnicas e métodos de prevenção *anti-spam*, introduzindo o conceito de filtros para fazer a classificação das mensagens. Através dessa revisão, notou-se a busca constante por soluções eficazes para controlar o recebimento de *spam*. A adoção das melhores técnicas representa um marco importante na tentativa de minimizar o volume de mensagens não solicitadas.

Capítulo 4

Ferramentas *Bayesianas Anti-Spam*

Este capítulo analisa algumas das principais ferramentas a*nti-spam* existentes usando filtros *Bayesianos* e as compara tanto às taxas de falso-positivos quanto às taxas de falso-negativos entre as ferramentas avaliadas.

4.1 Ambiente de Teste

O seguinte ambiente de teste foi montado para a realização da filtragem de *spam*: foi utilizado o pacote *Postfix* como servidor de e-mail (MTA) em uma máquina *Red Hat Linux* sem nenhum filtro de e-mail para coletar mensagens boas e ruins para início dos treinamentos e para a realização dos testes foi necessário arquivar um conjunto de mensagens e depois entregá-las a cada nova ferramenta a ser avaliada.

Buscando a eficiência das ferramentas foi criada uma conta de e-mail e divulgada em vários *sites* na Internet. Esse aspecto foi fundamental para que a conta de e-mail em questão recebesse muitas mensagens e as ferramentas pudessem ser avaliadas no quesito de acertos e erros na identificação de *spam*.

Devido ao fato de que as ferramentas *Bayesianas anti-spam* necessitam de treinamento inicial, foi preciso treiná-las, ou seja, 200 mensagens consideradas *spam* e 200 consideradas *não-spam* foram atribuídas às ferramentas antes de começar a filtragem das mensagens.

Portanto, para manter o formalismo quanto às mensagens utilizadas para treinamento inicial dos filtros, pode-se verificar nas figuras abaixo o arquivamento das mensagens utilizadas no treinamento. Através desse mecanismo, as ferramentas aumentam a precisão na classificação das mensagens (ver Figuras 4.1 e 4.2):

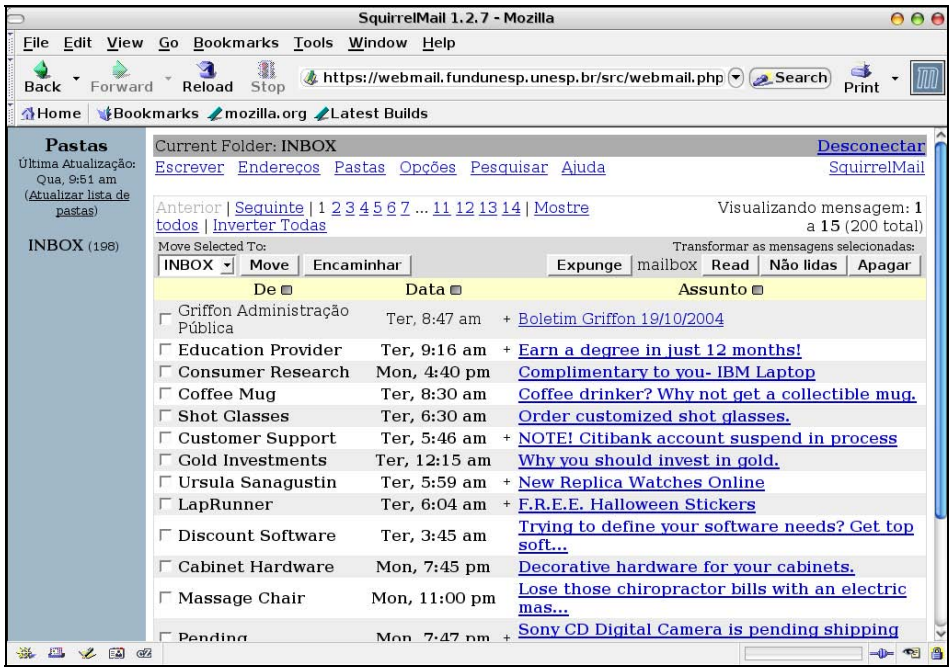

Figura 4.1: Treinamento inicial (200 mensagens *spam*)

De acordo com a Figura 4.1, pode-se observar no lado superior direito que foram coletadas em Outubro de 2004 um total de 200 mensagens consideradas *spam* para início dos treinamentos. Já a Figura 4.2, mostra as mensagens *não-spam*.

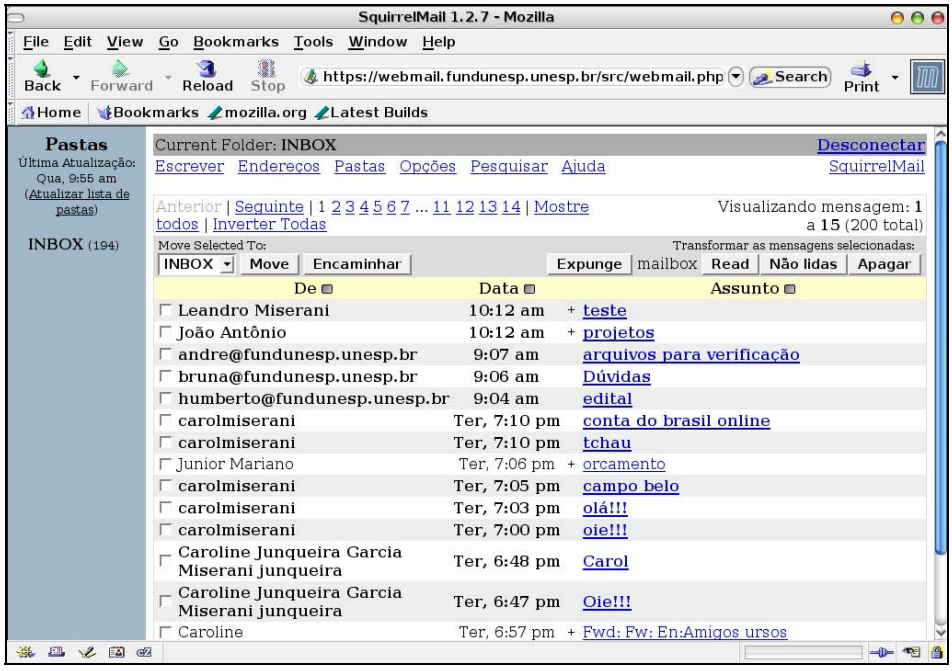

Figura 4.2: Treinamento inicial (200 mensagens *não-spam*)

Observando a Figura 4.2, nota-se que foram coletadas também um total de 200 mensagens consideradas *não-spam* no mesmo mês (Outubro de 2004), pois os filtros *Bayesianos* necessitam de treinamento inicial para melhorar a precisão na classificação das mensagens.

Após o treinamento inicial, começou-se então o arquivamento de um conjunto de mensagens para a realização dos testes. As mensagens foram arquivadas no período de outubro a novembro de 2004, totalizando 1136 mensagens onde, 1048 são consideradas *spams* e 88 são consideradas *não-spams* (ver Figura 4.3):

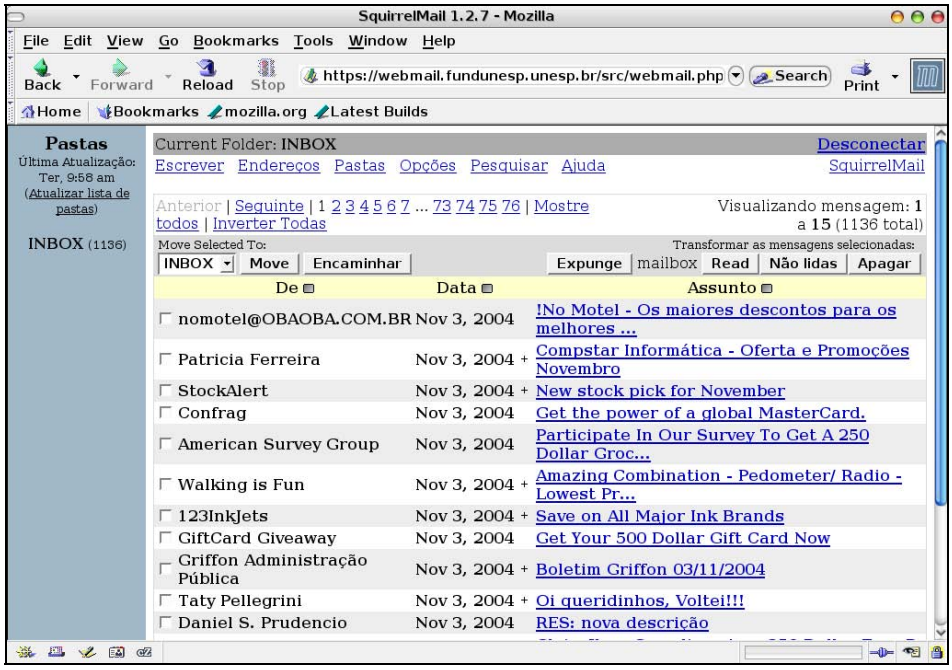

Figura 4.3: Mensagens utilizadas nos testes

A Figura 4.3 ilustra o mesmo conjunto de mensagens a serem processadas por todas as ferramentas avaliadas neste trabalho. Vale ressaltar que as mensagens utilizadas para treinamento não foram utilizadas nos testes, pois elas poderiam influenciar na acurácia das ferramentas.

As máquinas utilizadas nos testes foram: Intel Pentium III 800 MHz com 1GB de memória RAM utilizando *Red Hat Linux* (servidor de e-mail - MTA) e uma estação de trabalho Intel Pentium IV de 2.4 GHz com 512 MB de memória RAM utilizando *Gentoo Linux.*

O objetivo dos testes foi verificar a acurácia das ferramentas *anti-spam* na classificação das mensagens, observando tanto os casos de falso-positivos quanto falso-negativos.

4.2 Critérios de Seleção

A partir de 2002, as ferramentas *anti-spam* começaram a tomar um rumo semelhante à trajetória dos *anti-vírus* e *firewalls* **[24]**, ou seja, se transformando em produtos, comerciais em muitos casos, e necessários para a sobrevivência em virtude da enxurrada crescente de *spam*.

Pode-se afirmar que as ferramentas disponíveis hoje se encontram em avançado estágio de desenvolvimento e possuem recursos sofisticados, permitindo que o usuário leigo ou mesmo o especialista em ferramentas *anti-spam* implementem formas de combate ao *spam*. Para escolher quais ferramentas *anti-spam* seriam avaliadas, foram adotados os seguintes critérios:

- **Classificação em número de *downloads*:** foi dada preferência por ferramentas que são amplamente utilizadas e difundidas;
- **Classificação em satisfação na utilização:** preferência por ferramentas que são bem avaliadas pelos usuários;
- **Coleta de informações em listas de discussão:** através dessas listas é possível obter as ferramentas que estão sendo utilizadas com maior freqüência;
- **Escolha de ferramentas multiplataforma:** preferência por ferramentas que não são específicas para um único sistema operacional;
- **Opção por gratuidade:** preferência por ferramentas que sejam gratuitas.

