Capítulo 5

Técnicas de Defesa Contra *Spam*

Danilo Michalczuk Taveira[1], Igor Monteiro Moraes[1],
Marcelo Gonçalves Rubinstein[2] e Otto Carlos Muniz Bandeira Duarte[1]

[1]Grupo de Teleinformática e Automação - GTA
COPPE/Poli - Programa de Engenharia Elétrica
Universidade Federal do Rio de Janeiro

[2]Programa de Pós-Graduação em Engenharia Eletrônica
Departamento de Engenharia Eletrônica e Telecomunicações - FEN
Universidade do Estado do Rio de Janeiro

Abstract

Spams, or unsolicited electronic messages, represent more than half the e-mail traffic carried nowadays in the Internet and there is no evidence which points out the reduction of sending these messages. This situation increases the operational cost of service providers and also reduces the users trust in the e-mail application. The countermeasure to spams is the adoption of a set of techniques and procedures called anti-spam systems. These systems comprise all the process to fight spams from the prevention of e-mail harvesting and the inhibition of spamming to the messages characterization and filtering. This chapter presents the motivation and the mechanisms used to send spams and the techniques used to classify and filter them. Furthermore, different anti-spam systems proposed in the literature are analyzed in detail. At last, new proposals to inhibit the act of sending spams are discussed.

Resumo

Os spams, ou mensagens eletrônicas não solicitadas, já representam mais da metade do tráfego de correio eletrônico que circula atualmente na Internet e não há indícios que apontem para uma redução do envio destas mensagens. Tal situação aumenta o custo de operação dos provedores de serviço e diminui a credibilidade dos usuários na aplicação de correio eletrônico. A contramedida aos spams é a adoção de um conjunto de técnicas e procedimentos denominados sistemas anti-spam. Estes sistemas procuram atuar em todo o ciclo do processo de combate aos spams, desde a prevenção à construção da lista de destinatários, passando pela coibição do envio, até a caracterização e a filtragem das mensagens. Neste capítulo, são apresentados a motivação e os mecanismos utilizados para enviar os spams e as técnicas usadas para classificá-los e filtrá-los. Além disso, diferentes sistemas anti-spam encontrados na literatura são caracterizados. Por fim, novas propostas para coibir o envio de spams são discutidas.

5.1. Introdução

O combate ao envio de *spams*, é um dos grandes desafios na Internet. O *spam*, de forma simplificada, é toda mensagem eletrônica enviada sem a autorização do destinatário. Devido à simplicidade do protocolo SMTP (*Simple Mail Transfer Protocol*), o correio eletrônico é a aplicação mais afetada pelos *spams*. As estatísticas mostram que os *spams* já correspondem a pelo menos dois terços de todo o tráfego de correio eletrônico transportado pelos provedores de serviço, causando prejuízos da ordem de milhões de dólares [Pfleeger e Bloom, 2005]. Algumas previsões mais pessimistas estimam que, em poucos anos, as mensagens não solicitadas serão responsáveis por 95% do tráfego de correio eletrônico na Internet [Hoanca, 2006]. No Brasil esse problema também é bastante grave. Atualmente, o país é apontado como o quinto maior receptor e também como o quarto maior gerador de *spams* do mundo [Agência Globo, 2005, Spammer-X et al., 2004].

Além de causarem enormes prejuízos aos provedores de serviço, devido ao consumo de recursos tais como banda passante, memória e processamento, os *spams* também consomem inutilmente o tempo dos destinatários e reduzem a credibilidade dos usuários na Internet. A insatisfação entre os usuários é cada vez maior tanto pela perda de tempo na recepção e leitura das mensagens quanto pela possibilidade de disseminação de vírus e de outros programas que causam a perda de dados e o comprometimento da segurança de seus computadores.

A partir da popularização da Internet, um número cada vez maior de usuários tem acesso aos serviços de correio eletrônico, mensagens instantâneas e voz sobre IP (*Voice over IP* - VoIP). Em virtude de tal fato, o envio de *spams* é visto como uma atividade lucrativa. Como os *spams* em sua maioria possuem conteúdo comercial, grande parte dos custos de divulgação do anunciante são transferidos para os provedores de serviço, que são os responsáveis pelo encaminhamento das mensagens até os destinatários. Além disso, um único *spam* pode atingir milhares de destinatários. Um estudo da America Online [Krim, 2003] conclui que apenas dois *spammers* foram responsáveis por dois bilhões de *spams* que resultaram em oito milhões de reclamações de usuários. Um *spammer* é o indivíduo responsável por gerar e/ou enviar *spams*. O envio de *spams* também é estimulado pelo retorno obtido com as mensagens enviadas. Um estudo [Cukier et al., 2006] aponta que 39% dos usuários de correio eletrônico clicam em *spams* e que 11% dos usuários já compraram algum produto anunciado por *spams*. Outro estudo mostra que taxa de resposta às malas-diretas enviadas através de mensagens eletrônicas é doze vezes superior à taxa de resposta das malas-diretas impressas [Cullen, 2002]. Devido a este sucesso, já existem *spams* em serviços de mensagens instantâneas e de voz sobre IP. Além disso, já começam a aparecer os *spams* de vídeo que prometem ter efeitos nocivos ainda mais devastadores que os das mensagens de texto e de voz.

A adoção de sistemas anti-*spam* é a principal contramedida ao envio de mensagens não solicitadas. Os sistemas anti-*spam* são compostos por técnicas e procedimentos que buscam atuar na prevenção e na coibição de todas as etapas do processo de envio de *spams*. Estes sistemas tentam evitar a coleta de endereços de correio eletrônico para construção das listas de destinatários, coibir o envio e caracterizar e filtrar as mensagens. Para tentar classificar e reduzir o número de mensagens não solicitadas, diversos mecanismos foram propostos. A idéia básica destes mecanismos é tentar classificar as

mensagens como *spams* para, então, filtrá-las. As técnicas de combate ao *spam* existentes e os mecanismos propostos na literatura podem ser classificados em três grupos: os sistemas baseados em filtragem simples, os sistemas baseados na verificação da origem e os sistemas com auto-aprendizado. Nos sistemas por filtragem simples, novos dados ou regras são inseridos de forma manual no classificador de mensagens. A principal crítica a sistemas dessa natureza é a baixa eficiência, uma vez que tais sistemas dependem de constante atualização manual. O segundo grupo corresponde aos sistemas de verificação da origem das mensagens. Tais sistemas são essenciais uma vez que o endereço de origem do remetente pode ser facilmente falsificado, dificultando o rastreamento dos *spammers*. Desta forma, o objetivo dos mecanismos baseados na verificação da origem é confirmar a autenticidade do endereço de origem e determinar se o remetente não é um programa de envio automático de mensagens. Outra classe de sistemas anti-*spam* são os sistemas com auto-aprendizado, que são capazes de aprender sozinhos com as mensagens recebidas e, portanto, aumentar a sua eficiência no combate aos *spams*.

Apesar dos esforços para reduzir e até mesmo regulamentar o envio de *spams*, não existe hoje nenhum indício que permita inferir que tal atividade diminuirá nos próximos anos. Ao contrário, os *spammers* vêm se especializando e usando técnicas cada vez mais elaboradas para burlar os sistemas anti-*spam*. Vale ressaltar que os sistemas anti-*spam* estão em constante evolução já que para cada novo mecanismo criado, novas técnicas são desenvolvidas pelos *spammers* para enganá-los e permitir a passagem das mensagens não solicitadas. Também é fato que a maioria dos usuários da Internet não tem formação técnica em computação com capacidade para gerenciar e configurar seus computadores.

O objetivo principal deste capítulo é apresentar os conceitos e as técnicas usadas para classificar e filtrar as mensagens eletrônicas não solicitadas, os *spams*. Primeiramente, na Seção 5.2, discute-se quais características uma mensagem deve apresentar para ser considerada como um *spam*. Em seguida, são abordados alguns aspectos que motivam o envio de *spams*, bem como aspectos legais e iniciativas para regulamentar e coibir o envio de mensagens não solicitadas. Na Seção 5.3, são descritas técnicas usadas pelos *spammers* para obter endereços de correio eletrônico, enviar as mensagens e para burlar os sistemas anti-*spam*. Por sua vez na Seção 5.4, o funcionamento dos sistemas anti-*spam* é detalhado. São apresentados sistemas baseados em filtragem simples, como as listas negras, os sistemas com auto-aprendizado, como os que utilizam características de padrões sociais, e os sistemas baseados na verificação de origem da mensagem, como a verificação do DNS reverso. Por fim, na Seção 5.4.4, são apresentadas as novas propostas e as direções futuras no combate e na regulamentação do envio de *spams* na Internet.

5.2. Mensagens eletrônicas não solicitadas

Definir o que é uma mensagem eletrônica não solicitada é importante tanto para os sistemas anti-*spam*, que devem classificar tais mensagens, quanto para o desenvolvimento de leis que inibam ou até regulamentem o envio de *spams*. Nesta seção, são apresentadas definições e versões sobre a origem do termo *spam*. Também são apresentados alguns tipos de *spams* e as conseqüências e prejuízos causados por essas mensagens. São abordados ainda aspectos que tornam o envio de *spams* lucrativo e quais as medidas legais estão em discussão para regulamentar essa atividade.

5.2.1. Definição e classificação de *spams*

Alguns autores consideram *spam* toda mensagem comercial não solicitada (*Unsolicited Commercial E-mail* - UCE). Essa definição não inclui, por exemplo, mensagens que contêm fraudes e tentativas de golpe e que também são não solicitadas. Outros autores definem um *spam* como uma mensagem não solicitada enviada em batelada (*Unsolicited Bulk E-mail* - UBE), pois, na maior parte das vezes, inúmeras réplicas do mesmo *spam* são enviadas. Para muitos, um *spam* é simplesmente uma mensagem não desejada por um usuário. Entretanto, esta definição de *spam* é bem geral, conflitante e também possui um caráter subjetivo. Uma mensagem que pode ser considerada como um *spam* para um determinado usuário pode não ser para outro. Portanto, a definição do que pode ser considerado um *spam* já é um primeiro desafio. Devido à dificuldade para se afirmar o que é um *spam*, as regulamentações em vigor e as propostas de lei em discussão definem um conjunto de regras baseado em características comuns encontradas em mensagens classificadas como *spams*. Dessa forma, quando uma mensagem não atende às regras definidas, ela é considerada um *spam*.

Figura 5.1. A embalagem do SPAM extraída de http://www.spam.com.

Tão controversas quanto as definições são as explicações para a origem do termo *spam*. A palavra SPAM, escrita em letras maiúsculas, é uma marca registrada pela Hormel Foods LLC [Hormel Foods, 2000]. A Hormel Foods LLC é uma empresa de alimentos e o nome SPAM pode ter surgido de uma contração das palavras "*SPiced hAM*" que batizou um dos seus produtos. O SPAM, um presunto enlatado como mostra a Figura 5.1, se tornou conhecido em 1937 durante uma campanha publicitária e em seguida pela utilização durante a segunda guerra mundial pelo exército americano. Como a carne era racionada às tropas, o SPAM era o alimento largamente consumido. Ao retornar aos EUA, todos os soldados americanos recebiam uma medalha que ficou conhecida como medalha *spam*. Como a condecoração foi recebida por diversas pessoas, a palavra *spam* ficou associada a algo comum [Holmes, 2005]. Anos mais tarde, o grupo de comédia Monty Python filmou um esquete em que um cliente entra em um restaurante e pergunta à garçonete quais são os pratos do cardápio. A garçonete, então, cita cada um dos pratos e todos contêm SPAM. Dessa forma, a palavra *spam* é repetida muitas vezes em pouco tempo. Por isso, *spam* se tornou sinônimo de algo repetitivo e sem sentido como a maioria das mensagens eletrônicas não solicitadas [Hambridge e Lunde, 1999].

O registro histórico do primeiro *spam* é de 1978. Um anúncio de uma demonstração de produtos foi enviado na Arpanet por um funcionário do departamento de vendas

da Digital Equipment Corporation (DEC), uma fabricante de computadores. Devido à limitação de espaço destinado ao endereço dos destinatários dos programas de correio eletrônico usados na época, a mensagem foi encaminhada para "apenas" 320 destinatários. Na época, o anúncio da DEC causou surpresa e gerou um debate sobre se era correto ou não utilizar o correio eletrônico para tal finalidade. O segundo caso emblemático de envio de *spams* ocorreu em março de 1994 e foi uma propaganda enviada, de forma automatizada, para seis mil grupos de discussão de um fórum da Internet. Um casal de advogados enviou uma mensagem anunciando que o prazo de inscrições na loteria de vistos de trabalho americanos estava próximo e ofereciam seus serviços a imigrantes interessados. Tal mensagem é conhecida como o *spam* do Green Card [Canter e Siegel, 1994]. O fato de uma propaganda ter sido enviada para um fórum sem nenhuma relação com o tema em discussão revoltou grande parte dos usuários. Segundo alguns relatos, foi durante o debate sobre o anúncio dos advogados que surgiu a primeira associação entre o tipo de mensagem enviada e a palavra *spam*.