4.3 Ferramentas *Bayesianas* Avaliadas

Muito antes do artigo clássico de Paul Graham "*A Plan for spam*" sobre filtros *Bayesianos anti-spam* ter sido publicado, algumas pessoas já estavam testando classificadores de mensagens usando o Teorema de *Bayes*, mas foi a partir desse artigo que o mundo percebeu a utilidade de tal classificação aplicada aos *spams* **[35]**. Muitas ferramentas surgiram nesse meio tempo. Algumas utilizam o algoritmo descrito por Graham e outras desenvolveram o seu próprio algoritmo.

Com base nos critérios de seleção, foram escolhidas para teste e análise algumas ferramentas *Bayesianas anti-spam*. Basicamente, existem dois tipos de software que podem ser utilizados para bloquear *spam*: aqueles instalados nos *Servidores de e-mail* (MTA), que filtram os e-mails antes que cheguem até o usuário e aqueles instalados nos computadores dos *usuários* (MUA), que filtram os e-mails com base em regras individuais de cada usuário. Enfim, existem ferramentas para todos os gostos. As mais importantes são apresentadas nas seções subseqüentes.

4.3.1 Software *Anti-Spam* para (MTA)

Esta categoria de filtros atua em conjunto com o servidor de e-mail (MTA) e sua principal vantagem é filtrar o *spam* logo na sua chegada. Dentre os pacotes disponíveis em domínio público e desenvolvidos com a finalidade básica de atuar junto aos (MTAs), tem-se:

4.3.1.1 *Bogofilter*

O *Bogofilter* é um software livre para filtragem de *spam* escrito em C (linguagem de programação), que emprega um mecanismo *Bayesiano* para detecção de *spam* baseado no artigo "*A Plan for Spam*" escrito por Paul Graham em agosto de 2002, ou seja, ele classifica uma mensagem como *spam* ou não, por análises estatísticas no cabeçalho e corpo da mensagem **[36]**.

A versão utilizada foi o *Bogofilter 0.92.2* que suporta diferentes plataformas: *Linux*, *FreeBSD*, *Solaris*, *Mac OS X*, *HP-UX*, dentre outros. Ele pode funcionar diretamente com o

servidor de e-mail (MTA) evitando que os usuários tenham que tomar suas próprias providências.

Na Figura 4.4, o *Bogofilter* no momento da filtragem adiciona no cabeçalho da mensagem o campo ***X-Bogosity***, o qual informa se a mensagem foi considerada *spam*, e com que probabilidade, conforme mostra o exemplo abaixo:

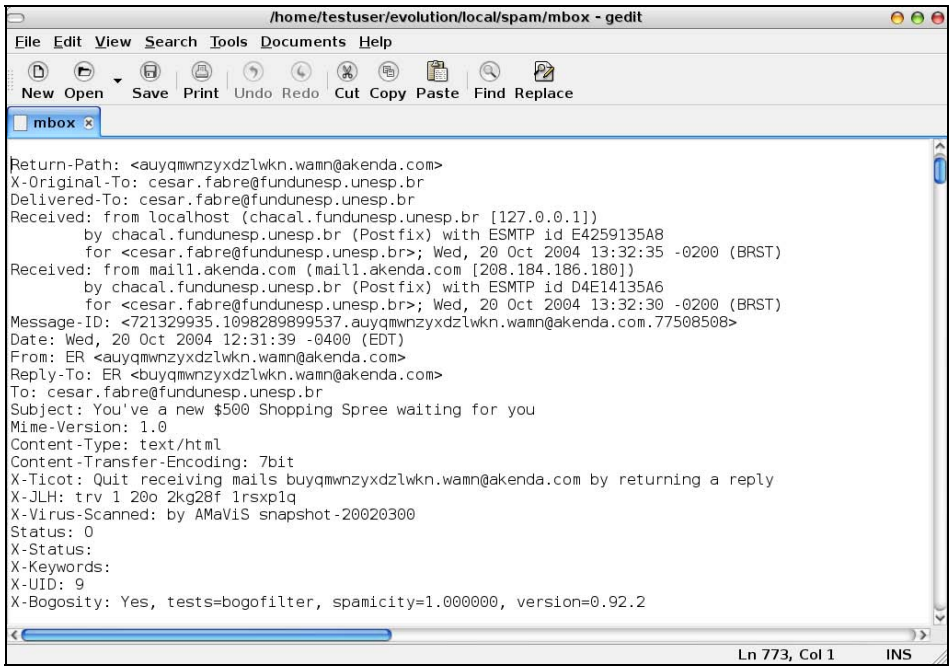

Figura 4.4: *Bogofilter* adicionando o cabeçalho *X-Bogosity*

Já a Figura 4.5 apresenta o cliente de e-mail *Ximian Evolution* com suas funcionalidades disponíveis para trabalhar em conjunto como o *Bogofilter*, basta ativá-la e usá-la.

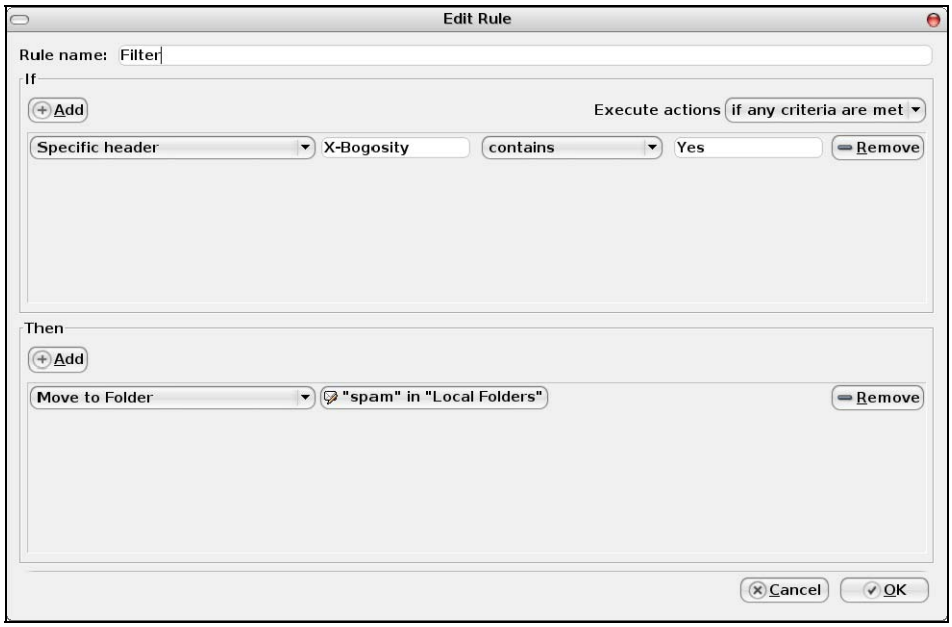

Figura 4.5: *Ximian Evolution* interagindo com o *Bogofilter*

Assim, os e-mails recebidos devem passar primeiramente pelo *Bogofilter* para que sejam devidamente marcados como *spam* ou *não-spam* e encaminhados para um cliente de e-mail, nesse caso, o *Ximian Evolution* que contém regras de filtragem para poder direcionar o e-mail considerado *spam* pelo *Bogofilter* para a pasta de "*spam*" (como mostra a Figura 4.5).

No início é preciso treiná-lo, pois o objetivo é construir a base de dados de mensagens boas e ruins. Após o treinamento inicial pode-se começar a filtragem das mensagens.

A Figura 4.6 mostra com detalhes os resultados obtidos com a união do *Bogofilter* e o cliente de e-mail *Ximian Evolution*.

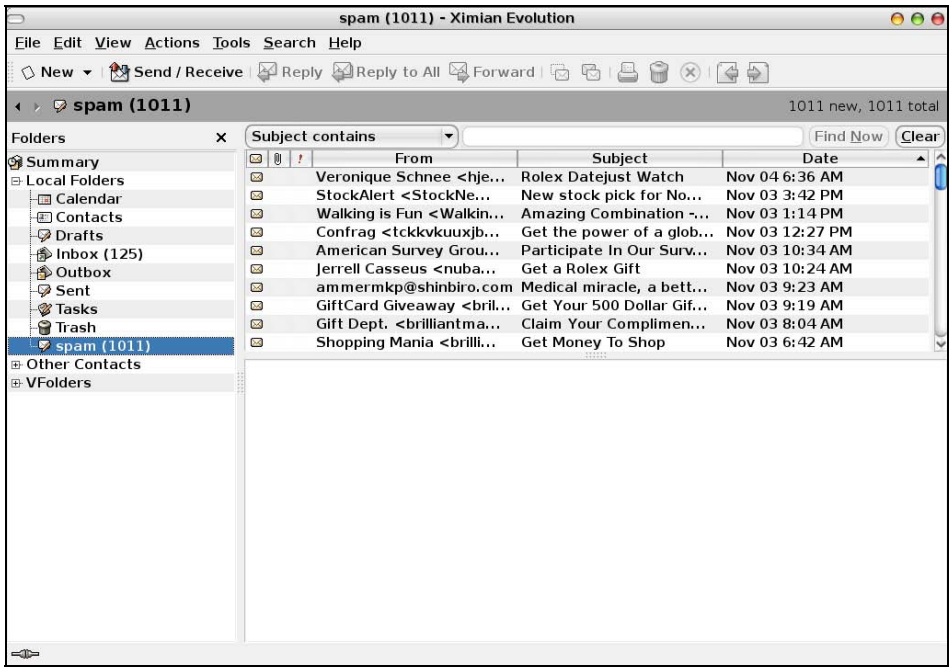

Figura 4.6: *Bogofilter + Ximian Evolution*

De acordo com a Figura 4.6, pode-se concluir que a união entre o *Bogofilter* e o *Ximian Evolution* resultou em números favoráveis. A Tabela 4.1 ilustra os resultados obtidos em relação ao número de mensagens classificadas:

	Vspam (Verdadeiro spam)	**VNspam (Verdadeiro não-spam)**	**Total**
spam	**1011**	**0**	**1011**
Nspam	**37**	**88**	**125**
	1048	**88**	**1136**

Tabela 4.1: Resultados obtidos com o *Bogofilter*

Observando a Tabela 4.1, nota-se que a pasta "*spam*" recebeu um total de 1011 mensagens, onde 1011 foram classificadas corretamente não permitindo a ocorrência de falso-positivos. Já a caixa de entrada ou "*inbox*" representada por "*Nspam, não-spam*" recebeu um total de 125 mensagens, onde 88 foram classificadas corretamente e 37 foram classificadas erradas permitindo a ocorrência de falso-negativos. O *Bogofilter* classificou um total de 1136 mensagens no período de outubro a novembro de 2004.

Tendo em vista os resultados obtidos, a acurácia nesta proporção é 96,7% de sucesso na classificação das mensagens.

Não se pode esquecer que, para funcionar corretamente, os filtros precisam de exemplos de *spam* e *não-spam*, e de mensagens recentes (eles devem ser constante ou periodicamente treinados). Usar a base de dados de outra pessoa (ou *site*) pode não funcionar, pois, se usar a base de dados de alguém que recebe muitas mensagens sobre o mercado financeiro, e o usuário receber muitas mensagens sobre música, vai estar indo contra a idéia fundamental do filtro, porque suas mensagens são estatisticamente diferentes das da outra pessoa.

A desvantagem do *Bogofilter* é pelo fato de ele recomendar muitas mensagens no início do treinamento para atingir índices favoráveis (cerca de 2000 mensagens boas e 2000 mensagens ruins). Vale ressaltar que neste trabalho foi adicionado aos filtros somente (200 mensagens boas e 200 mensagens ruins para treinamento inicial), mesmo assim, o *Bogofilter* gerou resultados satisfatórios e, além disso, o *Bogofilter* foi rápido e eficiente na classificação das mensagens.

4.3.1.2 *SpamAssassin*

O *SpamAssassin* é um software livre para filtragem de *spam* escrito em *Perl,* que interage com o servidor de e-mail (MTA) **[37]**.

A versão utilizada foi o *SpamAssassin 2.64* que suporta diferentes plataformas como: *Linux*, *FreeBSD*, *Solaris*, *HP-UX*, dentre outros. Ele aplica uma série de testes às mensagens, procurando por características que as identifiquem como *spam*. Entre esses testes, tem-se:

Análise de Cabeçalho: os *spammers* usam várias técnicas para mascarar suas identidades e esconder o servidor de origem das mensagens. O *SpamAssassin* tenta identificar indícios do uso desses truques;

Análise de Texto: o *spam* tem um estilo de texto próprio, geralmente destinado a lhe convencer de que um determinado produto ou serviço anunciado é uma oportunidade única na vida e que não deve ser desperdiçada, além de tentar lhe convencer de que está

recebendo essa mensagem porque se cadastrou em algum serviço ou porque um "amigo" o indicou. O *SpamAssassin* tenta identificar tal estilo, baseado em ocorrências comuns de palavras, frases, texto em MAIÚSCULAS ou E N T R E S P A Ç A D O, entre outros;

Listas Negras: o *SpamAssassin* suporta consulta a listas negras como "*mail-abuse.org* e *ordb.org*", e pode ignorar mensagens vindas de domínios reconhecidamente abusados por *spammers*;

Aprendizado na classificação: o *SpamAssassin* possui módulos para a utilização de filtros *Bayesianos*, ou seja, na classificação das mensagens faz uso de estatística e probabilidades. Para funcionar corretamente é necessário treiná-lo.

Razor: o *Vipul's Razor* é uma base de dados colaborativa para rastreamento de *spam*. Ela permite que um usuário reporte uma mensagem como *spam*, adicionando-a à base de dados do projeto, o que fará com que, automaticamente, todos dos outros usuários do *Razor* passem a ignorar a mensagem.

No total, há mais de uma centena de testes que são executados. Uma lista completa pode ser vista em http://spamassassin.apache.org/tests.html.

A mensagem recebe uma determinada nota para cada teste em que é reprovada. Essa nota varia conforme o teste, e pode até mesmo ser negativa, o que aumenta as chances da mensagem não ser um *spam*. Acima de uma pontuação limite, estipulada em 5 (cinco), por padrão, a mensagem é automaticamente considerada como *spam*.

A Figura 4.7 ilustra o comportamento do *SpamAssassin* no momento da classificação:

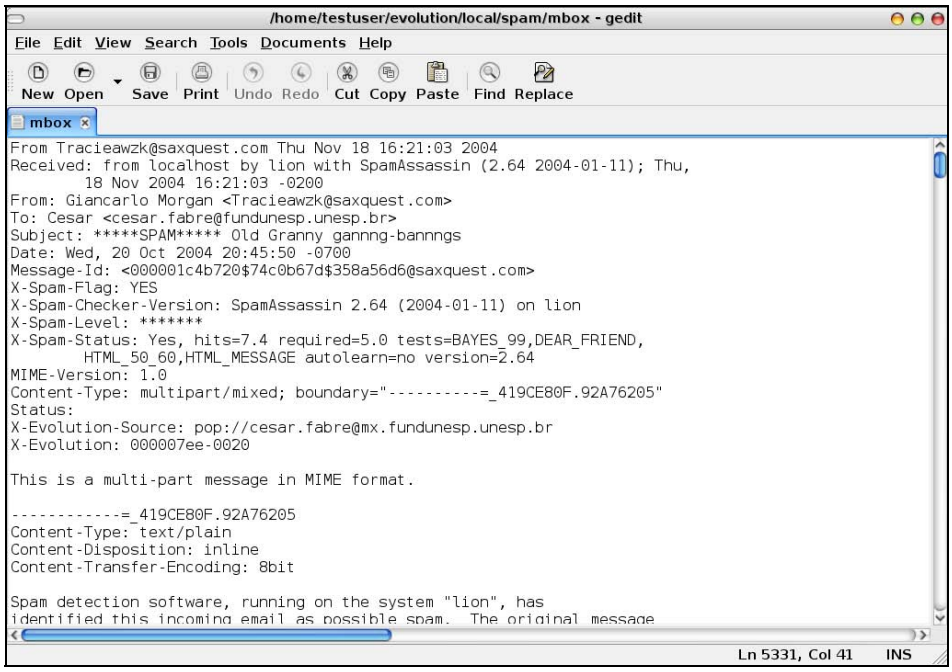

Figura 4.7: Cabeçalho do *SpamAssassin*

No cabeçalho da mensagem pode-se verificar o campo *X-Spam-Status*, indicando a mensagem como *spam* e a sua pontuação total nos testes; e, no seu início, um resumo dos testes executados e os resultados obtidos. O assunto (*Subject*) da mensagem também é modificado para conter o texto *****SPAM*****, o que facilita a filtragem no programa de e-mail (ver Figura 4.8):

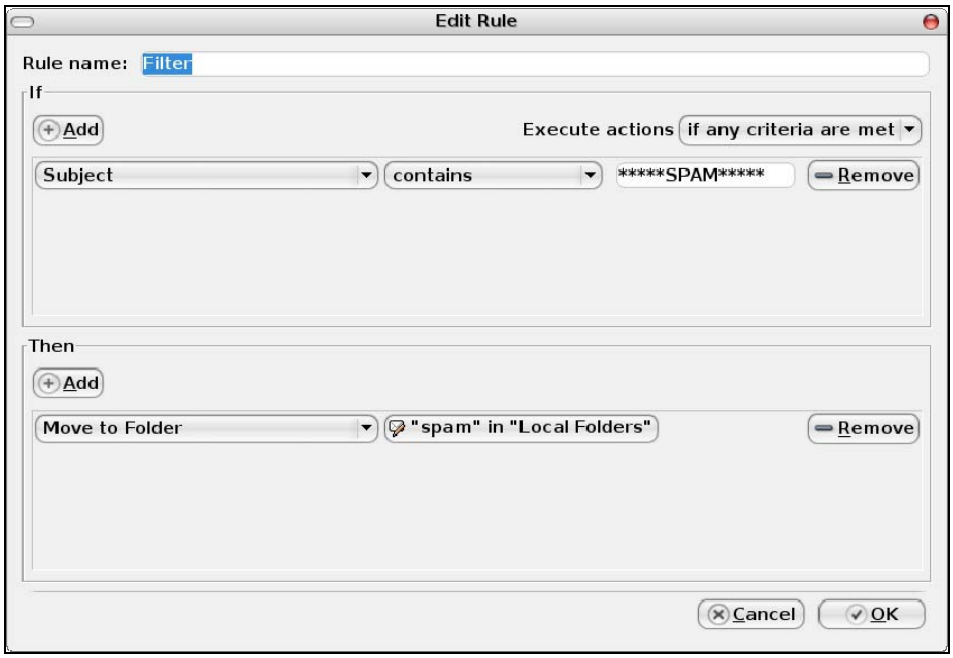

Figura 4.8: *Ximian Evolution* interagindo com o *SpamAssassin*

Agora é só checar os e-mails e esperar receber mais *spam*. Assim, as novas mensagens serão automaticamente movidas para a pasta indicada no filtro como mostra a figura acima. Poderia ser mais radical, em vez de mover a mensagem considerada *spam* para uma pasta, apagá-la automaticamente. Mas não é recomendável que faça isso logo de início por causa de um problema que pode acontecer com todo filtro de e-mail: falso-positivos (mensagens legítimas serem consideras *spam*).

Às vezes aquela mensagem em HTML, com fontes coloridas e o plano de fundo animado enviada por amigos ou aquele e-mail com as promoções de uma determinada loja virtual favorita podem acabar sendo confundidos com *spam*. Para evitar isso, é necessário criar exceções nos filtros do *SpamAssassin*, de modo que todas as mensagens que "casarem" com a exceção sejam ignoradas pelo programa.

A Figura 4.9 mostra com detalhes os resultados obtidos com a união do *SpamAssassin* e o cliente de e-mail *Ximian Evolution*.

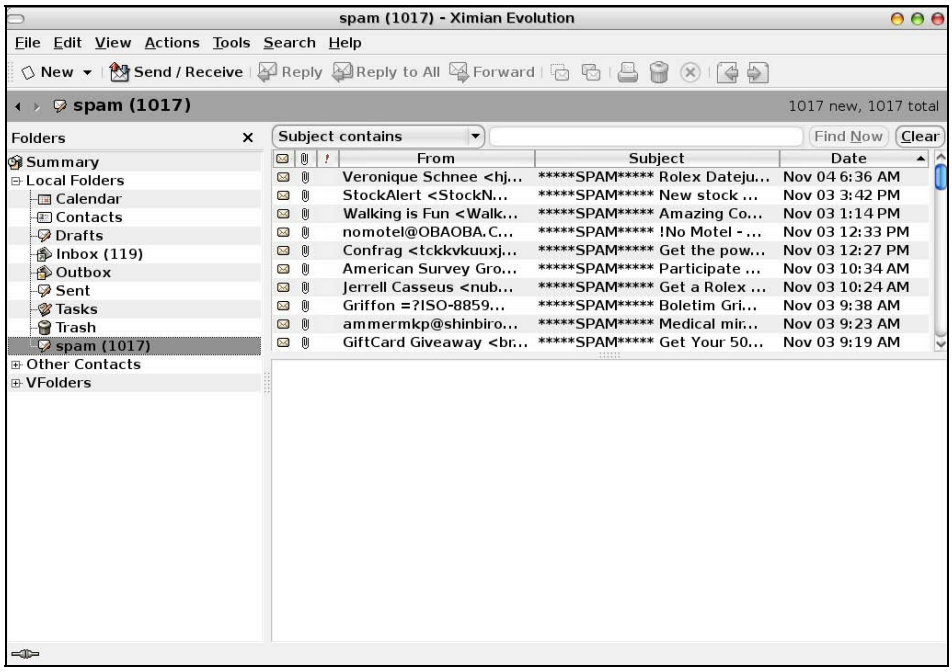

Figura 4.9: *SpamAssassin + Ximian Evolution*

De acordo com a Figura 4.9, pode-se concluir que essa união também resultou em números favoráveis e, comparando-se com o *Bogofilter*, pode-se observar que o filtro do *SpamAssassin* gerou resultados mais satisfatórios.

Através da Tabela 4.2, é possível verificar os resultados obtidos com o *SpamAssassim* em relação ao número de mensagens classificadas:

	Vspam (Verdadeiro spam)	**VNspam (Verdadeiro não-spam)**	**Total**
spam	**1017**	**0**	**1017**
Nspam	**31**	**88**	**119**
	1048	**88**	**1136**

Tabela 4.2: Resultados obtidos com o *SpamAssassin*

Observando a Tabela 4.2, nota-se que a pasta "*spam*" recebeu um total de 1017 mensagens, onde 1017 foram classificadas corretamente não permitindo a ocorrência de falso-positivos. Já a caixa de entrada ou "*inbox*" representada por "*Nspam, não-spam*" recebeu um total de 119 mensagens, onde 88 foram classificadas corretamente e 31 foram

classificadas erradas permitindo a ocorrência de falso-negativos. O *SpamAssassin* classificou um total de 1136 mensagens no período de outubro a novembro de 2004.