Atualmente, existem diversos tipos de *spams* [Cukier et al., 2006] que vão desde simples anúncios de produtos até tentativas de golpes financeiros contra os destinatários das mensagens. Os mais comuns são os de conteúdo comercial, que representam 74% do volume total de *spams*, como mostra a Figura 5.2. As mensagens comerciais anunciam, por exemplo, a venda de medicamentos sem prescrição médica, tratamentos estéticos, oportunidades de enriquecimento rápido e sítios de conteúdo pornográfico. Embora existam *spams* enviados por empresas conhecidas e que contêm ofertas verdadeiras, grande parte das mensagens tem origem e conteúdos suspeitos.

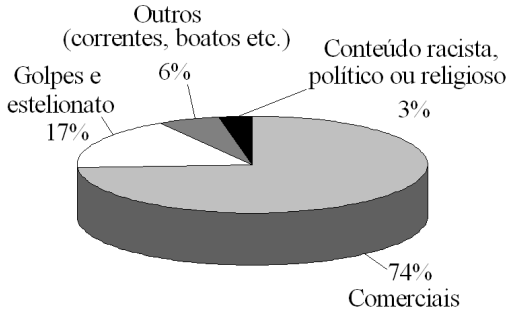
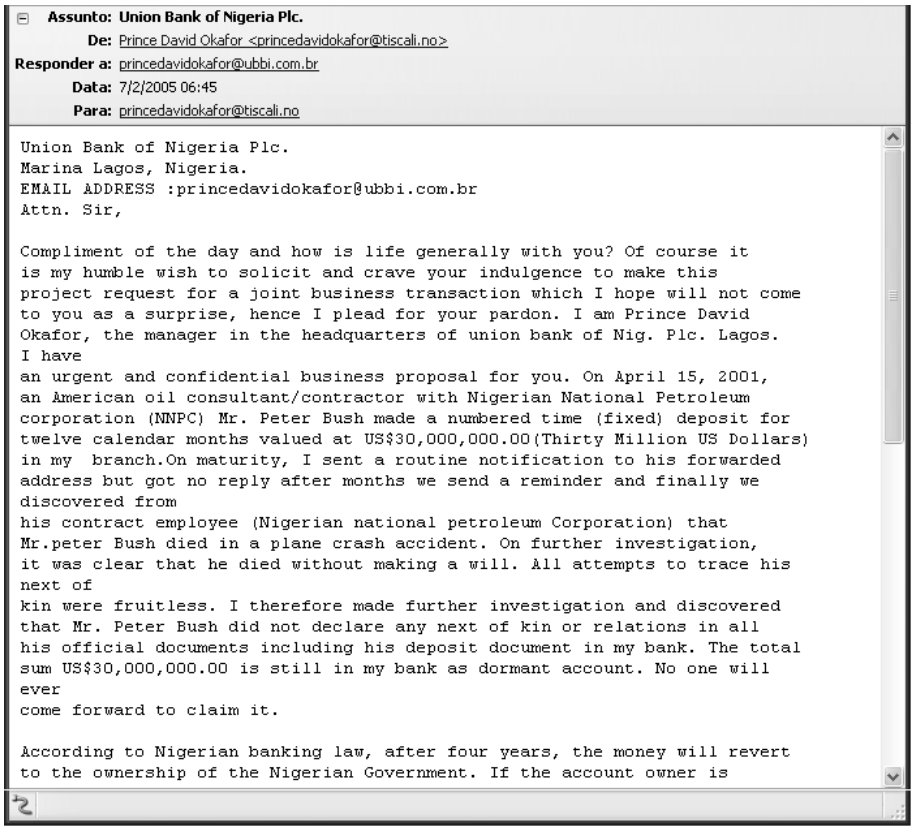


Figura 5.2. Estatísticas sobre os tipos de *spams* (adaptação) [Spammer-X et al., 2004].

Outro tipo de *spam* bastante comum é o que contém correntes ou boatos. As correntes são mensagens que prometem algum tipo de benefício, como dinheiro e saúde, a quem encaminhá-las para um determinado número de destinatários em um dado intervalo de tempo. Do contrário, se o usuário não encaminhar a mensagem sofrerá as conseqüências. Os boatos são *spams* que buscam impressionar os usuários com as falsas histórias que contêm. Essas histórias tratam, por exemplo, da busca por crianças desaparecidas, de ameaças de vírus de computador e da difamação de empresas e pessoas. O objetivo de quem envia tanto correntes quanto boatos é divulgar o conteúdo dessas mensagens para o maior número pessoas em um curto espaço de tempo e, conseqüentemente, torná-la uma lenda urbana.

Os *spams* com os efeitos mais nocivos para os destinatários são os que contêm ten-

tativas de golpes e fraudes e os que tentam disseminar vírus e outros códigos maliciosos. Toda mensagem não solicitada que contém alguma fraude ou tentativa de golpe é chamada de *scam*. Tais mensagens contêm histórias que servem de pano de fundo para que o destinatário execute uma determinada ação desejada pelo *scammer*. As histórias descritas nos *scams*, em sua maioria, tratam de ofertas de produtos que prometem resultados enganosos, oferecem oportunidades de negócios miraculosos ou até mesmo informam ao usuário que ele acabou de "ganhar" na loteria. Em troca, é solicitada alguma recompensa ao destinatário. Um dos golpes mais conhecidos da Internet, ilustrado na Figura 5.3, é a mensagem enviada por um suposto cidadão nigeriano que, por razões políticas e pessoais, está disposto a transferir uma grande quantidade de dinheiro para o destinatário. A condição para que o destinatário seja beneficiado é depositar uma pequena quantia em dinheiro como garantia em uma conta bancária a ser indicada. As mensagens com golpes semelhantes são classificadas como 419 em alusão ao número da lei nigeriana sobre a prática de fraudes.



Assunto: Union Bank of Nigeria Plc.
De: Prince David Okafor <princedavidokafor@tiscali.no>
Responder a: princedavidokafor@ubbi.com.br
Data: 7/2/2005 06:45
Para: princedavidokafor@tiscali.no

Union Bank of Nigeria Plc.
Marina Lagos, Nigeria.
EMAIL ADDRESS :princedavidokafor@ubbi.com.br
Attn. Sir,

Compliment of the day and how is life generally with you? Of course it
is my humble wish to solicit and crave your indulgence to make this
project request for a joint business transaction which I hope will not come
to you as a surprise, hence I plead for your pardon. I am Prince David
Okafor, the manager in the headquarters of union bank of Nig. Plc. Lagos.
I have
an urgent and confidential business proposal for you. On April 15, 2001,
an American oil consultant/contractor with Nigerian National Petroleum
corporation (NNPC) Mr. Peter Bush made a numbered time (fixed) deposit for
twelve calendar months valued at US$30,000,000.00(Thirty Million US Dollars)
in my branch.On maturity, I sent a routine notification to his forwarded
address but got no reply after months we send a reminder and finally we
discovered from
his contract employee (Nigerian national petroleum Corporation) that
Mr.peter Bush died in a plane crash accident. On further investigation,
it was clear that he died without making a will. All attempts to trace his
next of
kin were fruitless. I therefore made further investigation and discovered
that Mr. Peter Bush did not declare any next of kin or relations in all
his official documents including his deposit document in my bank. The total
sum US$30,000,000.00 is still in my bank as dormant account. No one will
ever
come forward to claim it.

According to Nigerian banking law, after four years, the money will revert
to the ownership of the Nigerian Government. If the account owner is

Figura 5.3. Um exemplo de um *spam* 419.

O estelionato também é um tipo de mensagem de golpe, em que é utilizada uma isca para roubar dados pessoais e/ou bancários do destinatário. Por utilizarem uma isca para "pescar" os dados do destinatário, essa atividade é chamada de *phishing* em referência ao verbo *fishing* do inglês. As iscas geralmente são mensagens com solicitações de recadastramento de dados em bancos, em administradoras de cartão de crédito e até mesmo em órgãos públicos, como a receita federal. Para que o usuário execute o que é pe-

dido, as iscas tentam se aproximar ao máximo de uma mensagem legítima, supostamente enviada por uma instituição acima de qualquer suspeita. Um exemplo de estelionato é uma mensagem enviada em nome do Banco do Brasil oferecendo ao destinatário um seguro contra fraudes. De acordo com o conteúdo da mensagem, o destinatário deve clicar em um atalho contido no *spam* para que o seguro seja ativado. O atalho o leva a uma página com um formulário que contém campos, como senha e conta corrente, a serem preenchidos. Para induzir o usuário a preencher os campos, o formulário falso é bastante semelhante ao formulário real utilizado pelo banco, como mostra a Figura 5.4. Uma vez que o destinatário preenche os campos e clica no botão de submissão, seus dados são enviados ao *spammer*.

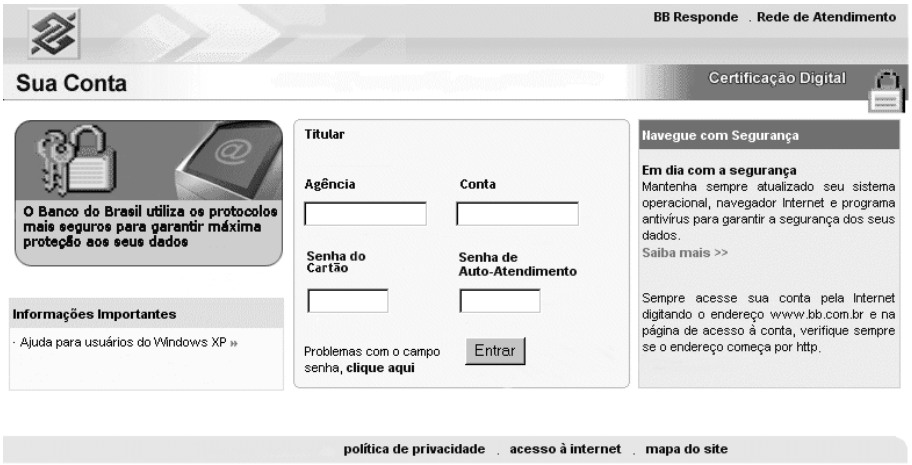
Os *spams* também são usados para difundir programas maliciosos como vírus, vermes e cavalos de tróia. Assim como nas mensagens de golpe e estelionato, os *spams* com programas maliciosos possuem algum tipo de isca para disfarçar o seu real conteúdo. Tenta-se com isso, induzir o destinatário a executar o programa enviado junto à mensagem ou fazer com que ele clique em um atalho que o leve a executar o programa hospedado em um dado sítio da Internet. Um dos objetivos da disseminação de programas maliciosos é recrutar máquinas zumbis para enviar cada vez mais *spams*, como será visto na Seção 5.3.4. Os programas maliciosos também são usados para capturar dados do destinatário. Um dos exemplos de *spam* dessa natureza é a mensagem que informa ao destinatário que ele está sendo traído, reproduzida na Figura 5.5. Para que o destinatário veja as fotos que comprovam a suposta traição conjugal, ele deve clicar no atalho indicado na mensagem. Ao clicar, o destinatário executa um cavalo de tróia que deixa o seu computador vulnerável às ações dos *spammers*. Deve ser observado na barra inferior à esquerda da Figura 5.5 que o atalho contido no *spam* é um programa executável do sistema operacional Windows.

Existem ainda mensagens com conteúdo racista, político ou religioso que correspondem a 3% do total das *spams* enviados.

5.2.2. Motivação para o envio de *spams*

Existem três razões para a proliferação dos *spams* na Internet: a facilidade para se obter endereços de potenciais consumidores, o baixo custo para enviá-los e o número de destinatários alcançados com apenas uma mensagem.

O correio eletrônico se tornou uma aplicação de grande popularidade por facilitar a comunicação entre pessoas e pelo baixíssimo custo para se enviar uma mensagem. Em virtude do sucesso dessa aplicação, a divulgação de endereços de correio eletrônico em sítios pessoais, de empresas e de instituições de ensino se tornou uma prática comum. Aproveitando-se da forma como estes endereços são divulgados eletronicamente, os *spammers* constroem a lista de destinatários de suas mensagens. Para isso, são utilizados programas, denominados robôs, que vasculham de forma automatizada os sítios da Internet em busca de endereços de correio eletrônico. Utilizando um desses programas, é possível se obter, em poucas horas e com um custo muito baixo, milhares de endereços de correio eletrônico. Mais detalhes sobre a aquisição de endereços são apresentados na Seção 5.3.2. Devido a eficácia dos *spams*, criou-se um grande comércio de listas de endereços de correio eletrônico. Uma lista com milhões de endereços pode custar de US$



(a) O formulário falso.

(b) O formulário verdadeiro.

Figura 5.4. Um exemplo de um *spam* de estelionato.

100,00 a US$ 1000,00 [Spammer-X et al., 2004]. Dessa forma, a criação e a comercialização dessas listas se tornaram uma fonte de renda atrativa, devido à sua lucratividade e ao fato de que nenhum ato ilegal está sendo cometido por quem pratica tal atividade.