Tendo em vista os resultados obtidos, a acurácia nesta proporção é 97,3% de sucesso na classificação das mensagens.

Para funcionar corretamente, o filtro *Bayesiano* do *SpamAssassin* precisa de exemplos de *spam* (200 mensagens) e *não-spam* (200 mensagens). O *SpamAssassin* tem como função ser um filtro de e-mail que tenta identificar ocorrências de *spam* utilizando-se de análise de texto e várias *Blacklists* atuais baseadas na Internet.

Seu procedimento é fazer um amplo mecanismo de testes heurísticos, verificando sua base de dados própria, nos cabeçalhos (*headers*) e corpo das mensagens de e-mail para identificar *spam*, calculando o total de *hits* das regras. Baseado no valor dos *hits*, ele classifica a mensagem como *spam*.

O *SpamAssassin* também integra o conceito de filtros *Bayesianos*, oferecendo muitos recursos que deixam o filtro com maior precisão na classificação das mensagens **[38]**. Os filtros *bayesianos* do *SpamAssassin* também são baseados no artigo "*A Plan for Spam*" escrito por Paul Graham em agosto de 2002.

A desvantagem do *SpamAssassin* é o fato de ele realizar vários testes no momento da classificação das mensagens, com isso, a filtragem das mensagens acaba sendo mais lenta, mas por outro lado, mostrou-se um filtro eficiente. Esse filtro requer do usuário um pouco de experiência em sistemas UNIX / Linux.

4.3.2 Software *Anti-Spam* para (MUA)

Esta categoria de filtros atua em conjunto com o cliente de e-mail do usuário final (MUA). Nesse caso, alguns software (MUAs) como o *Eudora*, *Pegasus Mail* e o *Mozilla Mail* possuem mecanismos para filtragem de *spam*. Vale ressaltar que os filtros no (MUA) podem ser usados em conjunto com os filtros no servidor de e-mail, o que aumenta a eficiência. Quando se fala em software de filtragem no (MUA), um ponto a ser considerado é o conhecimento técnico do usuário, visto que na maioria das vezes ele estará responsável por configurar e adequar os seus filtros de e-mail. Assim, software de filtragem com

interfaces intuitivas e boa documentação se destacam. Abaixo, seguem algumas ferramentas para esse fim, disponíveis em domínio público:

4.3.2.1 *Mozilla Mail*

A Fundação *Mozilla* promove a inovação na Internet através do desenvolvimento do software *Mozilla*, da premiada suíte com *Browser* e e-mail, além de produtos relacionados e outras tecnologias livres **[39]**.

A versão utilizada foi o *Mozilla 1.6* que contém opções para controle de lixo "*junk e-mail*" de *spam*. A principal característica do filtro *anti-spam* do *Mozilla* é a utilização do conceito de redes *Bayesianas* escrita no artigo de Paul Graham "*A Plan for Spam*". A técnica de filtragem *Bayesiana* requer primeiro que o programa de correio eletrônico seja treinado, mostrando o e-mail considerado *spam* e o e-mail não considerado *spam* **[40]**.

A utilização de redes *Bayesianas* permite que o filtro "aprenda" o que é considerado *spam* e, com isso, ele vá se adaptando e melhorando a performance conforme receber novos e-mails.

Uma vez configurado corretamente, o *Mozilla Mail* consegue filtrar as novas mensagens de forma estatística através das características individuais de cada palavra.

Quando o filtro estiver "treinado" o suficiente, o usuário poderá ativar a opção que move a mensagem considerada *spam* para outra pasta, como por exemplo, uma pasta com o nome de "*spam*". A partir desta pasta é possível encontrar milhares de e-mails considerados lixo em um arquivo de *log* chamado "*junklog.html*", como mostra a Figura 4.10:

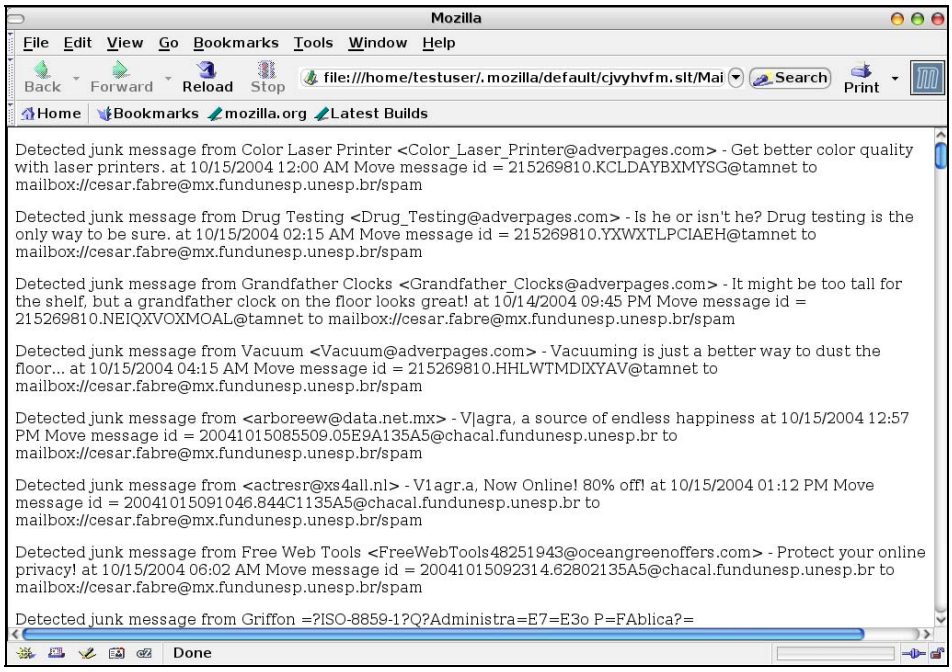

Figura 4.10: Arquivo "*junklog.html*" do *Mozilla Mail*

De acordo com a Figura 4.10, as mensagens filtradas como *spam* são movidas para a pasta "*spam*" e um *log* de lixo é automaticamente criado para cada mensagem considerada lixo no arquivo "*junklog.html*".

Para a realização dos testes foi preciso coletar mensagens boas e ruins no período de outubro a novembro de 2004. Vale ressaltar que as mensagens utilizadas nos treinamentos iniciais não foram aplicadas nos testes, ou seja, foram aplicadas mensagens diferentes para avaliar o aprendizado das ferramentas.

Os resultados foram gerados a partir de uma base de dados contendo cerca de 1136 mensagens, ou seja, na pasta "*inbox*" foram cerca de 115 mensagens e na pasta "*spam*" foram 1021 mensagens (ver Figura 4.11):

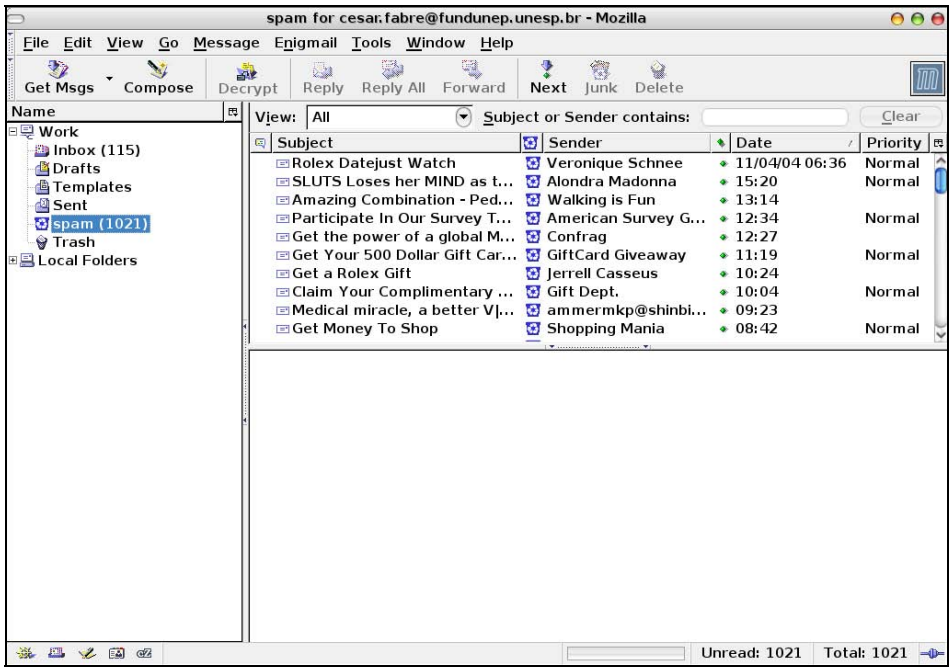

Figura 4.11: *Mozilla Mail* filtrando as mensagens

De acordo com a Figura 4.11, as mensagens classificadas como *spam* são movidas para a pasta "*spam*" e o total dessas mensagens podem ser vistos no menu inferior, que corresponde a 1021 mensagens do tipo lixo. A Tabela 4.3 ilustra os resultados obtidos em relação ao número de mensagens classificadas:

	Vspam (Verdadeiro spam)	**VNspam (Verdadeiro não-spam)**	**Total**
Spam	**1021**	**0**	**1021**
Nspam	**27**	**88**	**115**
	1048	**88**	**1136**

Tabela 4.3: Resultados obtidos com o *Mozilla Mail*

Verificando a Tabela 4.3, nota-se que a pasta "*spam*" recebeu um total de 1021 mensagens, onde 1021 foram classificadas corretamente não permitindo a ocorrência de falso-positivos. Já a caixa de entrada ou "*inbox*" representada por "*Nspam, não-spam*" recebeu um total de 115 mensagens, onde 88 foram classificadas corretamente e 27 foram classificadas erradas permitindo a ocorrência de falso-negativos. O *Mozilla Mail* classificou um total de 1136 mensagens no período de outubro a novembro de 2004.

Tendo em vista os resultados obtidos, a acurácia nesta proporção é 97,6% de sucesso na classificação das mensagens.

Quanto mais *spam* um usuário recebe, é menos provável que ele encontre e-mails inocentes na pasta de "*spam*". Mas, caso o *Mozilla Mail* tenha feito uma marcação errada, ou seja, a ocorrência de falso-positivos (são e-mails inocentes identificados equivocadamente como *spam*), o usuário deverá ser capaz de corrigir a identificação de *spam*. O processo de aprendizagem do filtro é contínuo.

O *Mozilla Mail* é gratuito e roda em diversas plataformas de computadores, como o *Microsoft Windows*, *Linux*, *Apple Macintosh*, etc. Ao contrário de outros software ditos gratuitos, o *Mozilla* não possui custos:

Sem restrições: os produtos (tecnologias) do Projeto *Mozilla* são completamente livres de ônus;

Sem propagandas: nenhum dos produtos que o Projeto oferece requer que o usuário veja propagandas para que possa usá-lo;

Sem programas espiões (spyware): os produtos não tentarão coletar informações pessoais do usuário ou sobre o uso do computador;

Sem comportamentos intrusivos: os programas não tentam instalar ou criar atalhos para software de parceiros;

Liberdade: o usuário pode usar quantas cópias do *Mozilla* quiser, ou seja, poderá copiá-lo para diferentes máquinas.

Na verdade, a liberdade é o núcleo da filosofia do Projeto *Mozilla*. Todos os programas do projeto são exemplos de código-aberto. Isso significa que eles são regidos por uma licença liberal a qual não impõe restrições na maneira como o software será usado. Além disso, qualquer um que use um programa de código-aberto está autorizado a receber e usar, alterar e, desde que de acordo com a licença, redistribuir o código fonte de tal programa.

A desvantagem do *Mozilla Mail*, contudo, é que ele ocupa 20 MB de memória quando carregado e o filtro *anti-spam* mostra-se lento em um computador antigo. Mas, por outro lado, os usuários iniciantes podem operá-lo sem restrições.

4.3.2.2 *POPFile*

O *POPFile* é uma ferramenta de classificação automática de e-mail. Uma vez corretamente configurado e treinado, trabalha em segundo plano, examinando os e-mails na medida em que chegam e classificando-os conforme o usuário desejar **[41]**.

A versão utilizada foi o *POPFile 0.20.1* que possui outras formas de filtragem, ou seja, pode-se dar a ele uma tarefa simples como separar apenas e-mails inúteis ou uma tarefa mais complicada como separar e-mails em várias pastas diferentes, chamadas de baldes (*buckets*). Pense nele como um assistente pessoal para a caixa de entrada. A Figura 4.12 ilustra a interface de autenticação do *POPFile Control Center* no momento da sua execução via *Web*:

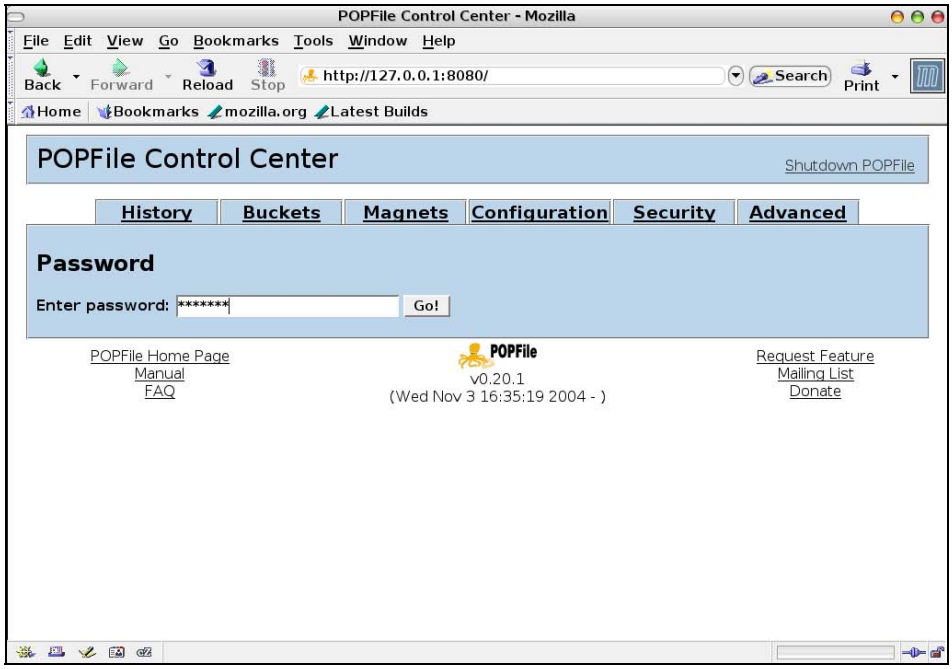

Figura 4.12: Interface de autenticação do *POPFile*

O *POPFile* funciona como um servidor *proxy*. O programa de e-mail do usuário conversa com o *POPFile*, que conversa com o servidor de e-mail **[42]**. Ao invés de o programa de e-mail do usuário pegar as mensagens diretamente no servidor de correio eletrônico, o *POPFile* as pega primeiro, e as analisa para decidir em qual balde (*bucket*) elas devem ser colocadas.

Uma vez configurado corretamente, o *POPFile* adiciona uma marcação utilizando colchetes no início da linha de assunto da mensagem (*subject*) e poderá incluir também o parâmetro ***X-Text-Classification: spam*** no cabeçalho do e-mail (como mostra a Figura 4.13):

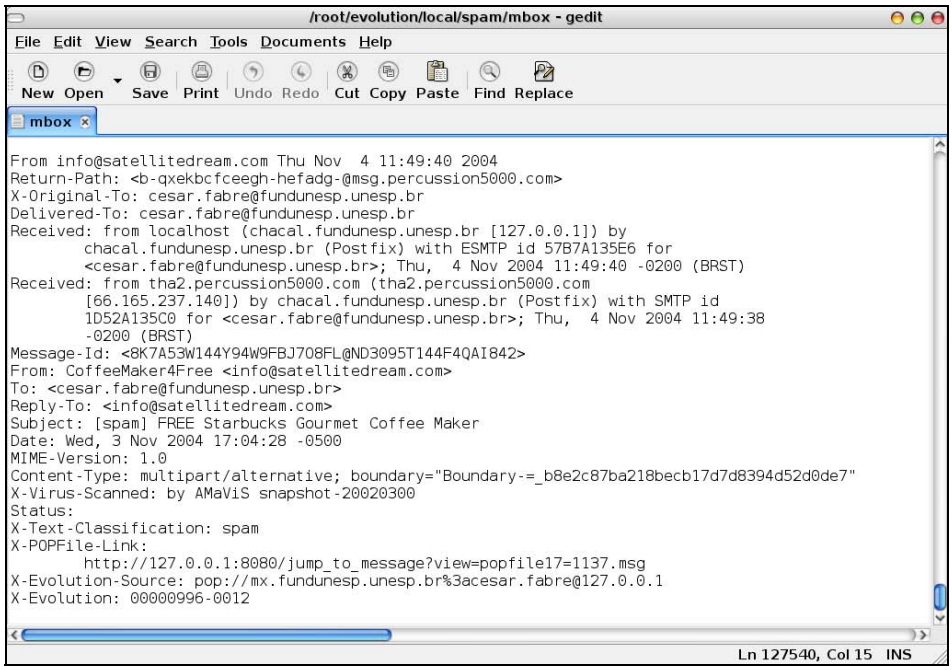

Figura 4.13: Cabeçalho da mensagem alterado pelo *POPFile*

Para realização dos testes, é necessário mostrar ao *POPFile* como classificar as mensagens de e-mail. Recém instalado, o *POPFile* não consegue filtrar as mensagens corretamente. Ele não sabe o que é *spam*, o que é e-mail, ou o que significam *buckets* que o usuário especificou. O sistema de classificação do *POPFile* precisa ser treinado por um tempo antes de se tornar efetivo, quanto mais ele for treinado, mais efetivo ele se torna. De fato, não vai nem mesmo classificar e-mail na primeira vez que o usuário usar, ou seja, vai deixá-lo como "não-classificado".

Os resultados foram gerados a partir de uma base de dados contendo 1136 mensagens, onde foram criados dois *buckets*: na pasta "inbox" foram recebidas 108 mensagens e na pasta "*spam*" foram 1028 mensagens no período de outubro a novembro de 2004 (ver Figura 4.14):

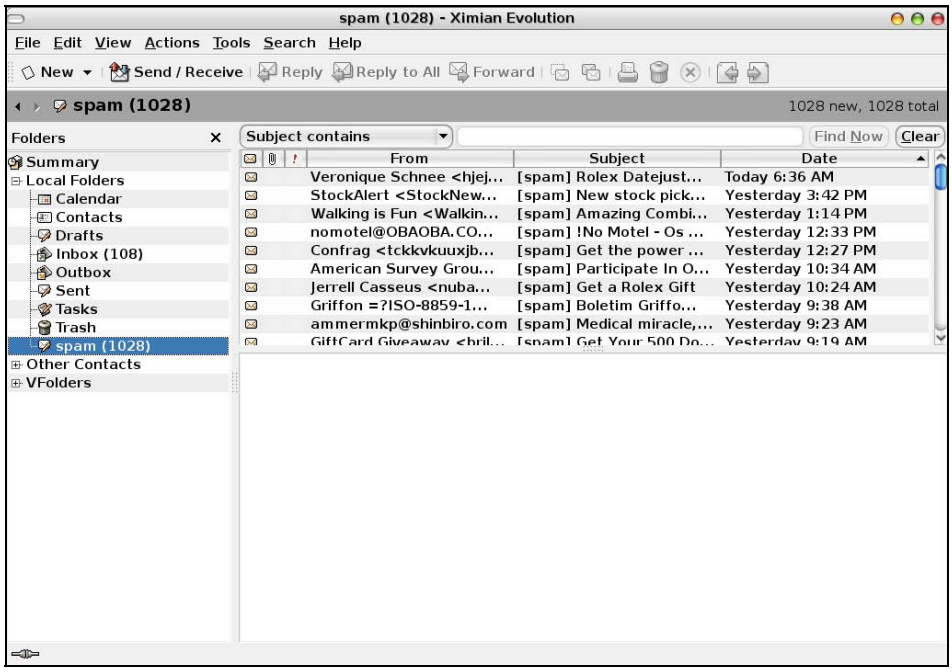

Figura 4.14: *POPFile + Ximian Evolution*

De acordo com a Figura 4.14, o *POPFile* classifica os e-mails fazendo uma análise estatística de cada palavra em cada e-mail que o usuário recebe, incluindo os cabeçalhos e move as mensagens para os *buckets* apropriados baseado em probabilidades, usando uma técnica estatística chamada *Naive Bayes*. Quando o *POPFile* classifica errado um e-mail, ou não o classifica, o usuário deverá ser capaz de corrigir o *POPFile* selecionando a classificação correta no *POPFile Control Center*. A Tabela 4.4 ilustra os resultados obtidos em relação ao número de mensagens classificadas:

	Vspam (Verdadeiro spam)	**VNspam (Verdadeiro não-spam)**	**Total**
Spam	**1028**	**0**	**1028**
Nspam	**20**	**88**	**108**
	1048	**88**	**1136**

Tabela 4.4: Resultados obtidos com o *POPFile*

Verificando a Tabela 4.4, nota-se que a pasta "*spam*" recebeu um total de 1028 mensagens, onde 1028 foram classificadas corretamente não permitindo a ocorrência de

falso-positivos. Já a caixa de entrada ou "*inbox*" representada por "*Nspam, não-spam*" recebeu um total de 108 mensagens, onde 88 foram classificadas corretamente e 20 foram classificadas erradas permitindo a ocorrência de falso-negativos. O *POPFile* classificou um total de 1136 mensagens no período de outubro a novembro de 2004.

Tendo em vista os resultados obtidos, a acurácia nesta proporção é 98,2% de sucesso na classificação das mensagens.

O *POPFile* é gratuito e roda na maior parte das plataformas de computadores, incluindo *Microsoft Windows*, *Apple Macintosh e Linux*.

A desvantagem do *POPFile* é o fato de ele não ser para qualquer tipo de usuário. Os que possuem habilidades medianas, ou seja, capazes de configurar uma instalação de um software e operá-la, acharão o *POPFile* fácil de usar, porém, os iniciantes ficarão confusos.