Em comparação com outros meios usados na divulgação de propagandas, o correio eletrônico é o que possui o menor custo e o maior alcance geográfico. Os anúncios impressos em papel são realizados por correio convencional ou através de mensageiros e requerem gastos com a criação, a reprodução e a distribuição das mensagens. A TV e o rádio são os veículos de massa mais utilizados por anunciantes para divulgarem produ-

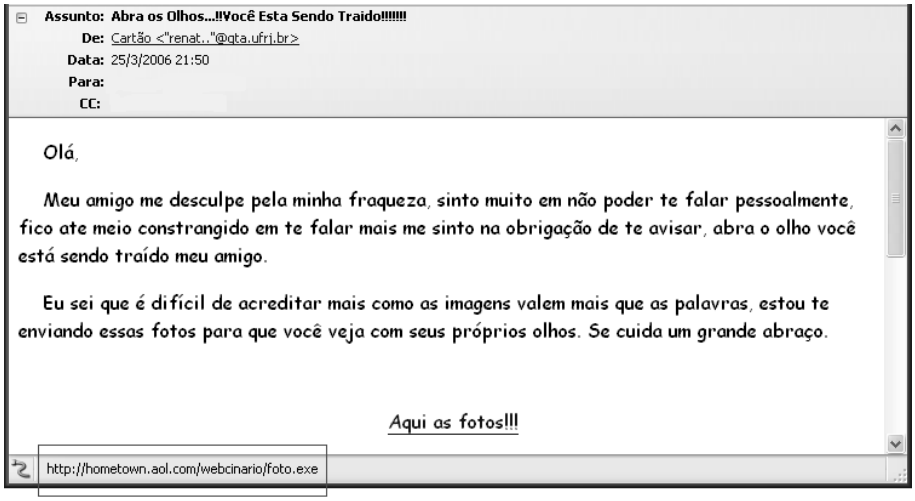


Figura 5.5. Um *spam* que induz o destinatário a executar um cavalo de tróia.

tos, serviços, ofertas e novas idéias. O impacto sonoro do rádio e o audiovisual da TV são imbatíveis. Porém, apesar do grande poder de penetração desses veículos, as propagandas na TV e no rádio estão limitadas geograficamente a regiões, estados e países e, além disso, requerem que o espectador esteja fisicamente em frente ao aparelho receptor durante a apresentação do anúncio. Deve-se considerar ainda as elevadas taxas pagas às agências e emissoras de radiodifusão pela produção e pela veiculação dos anúncios. Na busca por novos meios de divulgação, a Internet se apresenta como um veículo de grande alcance de potencias consumidores, de baixo custo e que cada vez mais possui impacto audiovisual devido às novas tecnologias. Por isso, mesmo não tendo sido criado para tal finalidade, o correio eletrônico se tornou um poderoso veículo de divulgação, em virtude da simplicidade do protocolo SMTP usado para enviar as mensagens. O protocolo SMTP foi desenvolvido pressupondo-se que as mensagens eletrônicas seriam trocadas entre remetentes e destinatários que confiavam uns nos outros e que as mensagens trocadas entre estes seriam relevantes para ambos. Essa situação foi logo explorada pelos *spammers*. Como são os responsáveis pelo encaminhamento das mensagens de correio eletrônico até os destinatários, os provedores de serviço herdam boa parte dos custos de divulgação das propagandas contidas nos *spams*. Dessa forma, a um custo reduzido, um único *spam* com uma propaganda pode alcançar milhares de destinatários, sem limites geográficos e sem grande força de trabalho. Estima-se que para enviar um milhão de *spams*, um *spammer* gaste US$ 250,00 [Pfleeger e Bloom, 2005].

O baixo custo de produção e de divulgação e o enorme número de destinatários alcançados são, sem dúvida, grandes estimulantes para o envio de *spams*. Porém, nenhuma característica motiva mais o envio de *spams* do que a efetividade dessas mensagens. Pode parecer surpreendente, mas, segundo uma pesquisa de uma empresa de consultoria em segurança da informação, 39% dos usuários de correio eletrônico entrevistados já clicaram em um atalho contido em *spams* [Cukier et al., 2006]. Foram entrevistados tanto usuários corporativos quanto domésticos. Dos usuários corporativos, 13% já compraram algum produto anunciado por um *spam*. Entre os usuários domésticos esse número é de

11%. Outro estudo [Pfleeger e Bloom, 2005] mostra que os catálogos de produtos enviados através de mensagens eletrônicas geram doze vezes mais respostas do que os catálogos impressos enviados através do correio tradicional. Além disso, em virtude do grande volume de mensagens enviadas, mesmo uma taxa de resposta baixa já representaria uma grande possibilidade de lucro para os anunciantes.

Recentemente, o envio de *spams* também tem sido motivado pela possibilidade de enriquecimento ilícito por parte dos *spammers*. Visto que muitos usuários seguem as determinações indicadas nos *spams*, o número de mensagens contendo tentativas de golpes cresce a cada dia e essas mensagens se tornam mais sofisticadas. É comum se ter notícia de pessoas que foram lesadas financeiramente após terem clicado em atalhos para sítios suspeitos contidos em mensagens eletrônicas. Sem saber, essas pessoas forneceram dados pessoais e financeiros aos *spammers* que de posse desses dados podem efetuar compras em cartões de crédito e transferências bancárias. Na maioria das vezes, quadrilhas bem organizadas estão por trás dessas fraudes, mas há alguns anos já é possível se notar a ação da polícia no combate a esse tipo de crime.

5.2.3. Prejuízos causados pelo envio de *spams*

O envio de mensagens eletrônicas não solicitadas acarreta em prejuízos para os usuários de correio eletrônico e para os provedores de serviço da Internet. Do ponto de vista dos usuários, os *spams* incomodam tanto pelo conteúdo quanto pela quantidade de mensagens recebidas. Estima-se que atualmente um usuário receba milhares de mensagens não solicitadas por ano [Pfleeger e Bloom, 2005]. Além do incômodo, os *spams* podem afetar a produtividade e a segurança dos usuários. Para cada mensagem não solicitada recebida, o usuário tem que baixá-la para o seu computador, abri-la e identificá-la como um *spam*. Isto provoca desperdício de tempo e gastos desnecessários, uma vez que o usuário paga para acessar a Internet. Em um ambiente de trabalho, os *spams* aumentam o tempo gasto pelos funcionários na leitura de mensagens eletrônicas. Estima-se que as empresas gastem cerca de US$ 1300,00 por ano com cada empregado por causa dos *spams* [Cukier et al., 2006]. Um usuário também pode sofrer com o não recebimento de mensagens legítimas em função dos *spams*. Na tentativa de reduzir o volume de mensagens não solicitadas, os provedores de serviço utilizam filtros para bloqueá-las. Como será visto em detalhes na Seção 5.4, esses filtros podem classificar mensagens legítimas como *spams* e, conseqüentemente, o usuário deixa de recebê-las. Além disso, em virtude do volume de *spams* recebidos, um usuário pode ter a sua caixa de correio completamente ocupada, o que o impede de receber novas mensagens. Porém, a ameaça mais grave aos usuários de correio eletrônico são as mensagens com tentativas de golpe. De acordo com uma pesquisa [Cukier et al., 2006], cerca de 4% dos entrevistados que utilizam correio eletrônico no trabalho e 11% dos que são usuários domésticos já perderam dinheiro com golpes enviados através de *spams*. Todos esses fatores contribuem para que a credibilidade dos usuários na aplicação de correio eletrônico e também na Internet diminua.

Para os provedores de serviço, os *spams* representam um prejuízo de bilhões de dólares. A cada um milhão de *spams* enviados, os provedores perdem US$ 2800, o que anualmente provoca um prejuízo de US$ 8,9 bilhões para as empresas americanas e US$ 2,5 bilhões para as empresas européias [Emery, 2003, Pfleeger e Bloom, 2005]. Os provedores de serviços são os responsáveis por encaminhar as mensagens eletrônicas até os

seus destinatários, inclusive as não solicitadas que atualmente correspondem a mais da metade do tráfego total de correio eletrônico. Além disso, muitos provedores oferecem gratuitamente serviços de correio eletrônico que são utilizados pelos *spammers* para enviarem suas mensagens. Assim, os provedores arcam com custos desnecessários de banda passante, memória e processamento para receber, armazenar e processar as mensagens eletrônicas não solicitadas que, ao serem recebidas, serão simplesmente movidas para a lixeira pela grande maioria dos destinatários. Portanto, além do aumento de custos com infra-estrutura, os provedores de serviço também têm que investir em soluções para combater os *spams* e, conseqüentemente, aumentam seus gastos com pessoal. É necessário aumentar o número de funcionários do serviço de atendimento ao cliente, que recebe inúmeras reclamações relativas ao recebimento de *spams*, e também formar uma equipe de manutenção para a partir das reclamações realimentar o sistema de combate aos *spams*. Para se ter uma idéia do custo de pessoal, a America Online estima que apenas dois *spammers*, que enviaram dois bilhões de *spams*, foram responsáveis por oito milhões de reclamações de usuários [Krim, 2003].

5.2.4. Legislação atual

Uma das etapas do processo para coibir o envio de mensagens não solicitadas na Internet é a definição de uma legislação sobre o tema. O estabelecimento de leis é fundamental para que se possa determinar quais mensagens podem ser consideradas como *spams* e também quais as punições devem ser aplicadas aos responsáveis pelo envio dessas mensagens.

Os Estados Unidos são pioneiros na discussão de leis anti-*spam*, provavelmente, porque cerca de 35% dos *spams* são originados no país [Commtouch, 2006]. Para tentar controlar e definir regras para o envio de *spams*, em 2004 entrou em vigor o estatuto CAN-SPAM (*Controlling the Assault of Non-Solicited Pornography and Marketing Act*). Este estatuto, válido em todo o território americano, foi definido pelo FTC (*Federal Trade Commission*), órgão responsável pela legislação anti-*spam* nos EUA, com base em leis estaduais já existentes. Para o FTC, um *spam* é qualquer mensagem eletrônica de conteúdo comercial enviada, geralmente em batelada, para um consumidor sem a requisição ou consentimento prévio desse consumidor [FTC, 2005]. Essa definição é a base das regras estabelecidas pelo CAN-SPAM. De acordo com o estatuto:

- uma mensagem eletrônica deve conter informações verdadeiras a respeito da sua origem;
- o campo de assunto da mensagem deve estar relacionado com o próprio conteúdo;
- caso seja uma propaganda, uma mensagem eletrônica deve indicar claramente o seu propósito;
- o endereço físico do remetente deve estar presente na mensagem para que o destinatário possa enviar reclamações ou denúncias;
- uma mensagem eletrônica deve fornecer ao destinatário a opção de não receber mais mensagens semelhantes do mesmo remetente.

O CAN-SPAM prevê penas criminais para os remetentes que não respeitarem as regras definidas no estatuto. As penas vão desde pagamentos de multa, para quem viola alguma das regras do estatuto, até a prisão do remetente, no caso, por exemplo, em que um *spammer* utiliza, sem autorização, o computador de terceiros para enviar as mensagens não solicitadas. As multas podem chegar a até US$ 11000. O CAN-SPAM também estipula multas para quem realizar ataques de dicionário para gerar endereços de possíveis destinatários dos *spams* e para quem coleta endereços de correio eletrônico em sítios da Internet (Seção 5.3.2). Ainda segundo o estatuto, os sítios devem conter mensagens explícitas informando sobre a proibição do uso dos endereços disponibilizados para o envio de *spams*. Também estão sujeitos a multa os indivíduos que usam programas para automatizar o registro de múltiplos endereços em sítios de serviço gratuito de correio eletrônico com o objetivo de enviar *spams*. O pagamento de multa também é previsto para quem se aproveita, sem autorização, de uma falha de configuração de servidores de correio eletrônico para enviar *spams*. O estatuto também pune com prisão quem tenta iludir e despistar destinatários e provedores de serviços sobre a verdadeira origem das mensagens não solicitadas, quem falsifica informações no cabeçalho de múltiplas mensagens e inicia a transmissão dessas mensagens, quem utiliza endereços IP falsos para enviar *spams* e quem, com o mesmo objetivo, cria contas de correio eletrônico e registra nomes de domínios usando informações falsas. De acordo com o CAN-SPAM, o termo "múltiplas" significa mais de 100 mensagens eletrônicas em um período de 24 horas, mais de 1000 mensagens em um período de 30 dias ou mais de 10000 mensagens durante o período de um ano.

A principal crítica ao estatuto CAN-SPAM é que ele mostra claramente aos *spammers* os limites nos quais eles podem atuar. Isso ocorre, pois o estatuto define legalmente o que é um *spam*. Dessa forma, é possível criar um *spam* que esteja de acordo com a regras definidas no estatuto e, conseqüentemente, dentro da lei [Spammer-X et al., 2004]. Além disso, mesmo que uma mensagem eletrônica viole as regras do estatuto CAN-SPAM, a identificação de quem a enviou é difícil, pois uma mensagem pode ser enviada com um remetente falso, como é visto na Seção 5.3.3. Outro ponto criticado no estatuto CAN-SPAM é a tentativa de balancear a manutenção da liberdade de expressão com a inibição ao envio de *spams*, o que deixou enormes brechas para que os *spammers* continuem suas atividades sem serem perturbados [Hoanca, 2006]. Também há quem diga que as medidas legais serão sempre ineficientes, pois não são ágeis o suficiente para acompanhar a evolução da tecnologia.

Por todos esses fatores, não houve redução no envio de *spams* após a entrada em vigor da legislação anti-*spam* americana. O volume de mensagens não solicitadas continua a crescer, o que comprova que as medidas legais ainda são pouco efetivas no combate aos *spams*.