4.4 Comparação dos Filtros *Bayesianos* Avaliados

Como foi visto na sessão anterior (4.3), os filtros *Bayesianos* requerem treinamento para aumentar a sua precisão na classificação das mensagens. Mesmo assim, é possível observar que os filtros *Bayesianos* não são perfeitos devido às taxas de falso-negativos encontradas nos testes realizados no período de outubro a novembro de 2004. Portanto, esta sessão apresenta uma breve comparação entre os filtros *Bayesianos* avaliados com relação às taxas de falso-positivos e falso-negativos. Lembrando que na realização dos testes todas as ferramentas avaliadas apresentaram o valor 0 (zero) para as taxas de falso-positivos. Isso é muito importante porque em geral as taxas de falso-positivos são mais graves que as taxas de falso-negativos.

4.4.1 Metodologia

Cada programa foi instalado de acordo com a sua documentação. No início, os filtros exigiram treinamento, ou seja, os dados foram fornecidos para que as ferramentas resultassem em níveis satisfatórios. A partir desse treinamento inicial foi possível avaliar as ferramentas no quesito acertos e erros na classificação das mensagens.

O objetivo principal era examinar a precisão das ferramentas *Bayesianas anti-spam* na classificação das mensagens, observando a capacidade de cada filtro com relação ao aprendizado.

Nos testes não foram utilizadas técnicas como *whitelist* e *blacklist*, apesar de que, na prática, algumas ferramentas *Bayesianas* avaliadas possuem essas técnicas embutidas, bastando ativá-las.

4.4.2 Resultados

O *spam* via e-mail tornou-se um problema gravíssimo que atinge milhares de usuários na Internet. Nos dias atuais, é praticamente impossível usar o correio eletrônico sem a recepção de mensagens não solicitadas em grandes quantidades. Para amortizar o problema, existem várias técnicas que podem ajudar no combate ao *spam* como: filtros baseados em Heurística, listas de bloqueio (*blacklists*) e listas de permissão (*whitelists*). Todas elas apresentam problemas, ou seja, os filtros baseados em Heurística necessitam constantemente de atualizações e não fazem classificação correta, as listas de bloqueio e permissão também necessitam de atualizações e quase sempre estarão atrás dos *spammers*.

Felizmente, apareceram os filtros *Bayesianos* para conter essas mensagens não solicitadas que recebemos todos os dias. Essa nova técnica é baseada no artigo de Paul Graham "*A Plan for Spam*" e pode-se concluir que essa seja a melhor esperança para livrar-se do *spam*. Os filtros *Bayesianos* são baseados em métodos estatísticos que fornecem probabilidades para cada mensagem pertencer a uma determinada classe (na maioria das vezes são duas classes, *spam* e *não-spam*), mas isso não é uma limitação da técnica, pois o *POPFile* atinge um número arbitrário de classes através da criação dos baldes (*buckets*).

A grande vantagem dos filtros *Bayesianos* é que eles podem ser treinados pelos usuários simplesmente categorizando cada mensagem recebida em *spam* e *não-spam*. Depois, com o passar do tempo, o filtro começa a fazer a categorização por si próprio e normalmente com um nível muito alto de precisão. Mas a desvantagem dos filtros *Bayesianos* poderá ocorrer caso o filtro não esteja treinado corretamente ocasionando falso-positivos, ou seja, mensagens legítimas serem classificadas como sendo *spam* ou falso-negativos, mensagens com conteúdo de *spam* serem classificadas como não-*spam*.

Entretanto, se o filtro cometer algum erro, o usuário poderá corrigi-lo e automaticamente o filtro aprenderá os erros cometidos.

Em geral, as taxas de falso-positivos são mais graves do que a ocorrência de falso-negativos, ou seja, visualizar um *spam* é melhor do que não ver uma mensagem importante. Uma solução seria verificar a pasta de *spams* uma vez por dia, mas isso se torna tedioso quando o ponto receptor recebe muitos *spams* e o propósito de filtrar *spam* não teria o mesmo valor.

Devido ao fato de que as taxas de falso-positivos e falso-negativos influenciarem no resultado da acurácia, decidiu-se compará-las entre os filtros *Bayesianos anti-spam* analisados.

Lembrando que os filtros *Bayesianos* necessitam de treinamento inicial, foram atribuídas aos filtros 200 mensagens consideradas *spam* e 200 consideradas *não-spam* para aumentar a precisão na classificação das mensagens. Mesmo assim é possível observar na Tabela 4.5 que os filtros *Bayesianos* não são perfeitos devido às taxas de falso-negativos apresentadas nos testes, mas ajudam a minimizar o problema ocasionado pelo *spam*.

Resultado da classificação de mensagens na categoria Spam				
	POPFile	Mozilla Mail	SpamAssassin	Bogofilter
Falso-negativos (spam na pasta inbox)	1,9%	2,6%	3,0%	3,5%
Verdadeiro-positivos (spam na pasta spam)	98,1%	97,4%	97,0%	96,5%

Tabela 4.5: Taxas de Falso-negativos e Verdadeiro-positivos

De acordo com a Tabela 4.5, os filtros analisados conseguiram índices em torno de 96,5% a 98,1% de acerto na classificação das mensagens com conteúdo de *spam* e o restante cerca de 1,9% a 3,5% representam a pequena quantidade de falso-negativos, ou seja, mensagens consideradas *spam* serem classificadas como *não-spam*. Analisando as mensagens que ocasionaram as taxas de falso-negativos, nota-se que elas não passam de um simples e-mail no formato texto solicitando que o usuário acesse um determinado site para efetuar a compra de produtos. Entretanto, essas mensagens podem ser marcadas pelo usuário como *spam* e automaticamente o filtro *Bayesiano* aprende que essas mensagens

possuem conteúdo de *spam*. Com isso, da próxima vez que forem enviadas a um determinado usuário, serão movidas para a pasta *spam*.

Através dos resultados da Tabela 4.5 pode-se concluir que o *POPFile* apresenta as melhores taxas, mas isso não quer dizer que o *POPFile* irá filtrar sempre 98,1% dos *spams* de uma organização. Portanto, com base nos testes realizados e no conjunto de mensagens coletadas, o filtro do *POPFile* foi o que apresentou resultados mais satisfatórios com relação aos outros filtros analisados.

Já a Tabela 4.6 apresenta as taxas de falso-positivos (*não-spam* na pasta *spam*) e verdadeiro-negativos (*não-spam* na pasta *inbox*).

Resultado da classificação de mensagens na categoria Não-spam				
	POPFile	Mozilla Mail	SpamAssassin	Bogofilter
Falso-positivos (não-spam na pasta spam)	0	0	0	0
Verdadeiro-negativos (não-spam na pasta inbox)	100,0%	100,0%	100,0%	100,0%

Tabela 4.6: Taxas de Falso-positivos e Verdadeiro-negativos

Como pode ser visto na Tabela 4.6 todas as ferramentas avaliadas conseguiram índices em torno de 100% de acerto na classificação das mensagens *não-spam* em relação ao conjunto de mensagens coletadas e, com isso, não houve ocorrência de falso-positivos, ou seja, mensagens legítimas serem classificadas como *spam*. Esse resultado é muito importante, pois em geral, as taxas de falso-positivos são mais graves do que as taxas de falso-negativos.

Esse resultado de 100% na classificação das mensagens legítimas não ocasionando taxas de falso-positivos se deve à pequena quantidade de mensagens recebidas no período (88 mensagens) e no conjunto de mensagens utilizadas nos testes não conter diversidade suficiente. Como a maioria das mensagens *não-spams* recebidas para os testes foram do mesmo domínio das mensagens utilizadas no treinamento, os filtros analisados conseguiram índices de 100% na classificação das mensagens, mas por conta do conjunto de mensagens coletadas os filtros não foram forjados o suficiente para ocasionar taxas de falso-positivos.

Vale ressaltar que as mensagens foram coletadas no período de outubro a novembro de 2004 totalizando 1136 mensagens classificadas pelos filtros analisados. Com relação aos resultados obtidos e no conjunto de mensagens coletadas, pode-se concluir que o *POPFile* apresenta resultados mais satisfatórios no quesito acertos e erros na classificação das mensagens.

Portanto, uma conclusão bastante ponderada é que existem muitos produtos que estão trabalhando para minimizar o impacto do *spam*, mas somente algumas ferramentas são capazes de solucionar o problema. Como exemplo de ferramenta eficaz pode-se considerar o *POPFile*, ou seja, o autor do software *POPFile*, John Graham-Cumming, tem realmente criado filtros eficazes no combate ao *spam*.

Com base nos dizeres acima, a recomendação sugerida no presente trabalho é a utilização do *POPFile* como filtro de *spam Bayesiano* neste momento.

4.5 Conclusão

Este capítulo analisou a automatização de algumas ferramentas *Bayesianas anti-spam*, buscando atingir resultados satisfatórios na classificação das mensagens. Observou-se que a técnica *Bayesiana* utiliza métodos estatísticos e probabilísticos na classificação das mensagens, fazendo com que os resultados sejam mais coerentes com relação às taxas de falso-positivos e falso-negativos. Os resultados obtidos apontam o filtro *Bayesiano anti-spam* do *POPFile* como uma alternativa promissora para filtrar mensagens com conteúdo de *spam* (como visto nos testes gerou pequenas taxas de falso-negativos). Já a questão das mensagens *não-spam* todos os filtros conseguiram índices em torno de 100% devido às poucas mensagens recebidas no período e a diversidade das mesmas.

Capítulo 5

Conclusão

Neste trabalho, com mais ênfase no capítulo 2, foram apresentados diversos problemas ocasionados pelo recebimento de mensagens não solicitadas (*spam*) e os custos mais comuns encontrados nas corporações devido à sobrecarga de mensagens desnecessárias todos os dias. Viu-se que existem empresas na Internet que fornecem ferramentas *online* capazes de calcular os custos referentes ao recebimento de *spam* dentro das corporações. Através dessas ferramentas os usuários podem inserir os dados de uma determinada empresa para poder verificar os resultados obtidos.

A enorme utilização da Internet nas últimas décadas trouxe consigo benefícios provocados pela popularização das novas tecnologias de informática, mas surgiram problemas motivados por esses serviços, dentre eles, o serviço de correio eletrônico, também conhecido como e-mail.

Em abril de 2004, o tráfego das mensagens não solicitadas chegou a 64% do total de mensagens que circulam na Internet, ou seja, todo esse lixo é acumulado nas caixas postais, comprometendo o desempenho dos servidores de correio eletrônico e da rede, além de fazer com que boa parte do nosso horário de trabalho seja destinado a limpar nossa caixa postal.

Estas vulnerabilidades e as novas ferramentas adotadas pelos *spammers* motivam um esforço cada vez maior em busca de soluções e tecnologias capazes de minimizar e, até mesmo, eliminar por completo a maioria desses problemas. Bons exemplos são: filtros *Bayesianos*, autenticidade do remetente, filtros baseados em Heurística, listas *blacklists* / *whitelists* e um sistema distribuído para calcular *checksum* (DCC).