No Brasil, ainda não há nenhuma legislação anti-*spam* em vigor. O que existem são projetos de lei em debate na Câmara e no Senado Federal, como o projeto 021/04 que recebeu parecer favorável na Comissão de Constituição, Justiça e Cidadania (CCJ) do Senado. Este projeto define um *spam* como sendo uma mensagem eletrônica com conteúdo comercial ou publicitário enviada a mais de 500 destinatários durante um período de 96 horas. Além disso, esta mensagem só pode ser enviada aos destinatários que autorizarem previamente o seu recebimento, deve deixar claro o seu objetivo, deve conter a verdadeira

identidade do remetente e deve possuir algum mecanismo para que o destinatário possa optar por não receber mais mensagens semelhantes do mesmo remetente. Como não há nenhuma regulamentação sobre o assunto em vigor no país, o objetivo de qualquer mensagem não solicitada que mencione estar em conformidade com as leis brasileiras é enganar os destinatários. Além do debate sobre projetos de lei, existem alguns grupos de discussão [CGI.BR, 2006, Grupo Brasil AntiSPAM, 2006b] que propõem cartilhas para informar aos usuários os perigos relativos aos *spams* e normas de conduta para os usuários de correio eletrônico. Um desses grupos é o Brasil AntiSPAM, que define um conjunto de regras para definir o que é uma mensagem não solicitada [Grupo Brasil AntiSPAM, 2006a]. Para este grupo, um *spam* é toda mensagem eletrônica com pelo menos duas das características a seguir:

- o remetente é inexistente ou possui identidade falsa;
- o destinatário não autorizou previamente o envio da mensagem;
- o destinatário não pode optar em não receber mais a mensagem;
- o assunto não condiz com o conteúdo da mensagem;
- a sigla NS (Não Solicitado) está ausente no campo de assunto de uma mensagem que não foi previamente requisitada;
- o remetente não pode ser identificado;
- uma mensagem semelhante foi recebida anteriormente em menos de dez dias apenas com os campos de remetente ou de assunto diferentes.

A proposta de estabelecer leis específicas para o combate aos *spams* no Brasil pode ter efeitos tão ineficientes quanto o estatuto CAN-SPAM nos EUA. Por isso, existe quem defenda a adaptação de artigos do código penal brasileiro [Decreto-lei n^o 2.848, 1940] para regulamentar a prática de enviar *spams* e aplicar punições aos eventuais infratores. Os artigos que tratam da usurpação (Artigo 161 do Capítulo III do Título II), de danos (Artigo 163 do Capítulo IV do Título II) e do estelionato (Artigo 171 do Capítulo VI do Título II) podem ser aplicados aos *spammers* que invadem máquinas de terceiros para enviar, sem autorização, mensagens não solicitadas e também aos que enviam mensagens com tentativas de golpe. A pena pode ir de detenção de 1 a 6 meses e multa no caso da usurpação, detenção de 6 meses a 3 anos e multa no caso de danos e, por fim, reclusão de um a cinco anos e multa no caso de estelionato. O texto destes artigos está reproduzido na íntegra no Apêndice A. Há também quem defenda a utilização dos Artigos 36 e 37 da Seção III do Capítulo V do Título I do código de defesa do consumidor [Lei n^o 8.078, 1990] que punem, respectivamente, a publicidade velada e a prática da propaganda abusiva. A reprodução destes artigos está no Apêndice B.

Como visto, as medidas legais não são suficientes para coibir a prática do envio de mensagens não solicitadas. Um dos fatores que contribuem para essa situação é a simplicidade do sistema de correio eletrônico da Internet, que possibilita o envio de mensagens

sem a confirmação da autenticidade do remetente. Dessa forma, a identificação de indivíduos que praticam atos ilícitos através de *spams* é bastante difícil e, por isso, a cada dia surgem novas técnicas para enviar mensagens não solicitadas. Algumas dessas técnicas são apresentadas na seção seguinte.

5.3. Técnicas para o envio de *spams*

O envio de *spams* engloba três fases principais. A primeira fase corresponde à obtenção de uma grande quantidade de endereços de correio eletrônico para a elaboração de uma lista de destinatários. A segunda compreende a criação da mensagem que será enviada. Por fim, é necessário um meio para enviar as mensagens. Nesta seção são apresentados o sistema de correio eletrônico da Internet, os aspectos de cada uma das fases de envio dos *spams* e como os *spams* evoluíram ao longo do tempo. Esta evolução é observada a partir de uma base de dados construída pelos autores com mais de oito mil mensagens não solicitadas recebidas nos últimos três anos. Vários exemplos de *spams* apresentados a seguir foram retirados dessa base de dados.

5.3.1. Sistema de correio eletrônico da Internet

O sistema de correio eletrônico da Internet é composto de agentes de usuário (*User Agents* - UAs), de servidores de correio ou agentes de transferência de mensagens (*Message Transfer Agents* - MTAs), de um protocolo simples de transferência de correio (*Simple Mail Transfer Protocol* - SMTP) e de protocolos de acesso a correio. A Figura 5.6 mostra os componentes do sistema de correio eletrônico e ilustra o envio de uma mensagem. Os agentes de usuário permitem que usuários leiam, respondam, reencaminhem, salvem e editem mensagens. Alguns dos principais agentes de usuário são o Outlook, o Eudora, o Thunderbird e o Mutt. Os servidores de correio armazenam as mensagens e se comunicam com outros servidores para realizar a transferência das mensagens. O protocolo SMTP transfere as mensagens entre servidores de correio e pode ser usado também nas comunicações entre o agente do usuário e o servidor de correio do usuário. O protocolo SMTP é normalmente executado em segundo plano como um *daemon* do sistema. Por último, os protocolos de acesso ao correio transferem mensagens do servidor de correio do usuário para o agente do usuário. O POP (*Post Office Protocol*), o IMAP (*Internet Message Access Protocol*) e o HTTP (*HyperText Transfer Protocol*) são exemplos destes protocolos.

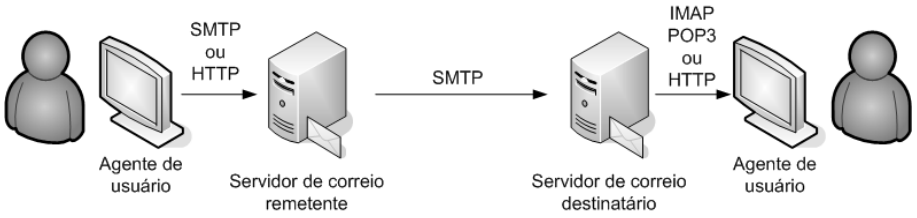


Figura 5.6. Os componentes do sistema de correio eletrônico da Internet.

Dentre os componentes de um sistema de correio eletrônico da Internet, o protocolo SMTP [Klensin, 2001] é o principal envolvido no envio de *spams*. O SMTP utiliza

o protocolo TCP (*Transmission Control Protocol*) e a porta 25 para enviar mensagens em ASCII de 7 bits[1] entre um cliente SMTP (transmissor) e um servidor SMTP (receptor).

O protocolo SMTP utiliza uma série de comandos para fazer a comunicação entre servidores de correio. Os principais comandos utilizados são o `HELO`, o `MAIL FROM`, o `RCPT TO`, o `DATA`, o `QUIT` e o `VRFY`. Os comandos, em sua maioria, são simples e auto-explicativos. O comando `VRFY` é utilizado para verificar a existência de um usuário ou de uma caixa de correio. Um exemplo de interação fictícia entre um cliente SMTP (C) e um servidor SMTP (S) é apresentado na Figura 5.7.

```
S: 220 servidor.br
C: HELO cliente.br
S: 250 Hello cliente.br, pleased to meet you
C: MAIL FROM: <usuario@cliente.br>
S: 250 usuario@cliente.br... Sender ok
C: RCPT TO: <usuario@servidor.br>
S: 250 usuario@servidor.br ... Recipient ok
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: From: usuario@cliente.br
C: To: usuario@servidor.br
C: Subject: Teste
C:
C: Teste de envio de correio.
C: .
S: 250 Message accepted for delivery
C: QUIT
S: 221 servidor.br closing connection
```

Figura 5.7. Um exemplo de troca de mensagens entre um cliente e um servidor SMTP.

Como no correio convencional, um correio eletrônico é formado por um envelope e uma mensagem. O envelope encapsula a mensagem e contém todas as informações necessárias para o transporte da mensagem do remetente até o destinatário. Os servidores de correio utilizam os envelopes para o transporte das mensagens.

Uma mensagem é composta de um cabeçalho e de um corpo. Os principais campos de cabeçalho são: `From:`, `To:`, `Subject:` e `Received:`. Um exemplo de uma mensagem fictícia simplificada é apresentado na Figura 5.8. Nesse exemplo, pode-se perceber que os campos `Received:` indicam a máquina de origem da mensagem (maquina.cliente.br) e os servidores SMTP pelos quais a mensagem passou, do último para o primeiro. Além disso, o cabeçalho é separado do corpo por uma linha em branco. O corpo da mensagem só diz respeito ao destinatário e em nada importa para os servidores.

```
Received: from cliente.br by servidor.br; 16 Jul 06 10:30:01 GMT
Received: from maquina.cliente.br by cliente.br; 16 Jul 06 10:29:58 GMT
From: usuario@cliente.br
To: usuario@servidor.br
Subject: Teste

Teste de envio de correio.
```

Figura 5.8. Um exemplo de mensagem eletrônica simplificada.

[1] Para transmitir dados binários ou em ASCII de 8 bits é necessário o uso de codificação.

Um dos principais problemas do uso do SMTP para enviar correios eletrônicos é a falta de um mecanismo de autenticação para assegurar a verdadeira identidade do remetente. Isso faz com que *spammers* possam facilmente fazer uma mensagem parecer que tenha sido originada de qualquer endereço de correio eletrônico. Para evitar isso, alguns provedores de serviço exigem o uso da extensão SMTP-AUTH [Myers, 1999], que permite a verdadeira identificação do remetente do correio eletrônico. Contudo, essa extensão não garante a autenticidade do remetente do envelope ou do campo `From:` da mensagem. Um usuário autenticado pode falsificar esses valores.

5.3.2. Coleta de dados

A coleta de dados é o primeiro passo a ser realizado antes do envio de um *spam*. É nessa etapa que os *spammers* constroem as listas de destinatários de suas mensagens. Quanto maior o número de endereços válidos conseguidos nessa etapa melhor será o resultado das etapas seguintes. As principais técnicas de obtenção de endereços de correio eletrônico se baseiam na varredura de lugares onde são divulgados endereços, na invasão ou exploração de sítios para obter o cadastro dos seus usuários, na utilização de ataques de força bruta, no uso de ofertas ou concursos e na compra de listas prontas de endereços.

A forma mais simples de obtenção de endereços eletrônicos corresponde à varredura de lugares onde endereços de correio eletrônico são divulgados, como grupos de notícias (*newsgroups*), listas de distribuição (*mailing lists*), salas de bate-papo, *bulletin boards*, bases de dados livres e páginas *web*. Essa técnica surgiu no fim da década de 90 e é conhecida como coleta de endereços de correio eletrônico (*harvesting*). Inicialmente a busca era realizada manualmente, através da procura pelo símbolo "@" e da extração do endereço correspondente. Com o passar do tempo, foram introduzidos robôs automatizados para realizar essa tarefa repetitiva com maior rapidez e menor esforço. O grande número de sítios e de listas de mensagens verificados em um curto espaço de tempo fizeram com que os desenvolvedores de páginas *web* começassem a utilizar uma série de artifícios para dificultar a coleta dos endereços. Alguns trocam o símbolo "@' por "at", enquanto outros procuram colocar espaços em branco entre os caracteres do endereço de correio eletrônico ou até mesmo utilizam figuras que contêm o endereço ao invés de texto. Contudo, em páginas *web* que contém uma forma de contato na qual basta ao usuário clicar sobre um atalho com o comando `mailto:`, o trabalho de busca pelos robôs é facilitado. De modo a evitar esses problemas, pode-se substituir o comando `mailto:` por um código em JavaScript com a mesma funcionalidade. As listas de discussão também são bastante usadas pelos *spammers* que podem procurar endereços em listas abertas ou se inscrever em uma lista fechada de modo a ter acesso às mensagens trocadas. As mensagens, pertencentes a listas de discussão ou não, com diversos endereços no campo `To:` facilitam a atuação dos *spammers*. Esse fato é agravado quando uma mensagem é encaminhada (*forwarded*) e, conseqüentemente, vários endereços são expostos para outros destinatários. Nesse caso, aconselha-se a utilização do campo `Bcc:` no lugar do `To:`, principalmente no caso do encaminhamento de uma mensagem com um número grande de destinatários [Walker, 2005]. A base de dados *whois* com dados de responsáveis por domínios também é explorada por programas automatizados que capturam endereços. Para os *spammers*, um dos principais problemas destas buscas em lugares onde endereços são divulgados é a falta de correlação dos usuários listados com os produtos oferecidos pelos

spammers, exceto quando a coleta é feita em uma lista sobre um determinado assunto e a oferta está relacionada ao tema.

Os ataques ou invasões de sítios também têm sido bastante explorados com o intuito de obter endereços eletrônicos. Por mais surpreendente que possa parecer, às vezes determinadas informações como números de cartões de crédito são deixadas de lado quando o objetivo principal é a obtenção de endereços de correio eletrônico e de informações relacionadas. Essas invasões procuram explorar falhas de segurança existentes nos sistemas operacionais e nos aplicativos. Um dos principais alvos dos *hackers* são as lojas *on-line*. Essas lojas contêm bases de dados que geralmente utilizam comunicação criptografada com o servidor, mas armazenam os dados dos clientes em texto simples (não formatado) [Spammer-X et al., 2004]. A invasão de servidores específicos, como de um sítio de um clube de futebol, pode dar ao *spammer* listas de usuários identificados com o produto a ser anunciado. Programas maliciosos como cavalos de tróia, vírus e vermes também são usados para obter a lista de endereços dos usuários do computador afetado. Esses programas acessam os dados, criam as listas de endereços e enviam as informações para os *spammers*.

Outra técnica bastante utilizada pelos *spammers* consiste no uso de dicionários para realizar ataques de força bruta ou verificação de massa, através da adivinhação dos nomes dos usuários em domínios da Internet. Domínios muito populares, principalmente dos provedores gratuitos de correio eletrônico, contêm milhares de usuários que geralmente utilizam seus nomes, apelidos e nomes de seus personagens favoritos do cinema e dos quadrinhos como identificadores de contas de correio eletrônico. Certamente um dos nomes mais comuns é jose@meudominio.com.br. Em função dessa previsibilidade dos nomes, pode-se usar dicionários de nomes ou de palavras mais gerais para gerar potenciais endereços de correio eletrônico. Essa idéia pode ser explorada de modo a usar combinações aleatórias de letras e números para buscar endereços incomuns, como por exemplo abcd@meudominio.com.br. Após a geração desses endereços, as mensagens eletrônicas são enviadas e aquelas que não retornarem com erro (usuário inexistente) terão o endereço de seus destinatários validados. A busca pode ser dificultada pelo limite de tentativas mal sucedidas de descoberta de destinatários (através do comando RCPT do SMTP) configurado em muitos servidores de correio. Outra forma de realizar a verificação consiste no uso do comando VRFY do SMTP. Em função disso, atualmente diversos administradores de rede desabilitam o uso do comando VRFY. Essa técnica de força bruta é a principal responsável pelo fato de uma pessoa receber *spams* mesmo sem divulgar o seu endereço de correio eletrônico [Levine et al., 2004].

Concursos ou ofertas gratuitas de produtos também têm sido usados com o intuito de obter endereços de correio eletrônico, através de uma solicitação de dados para participação no concurso ou na oferta. Na maioria das vezes, os anúncios são falsos. Existem alguns anúncios que em letras miúdas dizem que se o usuário concordar em participar da oferta, os dados do usuário podem ser repassados para parceiros (talvez *spammers*) e o usuário pode receber ofertas comerciais. Muitos afirmam que essa idéia é utilizada por grandes empresas de software para repassar os contatos dos usuários para parceiros [Spammer-X et al., 2004].

A compra de listas prontas de endereços de correio eletrônico também é utilizada

pelos *spammers*. Algumas são negociadas diretamente entre *spammers* enquanto outras são oferecidas livremente através de *spams*. A maioria das listas contém endereços já utilizados anteriormente de forma abusiva por *spammers*. Desse modo, a maioria dos usuários pertencentes às listas já se encontra em um estado de tolerância máxima para com os *spams*, geralmente apagando-os sem ler.

Após obter os potenciais endereços dos destinatários, é necessário saber se eles são válidos. Para saber se um usuário é ativo existem diversas técnicas utilizadas pelos *spammers*. Uma delas consiste em enviar no *spam* uma referência a uma figura armazenada remotamente com um atalho contendo o endereço do destinatário, como no trecho fictício apresentado na Figura 5.9. Nesse caso, ao abrir o *spam*, a figura é automaticamente buscada do sítio, permitindo a validação do endereço do destinatário através da consulta a arquivos de registro (*log*). Por isso alguns agentes de usuário como o Thunderbird não abrem automaticamente figuras que são referenciadas em mensagens eletrônicas. Ainda assim, basta o usuário clicar no atalho para o *spammer* validar o endereço. Nesse caso, o *spammer* pode ainda montar um perfil do usuário em função do tipo de figura baixada. A opção de remoção do endereço eletrônico de uma lista também pode ser usada para validar endereços. Nesse caso, é enviado um atalho dentro do *spam* com a opção de remover o endereço da lista, conforme o trecho de um *spam* real mostrado na Figura 5.10. Assim, regras estabelecidas por legislações anti-*spam*, como o CAN-SPAM, podem ser utilizadas em benefício do próprio *spammer*. Por último, outra forma de confirmar os endereços consiste no uso do comando VRFY como citado anteriormente.

```
<img src="http://dominiospammer.com.br/figs/sexo.jpg?email=usuario@meudominio.com.br"
width=60 height=40>
```

Figura 5.9. Um trecho de um *spam* com um atalho usado para identificar o destinatário.

```
<p align=3D"center"><font face=3D"Verdana, Arial, Helvetica, sans-serif" s=
ize=3D"1">
Caso deseje remover seu nome desta lista, <u><font color=3D"#0000ff">
<a href=3D"http://dominiospammer.com.br/out/outlist.asp?email=3D=
\%25TO_EMAIL">
clique aqui</a></font></u><a href=3D"http://www.dominiospammer.com.br/sair/out/o=
utlist.asp?email=3Dusuario@meudominio.com.br">.</a></font></p>
```

Figura 5.10. Um trecho de um *spam* modificado com a opção de remoção de uma lista.

Devido à natureza da atividade de enviar *spams*, os *spammers* tentam obter além dos endereços de correio eletrônico algo que identifique que um usuário é um potencial cliente para o produto anunciado pelo *spam*. O perfil de um usuário pode ser obtido facilmente quando o endereço é coletado em listas ou através da invasão de servidores específicos como o servidor de uma loja *on-line*, de um sítio de jogos etc. [Laufer et al., 2005].

5.3.3. Formato das mensagens

As mensagens não solicitadas eram inicialmente construídas sem nenhuma preocupação com o texto ou com as palavras utilizadas. Com o avanço das técnicas anti-*spam*, a criação das mensagens eletrônicas passou a ser mais elaborada de tal forma que determinadas expressões reconhecidas pelos filtros anti-*spam* não são mais utilizadas ou são

então criados artifícios para "escondê-las" dos mecanismos anti-*spams*. Os principais formatos utilizados para o envio de *spams* são o texto simples e o texto formatado em HTML (*HyperText Markup Language*). Grande parte dos *spams* enviados atualmente ainda utiliza o texto simples. Uma das principais razões para isso, é a maior dificuldade dos filtros anti-*spam* em identificá-los. Além disso, alguns programas de correio eletrônico mais simples, que funcionam em modo texto, não aceitam outros formatos como o HTML. Contudo, os *spams* com texto simples não costumam atrair muito a atenção dos usuários, levando a uma baixa taxa de retorno.

O formato HTML é o preferido dos *spammers* atualmente. Nesse formato, o *spammer* pode colocar textos que piscam ou figuras de modo a chamar a atenção dos leitores. No entanto, o uso incorreto da linguagem HTML permite que os anti-*spams* possam identificar mais facilmente essas mensagens como *spams* [Spammer-X et al., 2004].

Diversos artifícios vêm sendo utilizados atualmente para enganar os programas anti-*spams*. Ao verificar um grande número de mensagens que continham palavras muito utilizadas nos *spams* como Viagra e Cialis, os programas anti-*spams* passaram a filtrar essas mensagens. A reação efetuada pelos *spammers* foi trocar a maneira de escrever essas palavras fazendo com que o destinatário ainda percebesse a palavra. Por exemplo, existem várias variações utilizadas para a palavra Viagra, como Vi@gr@, V-i-@-g-r-@, \/iagra etc. Conforme essas variações vão se tornando mais comuns, os mecanismos anti-*spam* passam a filtrá-las. Com o uso do HTML para compor as mensagens, torna-se mais fácil iludir os programas anti-*spam*. Uma das formas empregadas consiste no uso de caracteres invisíveis ao olho nu, por exemplo figuras de 1 x 1 pixel, ou de comentários em determinadas palavras mais comuns em *spams*. A Figura 5.11 mostra um trecho de um *spam* que usa essa técnica. Na realidade, qualquer caracter entre os sinais "<" e ">" é considerado um comentário em HTML [Costales e Flynt, 2005].

```
Having <font>prob</font><font>lems</font> maintaining a full erection or=20
one at all?<br>
<font>\/l</font><font>agra</font> works excellently for your=20
<font>pro</font><font>blem.</font>
```

Figura 5.11. Um exemplo do uso de comentários em HTML para enganar anti-*spams*.

A codificação de entidade caracter (*character-entity encoding*), na qual os caracteres são descritos por uma palavra-chave ou por um "#" seguido de um número decimal de três dígitos, também é usada para que os *spammers* disfarcem o conteúdo das páginas. A codificação URL (*Uniform Resource Locator*) também pode ser usada para os mesmos fins. A técnica *snowflaking messages* [Gregory e Simon, 2005] também é usada pelos *spammers* para enganar os programas anti-*spam*, mesmo aqueles baseados em filtros bayesianos (Seção 5.4.1.3). A técnica consiste em incluir uma grande quantidade de texto aleatório, de modo a fazer que com o programa anti-*spam* pense que o correio eletrônico seja pessoal e individual. Essa técnica é utilizada nas várias versões do *spam* apresentado na Figura 5.12. Além disso, servidores de correio reportam a listas negras em tempo real (Seção 5.4.1.1) quando recebem uma grande quantidade de mensagens iguais [Spammer-X et al., 2004]. Com a aleatoriedade, fica muito difícil identificar as várias "cópias" das mensagens. A modificação dos campos `From:` e `Received:` também

é utilizada pelos *spammers*, de modo a dificultar a obtenção de suas identidades verdadeiras e iludir os programas anti-*spams*.

```
Hi,

CIjALIS
VALjIUM
AMBjIEN
VjIAGRA

http://www.notrogores.com

in a thick fence of them all round him-that at least was the spiders
idea. Standing now in the middle of the hunting and spinning insects
Bilbo plucked up his courage and began a new song:
```

Figura 5.12. Um exemplo de um *spam* com texto aleatório para enganar anti-*spams*.

5.3.4. Envio das mensagens

Em função dos *spams* poderem conter esquemas de fraudes, estelionato, ofertas de produtos inexistentes etc., os *spammers* procuram se manter no anonimato. Os *spammers* que praticam esses atos ilícitos não enviam suas mensagens através de seus servidores, pois mesmo utilizando campos falsos de `From:` e `Received:`, um usuário com algum conhecimento do formato de mensagens eletrônicas poderia chegar ao servidor do *spammer*. Além disso, para enviar mensagens para um grande número de usuários, é necessária uma grande banda disponível. Em função disso, servidores ou máquinas alheias são utilizados para que os recursos computacionais do *spammer* não fiquem indisponíveis e também para dificultar o rastreamento do verdadeiro remetente das mensagens. Com a mesma finalidade, contas são abertas em serviços de *webmail* gratuitos. As principais formas de envio de *spams* são apresentadas a seguir.

A primeira técnica de envio de mensagens não solicitadas utilizada pelos *spammers* consiste no uso *relays* abertos de correios eletrônicos. Um *relay* aberto é um servidor de correio que está configurado para encaminhar mensagens enviadas para ele de qualquer lugar, para qualquer receptor, sem a verificação do remetente. No protocolo SMTP original, essa era a configuração padrão. De modo a solucionar esse problema, o sendmail e outros servidores de correio eletrônico passaram a não permitir o *relay* de fora da rede. Contudo, às vezes surgem algumas falhas de segurança no sendmail que permitem o uso do *relay* [Spammer-X et al., 2004].

Por ter sido amplamente utilizada, a exploração dos servidores com *relay* aberto se tornou conhecida pelos administradores de rede que passaram a corrigir as falhas na configuração desses servidores. Além disso, atualmente, o envio de *spams* usando os *relays* abertos diminuiu. Esta técnica se tornou pouco utilizada pelos *spammers*, pois testes são proativamente realizados para verificar se servidores de correio eletrônico estão agindo como *relays* abertos e se for esse o caso, colocam o servidor em uma lista negra (Seção 5.4.1.1).

Em virtude da redução do número de servidores com *relay* aberto e do uso de listas negras, os *spammers* passaram a usar outras técnicas de envio como os servidores *proxy* abertos. Um servidor *proxy* é um servidor usado dentro de uma rede que outros compu-

tadores utilizam como um *gateway* para a Internet. Os *proxies* mais comuns usados na Internet são o Wingate e o Squid. Esses *proxies* utilizam protocolos como o Socks v4 e v5. Um servidor *socks* padrão permite que qualquer "cliente" utilize o servidor para encaminhar mensagens eletrônicas para servidores SMTP. Um dos problemas do uso de um servidor *proxy* vem do fato que assim que algum servidor de correio eletrônico detecta um número grande de mensagens vindas de um mesmo servidor (*proxy*), ele pode rejeitar as mensagens e reportar o endereço IP do servidor *proxy* para uma lista negra. Em função disso, os *spammers* procuram utilizar um grande número de servidores *proxy* que ainda não estejam em alguma lista negra. Outro artifício utilizado envolve servidores em países com línguas diferentes para dificultar a comunicação entre o servidor SMTP atacado e o responsável pelo *proxy* utilizado. Como a maioria dos servidores *proxy* pertencem a usuários de modem a cabo ou de linhas de assinantes digitais (*Digital Subscriber Lines* - DSLs), recentemente um grande provedor americano de acesso via cabo começou a bloquear todo o tráfego de saída na porta 25 (SMTP) para reduzir a quantidade de máquinas inseguras que enviam *spams* [Spammer-X et al., 2004]. Outro problema em usar os servidores *proxy* tem origem no compartilhamento dos servidores entre vários *spammers*, o que acelera a detecção e a inclusão em listas negras.