Este trabalho abordou no capítulo 3 as diferentes tecnologias *anti-spam*, utilizadas para filtragem dos e-mails recebidos, separando os *spams* dos e-mails válidos. Com tais tecnologias *anti-spam*, o presente trabalho abordou um estudo mais detalhado na técnica classificação de conteúdo, que utiliza abordagens diferentes das outras, pois, ao invés de analisar apenas o cabeçalho procurando identificar remetentes suspeitos, ela analisa todo o conteúdo da mensagem em busca de padrões suspeitos e, com base na identificação de

determinados padrões, utiliza estatística e probabilidade para fazer a classificação das mensagens. Essa técnica é baseada no Teorema de *Bayes* e por isso conhecida como Filtro *Bayesiano*.

No capítulo 4, foram analisadas algumas ferramentas *anti-spam* para servidor de e-mail (MTA) e usuário final (MUA) que implementam filtros *Bayesianos*. Como os filtros *Bayesianos* necessitam de treinamento inicial foi necessário atribuir aos filtros 200 mensagens consideradas *spam* e 200 consideradas *não-spam*. A partir dessa etapa de aprendizagem pode ser montada uma base de dados com as palavras (*tokens*) encontradas, e qual a incidência de cada *token* na mensagem, ou seja, são montadas listas de *tokens* bons (encontrados em *não-spams*) e *tokens* ruins (encontrados em *spam*). Com base nessas listas é possível estimar a probabilidade de uma mensagem ser ou não *spam*. Após o treinamento inicial arquivou-se um conjunto de mensagens no período de outubro a novembro de 2004 totalizando 1136 mensagens para a realização dos testes.

Com relação aos testes pode-se concluir que o ambiente de teste deste trabalho não é o mais adequado para a filtragem de mensagens, pois cientificamente apresentaram falhas que marcaram o trabalho dos filtros *Bayesianos anti-spam.* Pode-se dizer que 200 mensagens consideradas *spam* e 200 consideradas *não-spam* não são suficientes para início dos treinamentos, assim como o período de 2 (dois) meses para a coleta das mensagens não é um prazo considerável para a realização dos testes com sucesso. Além disso, a criação de uma conta de e-mail e divulgada em vários sites na Internet tornou a filtragem das mensagens artificial e não natural. O mais adequado seria utilizar-se uma *mailbox* pessoal que recebesse muitas mensagens com conteúdo de *spam* e *não-spam*. Portanto, este trabalho fornece mecanismos favoráveis que demandam a elaboração de um outro ambiente de teste a ser realizado futuramente em outras pesquisas.

Com relação às mensagens *não-spams*, os filtros analisados conseguiram índices de 100% na classificação das mensagens devido a dois motivos: (1) a pequena quantidade de mensagens recebidas no período de outubro a novembro de 2004; (2) ao fato de o conjunto de mensagens coletadas utilizadas nos testes não conter uma diversidade suficiente para ocasionar taxas de falso-positivos.

Com este trabalho foi possível extrair conhecimento das técnicas *anti-spam* e obter resultados satisfatórios com o uso de ferramentas *Bayesianas* para minimizar o problema

ocasionado pelo *spam*. Pode-se concluir que as ferramentas disponíveis hoje se encontram em avançado estágio de desenvolvimento e possuem recursos sofisticados, permitindo que o usuário leigo ou mesmo o especialista em ferramentas *anti-spam* implementem formas de combate ao *spam*

Como principais contribuições deste trabalho, pode-se destacar o estudo detalhado dos principais problemas ocasionados pelo recebimento de mensagens não solicitadas e os testes realizados com as principais ferramentas *Bayesianas anti-spam*. O leitor mais interessado pode encontrar ainda, no apêndice A, uma lista das ferramentas *anti-spam* que utilizam a técnica *Bayesiana*, no apêndice B, algumas configurações realizadas no arquivo do *Procmail* "*.procmailrc*" e no apêndice C, o *Script* utilizado para executar o arquivo "*.procmailrc*".

5.1 Trabalhos futuros

A continuidade deste trabalho envolve um estudo mais detalhado dos filtros *Bayesianos anti-spam* com relação às implementações que cada filtro utiliza na classificação das mensagens e demanda uma pesquisa mais aprofundada sobre a característica do *POPFile* de classificação em vários baldes (*buckets*), em cima do *SpamAssassin*, que consiste no mecanismo básico de filtragem (*spam* e *não-spam*). A idéia é fazer com que a classificação seja em faixas de probabilidade (e não de forma binária, sim e não), e depois depositar cada e-mail classificado na sua *mailbox* respectiva. Os e-mails mais bem classificados, em tese, deverão receber confiança alta e depositados na *mailbox* apropriada, a qual raramente o usuário deverá investigar em busca de falso-positivos. As mensagens com menos confiança irão para outras *mailboxes*, que deverão ser visitadas freqüentemente, conforme o grau de confiança. Além disso, merece destaque o estudo mais aprofundado da tecnologia *anti-spam* (Autenticidade do Remetente), que segundo o autor da ferramenta essa técnica de filtragem é promissora e consiste em filtrar os *spams* do futuro.

Outra sugestão para pesquisa futura seria estudar métodos que permitam detectar e acompanhar ataques a redes de computadores. Um dos métodos que tem sido utilizado é o desenvolvimento, implementação e monitoração de *honeynets*. *Honeynets* são ferramentas de pesquisa que consistem em uma rede projetada especificamente para ser comprometida.

Uma vez comprometida, a *honeynet* é utilizada para observar o comportamento dos invasores, possibilitando a realização de análises detalhadas das ferramentas utilizadas, de suas motivações e das vulnerabilidades exploradas. Em geral, a *honeynet* é composta por diversos *hosts*, que são *honeypots* com sistemas operacionais e arquiteturas variadas, de modo a permitir que seja possível observar o comportamento de invasores em diversas plataformas. Um desses *hosts* opera como servidor de nomes para a *Honeynet*, além de possuir o serviço de *syslog* habilitado, atuando como servidor central de *logs* para os demais *hosts*. Através do uso de *honeypots* pode-se analisar a constante procura por *proxies* abertos e servidores de e-mail mal configurados, que permitam sua utilização para envio de spam.

Referências Bibliográficas

[1] HUNT, L. *Past.com; An Internet timeline.* Computerworld, Framingham, MA, p.81, Mai.1999.

[2] GREENEMEIER, L. *Another Year And The Internet Can Legally Drink; In 1983, ARPANET officially switched from the Network Control Program protocol to TCP/IP, setting the stage for a revolution*. InformationWeek, Manhasset, NY, p.NA, Dez.2002.

[3] BRANZBURG, J. *How to fight spam. (In-Service)*. Technology & Learning, San Francisco, CA, v.23, i.8, p.39, Mar. 2003.

[4] MAGEE, J. F. *The law regulating unsolicited commercial e-mail: an international perspective*. Santa Clara Computer & High Technology Law Journal, Santa Clara, CA, v.19, i.2, p.333-382, Mai.2003.

[5] PC WORLD COMMUNICATIONS, Inc. *Forgotten pioneer. (20 Years of Online)*. PC World, San Francisco, CA, v.21, i.3, p.117, Mar.2003.

[6] MOODY, G. *Spam, spam, spam, spam. (junk e-mail)*. Computer Weekly, New York, NY, p.34, Out.2002.

[7] INFO-TECH RESEARCH GROUP. *Selecting a spam suppressor*. Info-Tech Advisor Newsletter, London, ON, p.NA, Fev.2003.

[8] GIBBS, M. *A Bayesian filter that nabs 99.75% of spam?*. Network World, Southborough, MA, p.38, Set.2003.

[9] PROVOST, J. 1999. *Naïve-Bayes vs. Rule-Learning in Classification of Email*. The University of Texas at Austin, Artificial Intelligence Lab. Technical Report AI-TR-99-284.

[10] RENNIE, J. D. M. 2000. *ifile: An Application of Machine Learning to E-Mail Filtering*. Proceedings of the KDD-2000 Workshop on Text Mining.

[11] SAHAMI, M.; DUMAIS, S.; HECKERMAN, D.; HORVITZ, E. 1998. *A Bayesian Approach to Filtering Junk E-Mail*. In *Learning for Text Categorization: Papers from the 1998 Workshop. AAAI Technical Report WS-98-05*.

[12] COHEN, W. W. 1996. *Learning Rules that Classify E-Mail*. In *Proceedings of the 1996 AAI Spring Symposium on Machine Learning in Information Access.*

[13] TEIXEIRA, R. C. *O Pesadelo do SPAM*. News Generation, Rio de Janeiro, RJ, v.5, n.1, Jan.2001.

[14] GRAHAM, P. *A Plan for Spam*. Ago.2003. Disponível em: <http://www.paulgraham.com/spam.html>. Acesso em: 20/05/2004.

[15] VIANNA, C. S. M. *Spam: uma abordagem crítica*. Jus Navigandi, Teresina, PI, a.6, n.59, Out.2002.

[16] SINGEL, R. *Spam Pitches Are Mutating Faster*. Wired News, San Francisco, CA, Out.2003.

[17] MICROSOFT, Corporation. *SPAM Como conter a praga dos e-mails?*. Revista Microsoft Business, São Paulo, SP, n.28, p.18-21, Jul.2003.

[18] CLABURN, T. *SpamCop Wins Round In Legal Battle; District court dissolves temporary restraining order, allowing site to continue to report complaints of spam*. InformationWeek, Manhasset, NY, p.NA, Mai.2004.

[19] CHATEL, M. *Classical versus Transparent IP Proxies*: RFC 1919. Annecy-Le-Vieux, France: Internet Engineering Task Force, Network Working Group, 1996.

[20] KLENSIN, J. *Simple Mail Transfer Protocol*: RFC 2821. Boston, MA: Internet Engineering Task Force, Network Working Group, 2001.

[21] VIJAYAN, J. *Groups that run real-time blacklists*. Computerworld, Framingham, MA, v.38, i.37, p.7, Set.2004.

[22] MANION, A. *Vulnerability Note VU#150227 – Multiple vendors' HTTP proxy default configurations allow arbitrary TCP conections*. CERT Advisory, Pittsburgh, PA, Mai.2002.

[23] PEPPERS & ROGERS GROUP. *Spam*. Inside 1to1 Newsletter, São Paulo, SP, Fev.2004.

[24] PICCOLINI, J. D. B.; TEIXEIRA, R. C. *Ferramentas Anti-Spam para o usuário final em Plataformas Windows*. NewsGeneration, Rio de Janeiro, RJ, v.7, n.4, Ago.2003.

[25] INFO-TECH RESEARCH GROUP. *Close the Door on E-Mail Relaying*. Info-Tech Advisor Newsletter, London, ON, p.NA, Abr.2002.

[26] MURRAY, A.C. *The Spam-Haters Club – ORBS and MAPS brand spammers with a scarlet S so that networkers know who to ostracize from their mail servers. But it the best way to stop junk e-mail?*. Network Magazine, Manhasset, NY, p.62, Abr.2001.

[27] GIBBS, M. *Math to fight spam*. Network World, Southborough, MA, p.30, Set.2003.

[28] DORNAN, A. *Lesson 188: Bayesian Spam Filtering*. Network Magazine, Manhasset, NY, p.64, Mar.2004.

[29] PEARL, J. *Bayesian Networks*. UCLA Cognitive Systems Laboratory, Technical Report (R-246), Revision 1, July 1997. In *MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA, 1999.