Computadores também podem ser invadidos de modo a serem utilizados para o envio de *spams*. Dependendo do número de computadores comprometidos, os *spammers* podem formar redes de máquinas zumbis (*zombie networks*), também chamadas redes de robôs (*botnets*). Essas máquinas estão sob total controle dos *spammers* e são usadas para enviar suas mensagens. As redes zumbis são coordenadas por máquinas mestras que também são controladas pelos *spammers*. A função das máquinas mestras é disparar o envio de *spams*. O grande problema de usar uma rede de máquinas zumbis para enviar *spams* é o risco de uma punição severa. Nessa situação, os *spammers* estão se apropriando de bens de terceiros para utilizá-los em benefício próprio, o que pode levá-los à prisão. Apesar dos riscos, a maioria dos *spams* é enviado dessa forma. Outro método popular de envio de *spams* utiliza o seqüestro (*hijacking*) de interfaces CGI (*Common Gateway Interface*). *Scripts* de CGIs são modificados através da adição e controle de variáveis de configuração ou de campos de entrada pelo usuário. Redes sem fio também podem ser invadidas com o propósito do envio de *spams*. Por último, o protocolo de roteamento BGP (*Border Gateway Protocol*) também pode ser usado para o envio das mensagens. Essa técnica é conhecida como injeção de rota BGP ou seqüestro de sistema autônomo. O *spammer* seqüestra faixas de endereços IP válidos que não estejam sendo utilizados e invade um roteador com falhas de segurança para ser responsável por essa faixa de endereços. Depois disso, o *spammer* manipula os pedidos de atualização de rotas e anuncia aos roteadores vizinhos que o roteador invadido é a única rota para a faixa de endereços IP adquirida maliciosamente. A partir daí, o *spammer* pode começar a enviar as mensagens usando um endereço IP da faixa. Quando um endereço IP é adicionado a uma lista negra, o *spammer* passa a usar outro endereço da faixa. Essa técnica necessita um bom conhecimento de roteadores e de protocolos de roteamento, porém é uma das mais efetivas no envio de *spams*. O envio dessa forma é muito difícil de ser rastreado e bloqueado.

Contas em serviços gratuitos de correio eletrônico via *web* são utilizadas para enviar *spams* ou receber respostas de usuários interessados nas ofertas [Wiki-Spam, 2006]. Em função da grande quantidade de mensagens enviadas pelos *spammers*, várias contas

são criadas por robôs. De modo a tentar evitar a ploriferação dessas contas, vários serviços de correio eletrônico via *web* passaram a exigir a identificação de uma palavra através de um gráfico sobre um fundo difícil de ser lido (Seção 5.4.3.3). Isso não impede que seres humanos identifiquem a palavra, mas torna difícil a leitura para os robôs. Com isso os *spammers* resolveram usar seres humanos para fazer essas leituras. Uma das maneiras de fazer isso é enviar mensagens a usuários pedindo que eles entrem com a palavra do gráfico e disponibilizando acesso a material pornográfico supostamente exclusivo.

Por último, existem também empresas especializadas no envio de *spams*, voltadas principalmente para os *spammers* que não possuem conhecimento técnico suficiente para enviar as mensagens. Essas empresas podem usar servidores hospedados em países que não proíbem o envio de *spams*.

5.4. Técnicas de combate aos *spams*

O objetivo de um sistema anti-*spam* é reduzir o número de *spams* recebidos por um usuário, classificando as mensagens para, então, filtrá-las. Esses sistemas estão em constante evolução já que para cada novo sistema, tenta-se criar técnicas para enganá-lo e permitir a passagem dos *spams*. As principais propriedades de um sistema anti-*spam* são a sua taxa de falsos positivos e de falsos negativos, ou seja, a taxa de mensagens legítimas classificadas como *spams* e vice-versa. Em geral, a taxa de falsos positivos tem um valor mais importante, já que uma mensagem legítima acaba sendo filtrada, o que pode gerar grandes transtornos e atrasos no processo de comunicação. Já os falsos negativos têm um impacto menor, já que o usuário irá receber o *spam*, mas provavelmente acabará apagando-o. Outro aspecto importante de um sistema anti-*spam* é a sua interferência com o usuário, seja ela por necessidade de configuração, manutenção, atualização ou desafios que são feitos ao usuário e que devem ser respondidos. Quanto maior o nível de interação com o usuário, mais complexo e menos amigável o sistema acaba se tornando, dificultando sua adoção em grande escala. Nesta seção, os sistemas anti-*spam* atuais e novos mecanismos propostos na literatura são classificados e analisados.

5.4.1. Sistemas baseados em filtragem simples

Nesses sistemas, novos dados ou regras são inseridos de forma manual. A principal crítica a sistemas dessa natureza é a baixa eficiência, uma vez que tais sistemas dependem de constante atualização manual, tornando o processo custoso para o usuário. O alto grau de evolução das técnicas de envio de *spam* também irão fazer com que esses sistemas precisem de uma grande taxa de atualização, o que aumenta ainda mais o trabalho do usuário. Alguns exemplos desse tipo de sistema são apresentados nesta seção.

5.4.1.1. Listas negras

Os primeiros sistemas anti-*spam* a surgirem baseavam-se na utilização de listas negras, que são listas com endereços de origem ou endereços IP de remetentes que reconhecidamente são fontes de *spam*. Já nas listas brancas, são colocados endereços de pessoas ou servidores confiáveis. Dessa forma, quando uma mensagem é recebida as duas listas são analisadas. Se a presença do endereço de origem for detectada na lista

negra, a mensagem é diretamente classificada como *spam*, sem sequer ser analisada por quaisquer outros mecanismos que porventura estejam sendo utilizados. Por outro lado, caso o endereço esteja na lista branca, a mensagem é aceita diretamente. Caso o endereço não esteja em nenhuma das duas listas, a mensagem pode ser aceita ou então são utilizados outros mecanismos anti-*spam* que estejam disponíveis, para tentar classificar a mensagem.

Inicialmente, essas listas eram feitas pelos próprios usuários e cada um ficava responsável por adicionar e remover os endereços das duas listas. Isso demandava um grande esforço do usuário, pois ele tinha que separar as mensagens e determinar se o endereço deveria ser colocado na lista branca na negra ou, em caso de dúvidas, em nenhuma das duas listas. A evolução natural foi tornar o sistema centralizado, onde entidades centrais controlam a entrada e saída de endereços das listas. A primeira implementação desse tipo de sistema distribuído foi chamada de *Real-time Blackhole List* (RBL) e foi criada por Paul Vixie. Na RBL eram listados os endereços IP de servidores utilizados para enviar *spam*. Contudo, esses endereços eram adicionados manualmente. Em seguida a RBL, surgiu outra proposta chamada ORBS (*Open Relay Behavior-modification System*) cuja principal diferença para a RBL é a realização automática de testes para identificar servidores que permitem o envio de mensagens sem nenhum controle. Os servidores com essa característica são adicionados automaticamente na lista negra, bloqueando as mensagens originadas por eles. O processo de remoção dessa lista, no entanto, é realizado de forma manual, através do contato do administrador do servidor listado devia com a entidade responsável pela manutenção da lista.

A consulta a estas listas distribuídas é geralmente feita através do protocolo DNS (*Domain Name System*), fazendo com que elas sejam chamadas genericamente de DNSBLs (*Domain Name Sysytem Black Lists*). Para realizar a consulta a essas listas, o servidor verifica se o endereço IP do cliente ou de outro servidor que se conectou a ele está presente na lista DNSBL anteriormente configurada. A verificação utilizando o protocolo DNS é realizada fazendo uma consulta DNS ao endereço formado invertendo-se os bytes do endereço IP do cliente e adicionando o nome do domínio da entidade responsável pela DNSBL. Um exemplo desse processo é mostrado na Figura 5.13, onde o servidor A quer enviar uma mensagem eletrônica para o Servidor B. Nesse caso, o Servidor B verifica se o Servidor A está presente ou não na lista negra. Essa verificação é realizada através do envio de um pedido DNS para um servidor DNS que fará a consulta à lista DNSBL. Caso a resposta do pedido de DNS seja positiva, isso significa que o endereço testado está presente na lista negra. Essas listas são também um alvo freqüente de ataques de negação de serviço por parte de *spammers* para tentar neutralizar esse mecanismo de proteção anti-*spam*.

As listas negras podem ser efetivas no bloqueio de servidores utilizados por *spammers*. Estudos mostram que até 80% dos *spams* podem ser evitados por meio do uso desses mecanismos [Jung e Sit, 2004]. Essas listas, no entanto, sofrem o problema de falsos positivos, quando um endereço é incorretamente adicionado à lista. O processo de retirada da lista pode demorar, fazendo com que mensagens legítimas acabem sendo perdidas. Máquinas contaminadas por vírus também podem ser afetadas por esse sistema. O vírus pode aproveitar-se dos recursos da máquina e instalar um servidor de correio eletrônico para ser usado por *spammers*. Essas máquinas podem acabar nas listas negras, fazendo

com que as mensagens do usuário da máquina contaminada acabem sendo recusadas.

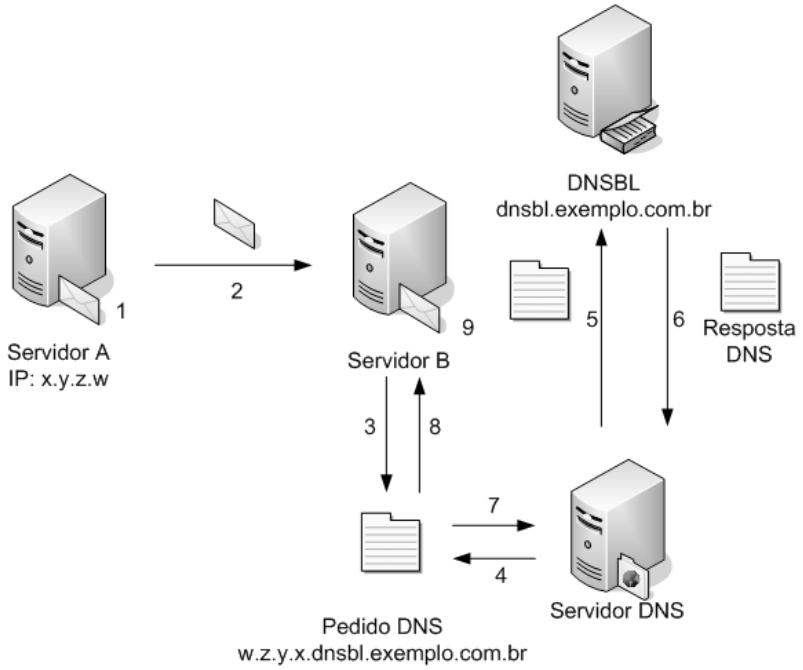


Figura 5.13. Verificação de listas negras.

5.4.1.2. Uso de pesos e regras

Nesses sistemas são utilizadas regras lógicas para a classificação das mensagens como *spam* ou não. Cada uma das regras define um teste que deve ser realizado na mensagem. Caso o resultado de um teste seja positivo, esse resultado é usado de forma ponderada para determinar a probabilidade da mensagem ser ou não *spam*. Os pesos de cada um dos testes podem ser positivos ou negativos, indicando uma maior ou uma menor probabilidade da mensagem ser *spam*. Quando uma mensagem é recebida, o mecanismo realiza todos os testes previamente definidos, somando os pesos de todos os testes cujo resultado foi positivo. Com base no valor final da soma de todos os pesos, são tomadas ações, podendo a mensagem ser encaminhada para o destinatário, marcada como sendo provavelmente um *spam* ou então descartada, o que geralmente acontece se o valor da soma de todos os pesos for muito alta.

Uma das principais e mais delicadas etapas desse mecanismo é a construção de regras, que ao mesmo tempo tem que balancear a generalidade para a adaptação a novos tipos de *spam* e a especificidade, para não classificar mensagens legítimas como *spam*. O processo de criação de regras é geralmente feito extraindo-se frases ou palavras características de mensagens de *spam*. Em seguida é feita uma análise utilizando uma grande base de mensagens manualmente classificadas como *spam* ou não, para avaliar a taxa de falsos positivos e negativos. Se alguma das taxas for alta, a regra é refinada, até que as duas taxas sejam baixas o suficiente. A determinação do peso da regra geralmente é baseada na percentagem das mensagens de *spam* onde a regra apresentou um resultado

positivo, sendo maior quanto maior for essa percentagem. Um dos principais sistemas anti-*spam* baseados nessa proposta é o *SpamAssassin* [Apache, 2006] que atualmente é um dos mecanismos mais utilizados. O *SpamAssassin*, no entanto, também utiliza outros métodos como listas negras, filtros bayesianos, verificação do DNS reverso e verificação SPF, criando regras especiais que representam esses outros tipos de teste e não são regras estáticas.

As soluções baseadas apenas em regras estáticas, no entanto, tem o inconveniente da construção das regras, tornando o processo muito complexo para usuários não experientes. Além do problema da construção das regras, elas também devem ser atualizadas freqüentemente, já que os *spams* estão em constante evolução. Esse problema da construção manual de regras levou ao surgimento de sistemas adaptativos, que são capazes de se adaptar a mudanças nas características dos *spams* ao longo do tempo como serão mostrados na Seção 5.4.2. Como a maioria dos usuários não tem conhecimento e/ou tempo suficiente para criar suas próprias regras, geralmente um grupo de pessoas fica responsável por criar as novas regras. Esse processo, no entanto pode causar problemas em situações específicas, principalmente para pessoas que trabalham com assuntos que geralmente estão presentes em mensagens de *spam*, fazendo com que as mensagens legítimas acabem sendo classificadas como *spam*. Analisando as regras [Project, 2006] utilizadas pelo sistema anti-*spam* SpamAssassin, se uma mensagem contiver nomes de certos medicamentos, a probabilidade da mensagem ser classificada como *spam* já será bem alta, o que pode causar grandes transtornos para pessoas que atuam no segmento de farmácias *on-line*.

5.4.1.3. Filtros bayesianos

Os sistemas que utilizam filtros bayesianos funcionam com base em métodos estatísticos que levam em conta a freqüência de ocorrência de determinadas palavras ou frases em mensagens classificadas ou não como *spam*. Um usuário de correio eletrônico quando recebe uma nova mensagem *spam* faz a sua leitura e acaba identificando e memorizando algumas palavras que fizeram com que ele decidisse que aquela mensagem é um *spam*. Da próxima vez que este usuário ler e identificar essas mesmas palavras em outra mensagem já terá uma suspeita maior que a nova mensagem recebida também seja uma mensagem *spam*. Dessa forma, algumas palavras-chave acabam se constituindo em características dos *spams*. Os filtros bayesianos têm por finalidade repetir este mesmo procedimento humano de identificação de *spam* realizado pelo usuário, só que de forma automatizada. Para a identificação automatizada de *spam* um filtro bayesiano deve ser construído. O primeiro passo é um processo de aprendizagem, onde o usuário identifica manualmente mensagens legítimas e mensagens *spam*, para a construção de um filtro que permita classificar mensagens futuras como legítimas ou *spams*. Uma característica importante desses filtros é que como eles são geralmente feitos com base nas informações de identificação de mensagens *spams* passadas pelo usuário, eles se adaptam ao padrão de *spams* e de mensagens legítimas recebidas pelo próprio usuário.

Para construir os classificadores usados para filtrar os *spam*, são geralmente utilizadas redes bayesianas. Uma rede bayesiana é um grafo acíclico, direcionado que re-

presenta uma distribuição de probabilidade [Pearl, 1988]. Nesse grafo, cada variável aleatória X_i é representada por um nó. Uma aresta entre dois nós indica a probabilidade de influência do nó pai para o nó filho. Além disso, cada nó X_i da rede é associado a uma tabela de probabilidade condicional, que determina a distribuição de X_i dados os valores dos seus pais. Um classificador bayesiano utiliza uma rede bayesiana onde existe um nó C representando uma das possíveis classes c_k que representam as possíveis classificações que o filtro pode realizar e vários filhos X_i, para cada uma das características testadas. Dessa forma, dado um certo conjunto de valores de X_i, pode-se calcular a probabilidade de cada classe c_k de acordo com a Equação 1, onde o termo $P(X = x|C = c_k)$ é dado pela Equação 2. Os filtros bayesianos podem permitir a dependência entre as características X_i ou não, nesse caso tornando a Equação 2 mais fácil de ser resolvida.

$$P(C = c_k|X = x) = \frac{P(X = x|C = c_k)P(C = c_k)}{P(X = x)} \tag{1}$$

$$P(X = x|C = c_k) = \prod_i P(X_i = x_i|C = c_k) \tag{2}$$

Para a utilização dos filtros bayesianos na classificação de mensagens, é necessário que, em primeiro lugar, as mensagens sejam representadas como vetores de características X_i a serem testadas, chamadas de símbolos. Esses vetores são construídos com base na separação das palavras da mensagem em vários símbolos. Além de considerar as palavras para a classificação das mensagens, podem ser utilizadas outras características que geralmente estão presentes em *spams*, como o uso exagerado de exclamações e outras informações características, como o uso de *tags* html. Após a separação da mensagem em vários símbolos, cada um deles é comparado com sua freqüência de ocorrência em mensagens *spams* e não *spams* anteriores. Se o símbolo apareceu predominantemente em mensagens *spams*, é atribuída uma alta probabilidade dele representar uma característica de *spam*. A probabilidade do símbolo é geralmente calculada como sendo a proporção entre o percentual de aparição do símbolo em mensagens *spams* e legítimas. Para economizar espaço no filtro, geralmente somente os símbolos com probabilidade muito alta e muito baixa são armazenados, já que eles podem determinar as principais características. O processo de divisão da mensagem em vários símbolos é de extrema importância no processo de classificação. Os primeiros mecanismos separavam as palavras utilizando-se sinais de pontuação, fazendo com que *spammers* explorassem essa vulnerabilidade, criando frases como *C/A/L/L/ N-O-W - I/T/S F_R_E_E*, onde as letras são separadas por caracteres que geralmente são utilizados para separar frases ou palavras. Se for utilizado um método convencional de separação em palavras, a frase acima produzirá apenas símbolos que são compostos de uma letra, dificultando a determinação se o símbolo corresponde a uma característica de *spam* ou não. Atualmente, no processo de separação de símbolos das mensagens são utilizados processos mais complexos, como a junção de símbolos formados por apenas uma letra, a agregação de símbolos, e a tentativa de reduzir novos símbolos a símbolos que tenham uma similaridade com os já conhecidos, com o objetivo de reduzir o número de símbolos diferentes que são armazenados e tratar da mesma forma símbolos que tenham grande similaridade.

Outro mecanismo que pode ser utilizado juntamente com os filtros bayesianos é

a redução bayesiana de ruído (*Bayesian Noise Reduction* - BNR) [Zdziarski, 2004]. Essa técnica tem como objetivo remover palavras fora do contexto, que muitas vezes são utilizadas por *spammers* para tentarem enganar os filtros bayesianos, conforme foi mostrado na Seção 5.3.3. Para analisar o contexto, é verificada inicialmente a probabilidade de cada símbolo encontrado na mensagem. Em seguida, é utilizada uma janela deslizante que agrupa as probabilidades de cada um dos símbolos serem característicos de *spam*. Cada um desses agrupamentos é chamado de contexto artificial. O contexto artificial é então analisado utilizando um outro filtro bayesiano, fornecendo uma probabilidade do contexto artificial criado estar presente em mensagens *spam*. Se o contexto for identificado como sendo característico de *spam*, os símbolos que fazem parte do contexto e tem uma baixa probabilidade são removidos, por se tratarem provavelmente de palavras aleatórias que não são freqüentemente encontradas em *spams* e que foram adicionadas para tentar enganar o filtro.

5.4.2. Sistemas com auto-aprendizado

Estes sistemas são capazes de aprenderem sozinhos e de aumentar sua eficiência no combate aos *spams*. Um sistema com auto-aprendizado possui como principal característica a falta de necessidade de intervenção do usuário, pois o próprio sistema aprende com o seu passado e presente, para se adaptar as mudanças do futuro. Em função de não necessitar de nenhuma intervenção do usuário, geralmente suas taxas de falsos positivos e falsos negativos podem ser maiores. Atualmente, a maior parte das pesquisas na área se concentra nesse tipo de sistema. Alguns exemplos desse tipo de sistema são apresentados a seguir.

5.4.2.1. Caracterização de tráfego de *spams*

Um dos mecanismos mais básicos de caracterização do tráfego de *spams* é monitorar a quantidade de mensagens enviadas pelos usuários. Se um determinado usuário estiver enviando uma quantidade muito grande de mensagens, provavelmente trata-se de um *spammer* e medidas podem ser tomadas, como proibir o envio de novas mensagens. Esse mecanismo, no entanto, não impede que o *spammer* utilize seu próprio servidor que não irá impor nenhum limite ao envio de mensagens. A verificação do tráfego excessivo de mensagens entre servidores é difícil de ser realizada, pois mensagens legítimas podem gerar um tráfego muito grande.

O tráfego de *spam* tem várias características que o tornam diferente do tráfego gerado por mensagens legítimas, já que os *spams* são geralmente enviados por um mecanismo automatizado que visa enviar o máximo de mensagens possíveis para o máximo de destinatários. Essa característica torna possível a identificação de padrões e anomalias no tráfego de mensagens, que podem ser utilizados para caracterizar uma mensagem ou um fluxo de mensagens como *spam*. Mecanismos baseados nesse sistema ainda não estão disponíveis comercialmente. Por enquanto essa é apenas uma área de pesquisa que busca encontrar e entender as diferenças entre os *spams* e mensagens legítimas.

Uma das características de um *spam* corresponde a sua distribuição constante ao longo dos dias de semana e horas do dia [Gomes et al., 2004]. Essa distribuição já não

ocorre para as mensagens legítimas, cujos envios se concentram nos horários entre a manhã e a tarde ao longo do dia e durante os dias da semana, excluindo-se o final de semana. Essa característica tem origem no próprio comportamento humano, já que a grande parte da população trabalha e troca mensagens mais freqüentemente durante esse período. Já para os *spammers*, quanto mais *spams* enviados, maiores serão seus lucros, fazendo com que essas mensagens sejam enviadas sempre que possível, incluindo os períodos da madrugada e os finais de semana, em que a grande maioria das pessoas não trabalha.

Outra característica temporal do tráfego de *spam* é o intervalo entre as mensagens enviadas, que geralmente é muito curto, novamente devido à forma pela qual são enviados. Para mensagens legítimas, esse tempo geralmente é maior, já que é necessário ler a mensagem, tomar alguma ação, caso necessário, e escrever uma resposta. Todo esse processo dura, em geral, um tempo maior do que o tempo de envio de dois *spams*.

A análise do número e da freqüência dos remetentes e dos destinatários das mensagens também pode fornecer outra característica das mensagens de *spam*. Um remetente que envie mensagens para um número muito grande de destinatários é candidato a ser classificado como *spammer*, já que esse comportamento não é geralmente encontrado entre usuários legítimos. Esse caso geralmente ocorre quando um determinado endereço está em uma lista utilizada por *spammers*, fazendo com que um mesmo usuário receba uma quantidade de mensagens muito grande de diferentes origens. Outra situação onde esse processo pode ocorrer é quando *spammers* utilizam ataques de dicionário, onde pessoas com nomes comuns acabam recebendo *spams* de vários *spammers* que utilizem esse método.

O tamanho das mensagens também é outro aspecto que pode ser utilizado para caracterização de *spam*, já que estudos [Gomes et al., 2004] mostram que apenas 1% das mensagens de *spam* tem mais do que 60Kb. O principal motivo para que os *spammers* utilizem mensagens pequenas e não enviem anexos nas mensagens é que quanto menor for a mensagem, menos banda ela irá utilizar, permitindo então que mais mensagens sejam enviadas em menos tempo, um dos principais objetivos dos *spammers*.

As características do comportamento humano se refletem na característica do tráfego gerado pelas mensagens legítimas, diferenciando-o do tráfego gerado por *spams*. As características do tráfego, no entanto, são muito pessoais e nem sempre podem ser generalizadas. Embora o comportamento humano na média possibilite identificar as anomalias em tráfegos de *spams*, diferenças em relação ao padrão podem ser resultados de casos especiais. Um exemplo dessa situação é uma pessoa que trabalha com grupos de pessoas em diferentes fusos horários, fazendo com que a distribuição temporal não siga o comportamento da média, tendo picos em horários não convencionais, como as madrugadas. Essas limitações poderiam ser contornadas com o aprendizado específico para cada usuário, ao invés de serem utilizadas médias de comportamento de tráfego.

5.4.2.2. Listas cinzas

A proposta dos sistemas baseados em listas cinzas é criar uma lista intermediária, entre as listas brancas e as listas negras, apresentadas na Seção 5.4.1.1. Sua eficácia baseia-se no fato de que *spams* geralmente não são retransmitidos quando o servidor noti-

fica algum erro no envio das mensagens. Esse comportamento visa aumentar a capacidade dos *spammers* enviarem mensagens com sucesso, uma vez que se um determinado endereço reportou um erro quando a mensagem foi enviada, um *spammer* não tentará enviar a mensagem uma próxima vez, para não perder tempo. Endereços para os quais o *spam* não foi entregue na primeira vez podem ser excluídos das listas de *spam*, assumindo que esses endereços não existem mais ou são inválidos.

O mecanismo de listas cinzas trabalha com duas listas, uma lista branca e uma lista cinza. Quando uma mensagem é recebida, o sistema consulta se o par destinatário/origem encontra-se na lista branca. Caso positivo, a mensagem é encaminhada para o destinatário sem restrições. Por outro lado, quando o par destinatário/origem não se encontra na lista branca, o servidor reporta um erro para o remetente da mensagem informando que houve um problema temporário no envio da mensagem. Servidores de mensagem legítimos devem seguir a RFC821 [Postel, 1982] e tentar reenviar a mensagem após um tempo, recomendado pela norma de quatro horas. Esse par destinatário/origem que teve a mensagem recusada é inserido na lista cinza juntamente com a informação temporal de quando foi adicionado. Quando o servidor do remetente tenta realizar novamente a entrega da mensagem, o servidor do destinatário irá verificar que o par destinatário/origem já foi adicionado à lista cinza. Se o par estiver na lista cinza por um período maior que um determinado valor configurado, denominado "quarentena", a mensagem é aceita e encaminhada para o destinatário. Nesse caso, a entrada que estava na lista cinza é retirada e adicionada à lista branca, fazendo com que futuras comunicações entre esse destino e origem não precisem passar novamente pelo processo de inclusão e exclusão da lista cinza. No caso da mensagem retransmitida ser recebida antes de se esgotar o período de quarentena especificado, a mensagem é recusada novamente. O período de quarentena é adotado para não permitir que um servidor retransmita logo em seguida à recepção de uma mensagem de erro temporário causado pelo processo de listas cinzas e a mensagem acabe sendo aceita. Isto dificulta e aumenta os custos de envios automáticos de mensagens.