[30] READY, J. *The big squeeze: closing down the junk e-mail pipe*. Computer Technology Review, Beverly Hills, CA, v.23, i.12, p.34(2), Dez.2003.

[31] WIVES, L. K. *Técnicas de Descoberta de Conhecimento em Textos Aplicadas à Inteligência Competitiva*. 2001. Exame de Qualificação. Porto Alegre: PPGC/UFRGS.

[32] BILLSUS, D.; PAZZANI, M. *A Hybrid User Model for News Story Classification*. 1999. Proceedings of the Seventh International Conference on User Modeling (UM'99), Banff, Canada.

[33] CROCKER, D. *Mailbox Names for Common Services, Roles and Functions*: RFC 2142. Santa Cruz, CA: Internet Engineering Task Force, Network Working Group, 1997.

[34] HAMBRIDGE, S. *Netiquette Guidelines*: RFC 1855. Santa Clara, CA: Internet Engineering Task Force, Network Working Group, 1995.

[35] GRAHAM, P. *Better Bayesian Filtering*. Jan.2003. Disponível em: <http://www.paulgraham.com/better.html>. Acesso em: 14/07/2004.

[36] RAYMOND, E. S. *Communications and Internet (Bogofilter)*. Set.2002. Disponível em: <http://www.catb.org/~esr/software.html>. Acesso em: 14/07/2004.

[37] COUSINS, A. *Advanced SpamAssassin techniques. (Linux)*. Australian PC World, St Leonards, Australia, p.140, Ago.2003.

[38] SYNDER, J. *Where's SpamAssassin?*. Network World, Southborough, MA, p.38, Dez.2004.

[39] RAPOSA, J. *Mozilla 1.4's Key Improvements Out of Sight*. eWeek, Woburn, MA, p.NA, Jul.2003.

[40] CRUICKSHANK, A. *Switch to Mozilla from Internet Explorer: change your default browser with the minimum of hassle*. Internet Magazine, London, ON, i.115, p.78(2), Mar.2004.

[41] HOLZMAN, C. *Bayesian Filters To the Rescue – Three free programs can help fight spam*. VARbusiness, Manhasset, NY, p.84, Dez.2004.

[42] BLASS, S. *ask dr. internet. (Bayesian spam filter for Windows and Linux)*. Network World, Southborough, MA, p.43, Out.2003.

Apêndice A

Lista das Ferramentas *Bayesianas Anti-Spam*

Este apêndice apresenta uma lista mais abrangente das ferramentas *anti-spam* que utilizam a técnica *Bayesiana* para a classificação das mensagens.

A.1 Ferramentas *Bayesianas* para (MTA)

Nome	Classificação	Plataforma	Solução	Implementação
Anti-Spam SMTP Proxy Server	Inclui filtro Bayesiano	MacOS, Windows, OS/2, OS Independent, POSIX	Gratuita	Perl
Bayesian Mail Filter	Bayesiana	POSIX	Gratuita	C
Bogofilter	Bayesiana	MacOS X, AIX, FreeBSD, HP-UX, Linux, SunOS/Solaris	Gratuita	C
Quick Spam Filter	Bayesiana	Windows, BSD, Linux, SunOS/Solaris	Gratuita	C
SpamAssassin	Inclui filtro Bayesiano	OS Independent, POSIX	Gratuita	Perl
SpamProbe	Bayesiana	MacOS X, AIX, FreeBSD, Linux,, SunOS/Solaris	Gratuita	C++

Tabela A.1: Filtros *Bayesianos anti-spam* para (MTA)

A.2 Ferramentas *Bayesianas* para (MUA)

Nome	Classificação	Plataforma	Solução	Implementação
JoeEmail	Bayesiana	Windows	Gratuita	Visual Basic
Mozilla Mail	Bayesiana	Windows, *Linux*, MacOS X	Gratuita	C++
Pop3 Agent	Bayesiana	Windows 95/98, Windows NT/2000	Gratuita	Delphi/Kylix
POPFile	Bayesiana	MacOS, Windows, OS Independent, POSIX	Gratuita	Perl
SpamBayes	Bayesiana	Windows, OS Independent	Gratuita	Python
SpamPal Bayesian Plugin	Bayesiana	Windows 95/98/2000	Gratuita	C++

Tabela A.2: Filtros *Bayesianos anti-spam* para (MUA)

Apêndice B

Configurações realizadas no arquivo "*.procmailrc*"

Este apêndice apresenta as configurações que foram realizadas junto ao arquivo do *Procmail* "*.procmailrc*". Esse arquivo foi essencial para a realização da filtragem nas ferramentas *Bayesianas* para (MTA).

B.1 Arquivo "*.procmailrc*" do *Bogofilter*

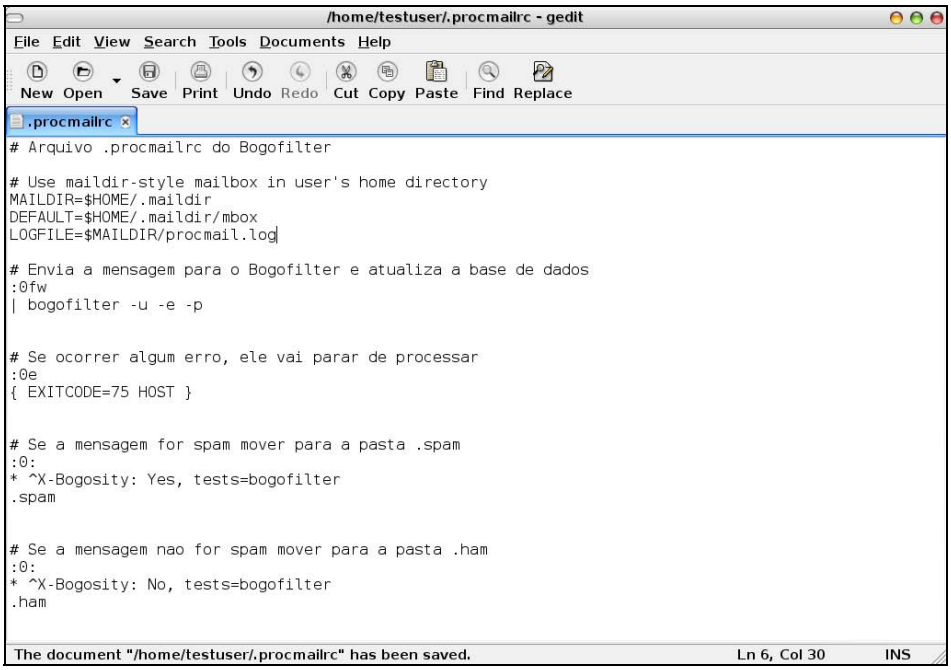

Figura B.1: Arquivo "*.procmailrc*" do *Bogofilter*

B.2 Arquivo "*.procmailrc*" do *SpamAssassin*

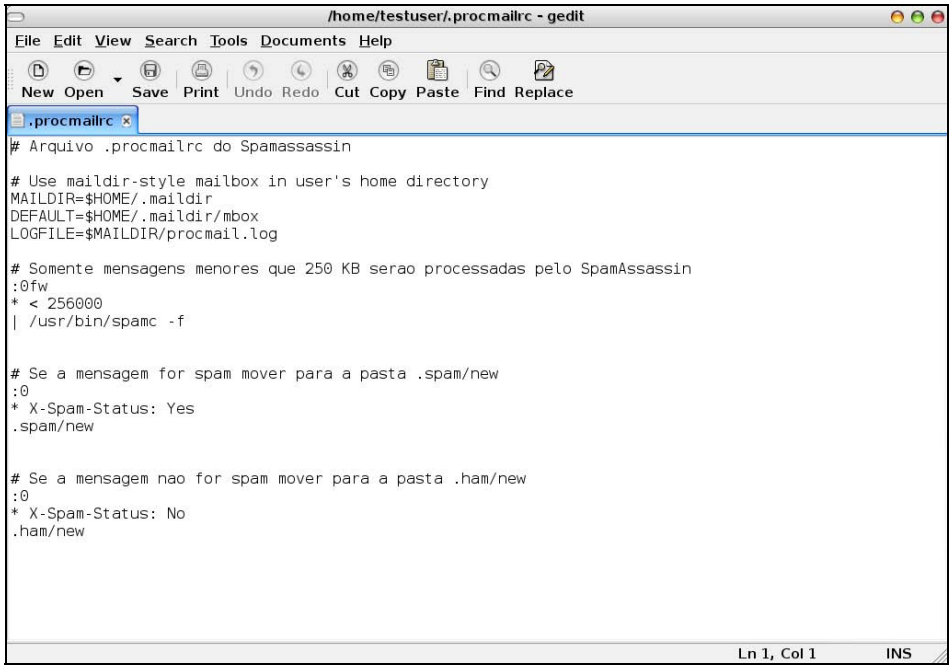

Figura B.2: Arquivo "*.procmailrc*" do *SpamAssassin*

Apêndice C

Shell Script

Este apêndice apresenta o *Script* utilizado nos testes das ferramentas *Bayesianas anti-spam* para (MTA). Através desse *Script* é possível executar o arquivo do *Procmail* "*.procmailrc*" com apenas um comando no terminal do Linux.

C.1 *Script* que executa o arquivo "*.procmailrc*"

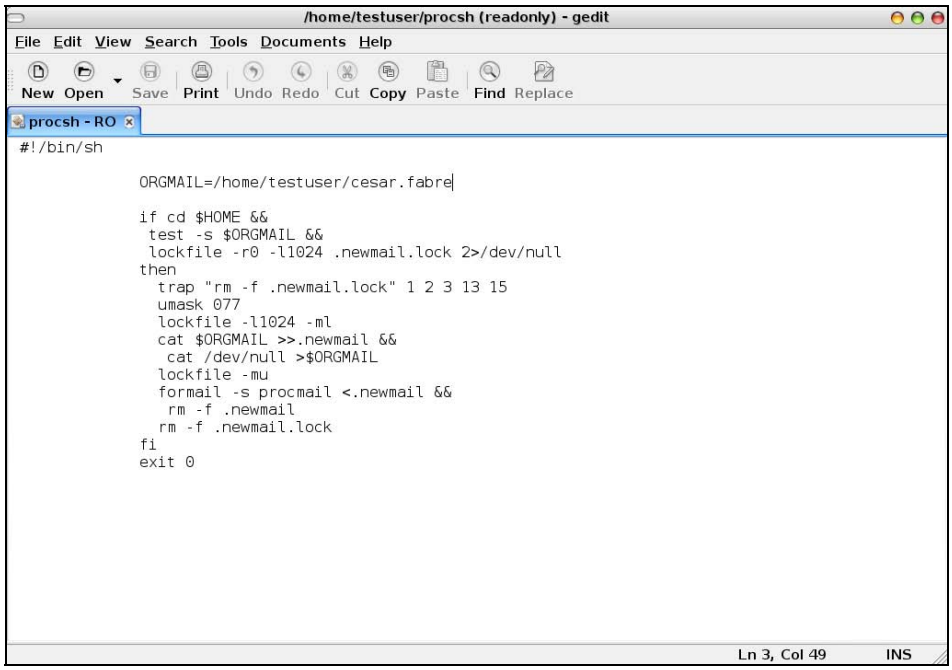

Figura C.1: *Script* que executa o "*.procmailrc*"