A origem é identificada não apenas pelo endereço do remetente da mensagem, mas também pelo endereço do servidor que enviou a mensagem. O objetivo desse mecanismo é amenizar o problema da falsificação de endereços do remetente nos *spams*. Se a informação do servidor de origem não fosse guardada junto com o endereço eletrônico, um *spammer* poderia enviar *spams* de outros servidores se passando por um usuário que já se encontra na lista branca, não precisando passar pelo processo da lista cinza. Entretanto, essa característica do mecanismo, entretanto, pode causar problemas para servidores de correio eletrônico que utilizam vários servidores para enviar as mensagens. Quando a mensagem for enviada para o servidor de destino pela primeira vez, ela será enviada por algum dos servidores de correio que estejam disponíveis. Supondo o caso do servidor do destinatário não ter ainda recebido nenhuma mensagem desse remetente e servidor, sua informação será armazenada na lista cinza e a mensagem não será aceita. O servidor do remetente irá receber o erro e esperar um determinado tempo para reenviar a mensagem. Quando a mensagem for reenviada, um servidor diferente pode ser utilizado, fazendo com que o servidor do destinatário coloque essa nova combinação de origem/servidor novamente na lista cinza, atrasando ainda mais o envio da mensagem. Quanto maior o número de servidores de correio eletrônico que o provedor dispuser, mais grave esse problema se tornará, já que a cada vez a mensagem pode ser enviada por um servidor diferente.

Esse caso pode chegar até o extremo da mensagem não ser entregue devido a um número excessivo de tentativas de retransmissão.

Esse mecanismo tem a vantagem de necessitar de poucos recursos computacionais para ser executado, já que não é necessário realizar nenhum procedimento complexo. Além disso, a mensagem é recusada na primeira vez antes de ser recebida, fazendo com que os custos com armazenamento e banda passante acabem sendo menores, uma vez que se a mensagem não for retransmitida, trata-se provavelmente de um caso de *spam* e nesse caso a mensagem não é recebida ou armazenada.

A principal desvantagem desse método consiste no atraso de mensagens legítimas. Isso pode causar descontentamento dos usuários em função do tempo muito grande de espera para receber mensagens de certos remetentes, principalmente daqueles que usem provedores maiores que geralmente contam com um grande número de servidores de envio de correio eletrônico. Servidores mal configurados que tentam retransmitir mensagens muito rápido ou um pequeno número de vezes, podem gerar falsos positivos, fazendo com que mensagens legítimas não cheguem aos destinatários.

5.4.2.3. Potes de mel

A proposta de utilizar potes de mel (*honeypots*) não tem como objetivo principal atuar diretamente na filtragem de *spam*, mas servir como um mecanismo que auxilie o aprendizado sobre o comportamento de *spammers* e também para detectar padrões utilizados por *spams*. A partir deste aprendizado elaborar técnicas de combate aos *spams*.

Os potes de mel são compostos por estruturas que têm como principal finalidade enganar de alguma forma os *spammers* [Andreolini et al., 2005]. Os três principais componentes de um pote de mel para *spams* são:

- Servidores de Internet Falsos - Esse componente simula um servidor de páginas de internet falso, gerando páginas aleatórias com um grande conjunto de endereços eletrônicos criados aleatoriamente, mas cujo domínio pertence ao mesmo domínio do pote de mel. Esse procedimento faz com que robôs que automaticamente vasculham páginas da internet à procura de endereços de correio eletrônico adicionem uma grande quantidade de endereços à sua lista de envio de *spams*. Alguns destes endereços podem ser usados como recipientes de *spams* e, desta forma, aprender as características dos *spams*. O objetivo de se colocar uma enorme quantidade de endereços falsos é de diminuir a eficácia da ação do *spammer* uma vez que se diminui taxa de sucesso do *spam*.chegar a usuários legítimos. A construção das páginas geralmente é feita com base em vários textos aleatórios, de forma que torne mais difícil a detecção de um pote de mel. Uma técnica possível é criar textos com os endereços falsos com a mesma cor do fundo da página, fazendo com que uma pessoa que visite a página não visualize esses endereços, mas que os robôs os adicionem às suas listas;

- Servidores de Envio de Mensagens - Para enganar os *spammers* são criados servidores de *email* que aparentemente podem enviar mensagens sem nenhuma restrição,

passando-se por servidores mal configurados ou, por exemplo, máquinas infectadas por vírus que passam a ser servidores de correio eletrônico. Essas máquinas acabam sendo encontradas por *spammers* que fazem ataques de varredura à procura desse tipo de servidor. Assim que uma máquina desse tipo seja descoberta por *spammers*, ela pode servir para envio de *spams*. Este servidor é configurado para reportar ao remetente que a mensagem foi enviada com sucesso, quando na verdade as mensagens não são transmitidas. Dessa forma, esses servidores podem obter várias informações sobre os *spammers* como, por exemplo: a sua origem, os tipos de mensagens enviadas e características do tráfego gerado;

- Servidores para Receber Mensagens - Esses servidores são responsáveis por receber as mensagens que foram enviadas para os endereços divulgados com o intuito de serem adicionados às listas de *spam*. Já que esses endereços não são utilizados para nenhum outro propósito, todas as mensagens recebidas são *spams*, que podem ser utilizadas nos mecanismos de aprendizado de outros sistemas anti-*spam*.

Os potes de mel podem ser compostos de um ou mais componentes apresentados anteriormente. Podem ser utilizadas várias máquinas virtuais, representando cada um dos servidores, mas todas sendo executadas na mesma máquina física. É importante que se tome o cuidado de separar o pote mel do resto da rede para que o comprometimento do pote de mel não atinja toda a rede. Uma solução para esse problema consiste em utilizar uma arquitetura na qual os servidores do pote de mel estão localizados em uma zona desmilitarizada (DMZ) e separados dos servidores públicos por um roteador e um *firewall* (Figura 5.14) responsável por limitar o acesso dos servidores do pote de mel à Internet, não permitindo, por exemplo, que as mensagens enviadas por *spammers* através dos falsos servidores sejam realmente enviadas para o destino final.

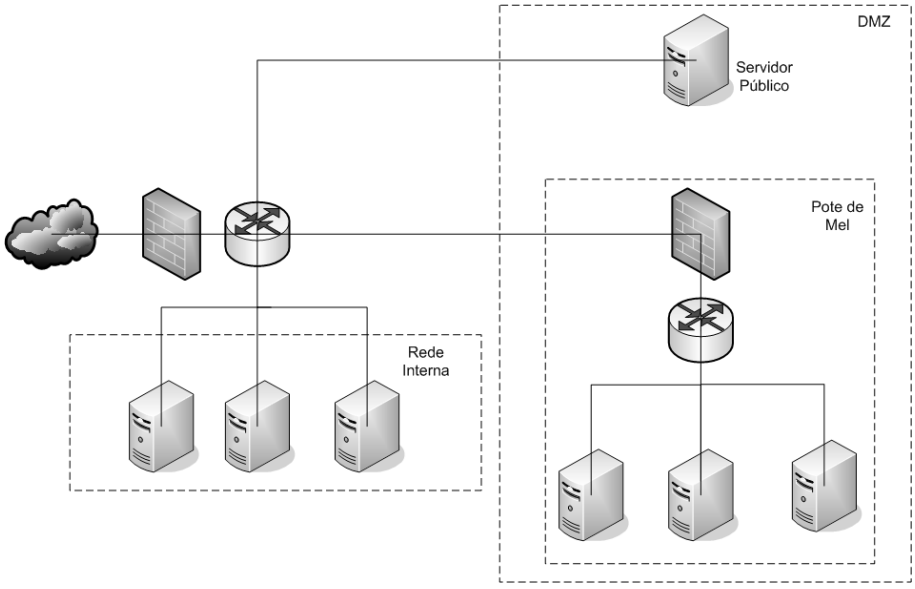
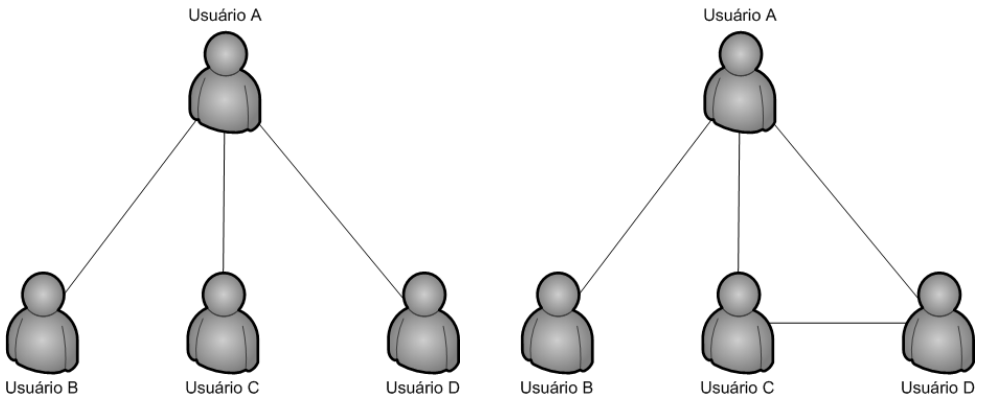


Figura 5.14. A estrutura de um pote de mel.

5.4.2.4. Padrões sociais

Uma característica importante que pode ser utilizada para a classificação das mensagens como *spam* é o uso de padrões de redes sociais. Devido ao instinto humano de formação de grupos na sociedade, a análise da rede social de determinado indivíduo pode auxiliar no processo de classificação de *spams*. Em grupos em que as pessoas se conhecem, a probabilidade de uma mensagem enviada por um indivíduo desse grupo para outro indivíduo do grupo ser um *spam* é baixa. Portanto, o objetivo desse mecanismo é determinar quais são as redes sociais formadas pelos usuários, classificando como *spam* as mensagens que não pertencem a remetentes participantes da rede social.

Uma forma de estabelecer uma rede social é construir um grafo onde cada nó representa um endereço de correio eletrônico e cada aresta representa a ocorrência de uma comunicação prévia entre os dois usuários correspondentes aos nós na extremidade da aresta. Na Figura 5.15(a) é mostrado o grafo que se forma ao se considerar uma mensagem que foi enviada do usuário A para os usuários B, C e D. Considerando que C envia uma mensagem para D, é adicionada mais uma aresta ao grafo ligando os nós C e D 5.15(b). Repetindo o processo de analisar os remetentes e destinatários das mensagens e atualizando o grafo, é criada uma representação das relações entre os diversos usuários. A construção da rede de relacionamentos pode ser feita apenas levando-se em consideração as mensagens recebidas por um usuário [Boykin e Roychowdhury, 2005] ou então utilizando as mensagens recebidas por todos os usuários do domínio [Gomes et al., 2006].



(a) Rede inicial. (b) Rede final.

Figura 5.15. Um exemplo de rede social.

Depois da rede de relacionamentos ter sido construída, ela é dividida em várias componentes, que são subgrafos do grafo completo que não apresentam arestas entre si. Porém, antes dessa divisão em componentes, o nó do qual se deseja analisar as relações é removido do grafo, já que ele acabaria ligando todas as componentes e não poderia ser feita a separação. Cada componente é então analisada para descobrir se ela corresponde a um relacionamento social ou corresponde a uma componente formada pelo processo de envio de mensagens de *spam*. A principal característica usada para classificar as compo-

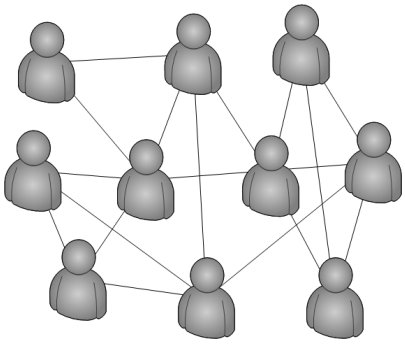
nentes nessas duas classes é o grau de agrupamento [Boykin e Roychowdhury, 2005], que é definido pela Equação 3, onde N_2 é o número de nós com mais de dois vizinhos, k_i é o número de vizinhos do nó i e E_i é o número de arestas que existem entre os k_i vizinhos do nó i. O significado do coeficiente de agrupamento é medir o nível de ligações entre usuários diferentes. Se cada nó do grafo possui uma ligação com cada um dos outros nós, pode-se observar pela Equação 3 que o grau de agrupamento é um. Já no caso de não existirem interligações entre usuários, o termo E_i será zero, fazendo com que o coeficiente de agrupamento também seja zero.

$$C = \frac{1}{N_2} \sum_i \frac{2E_i}{k_i(k_i - 1)} \tag{3}$$

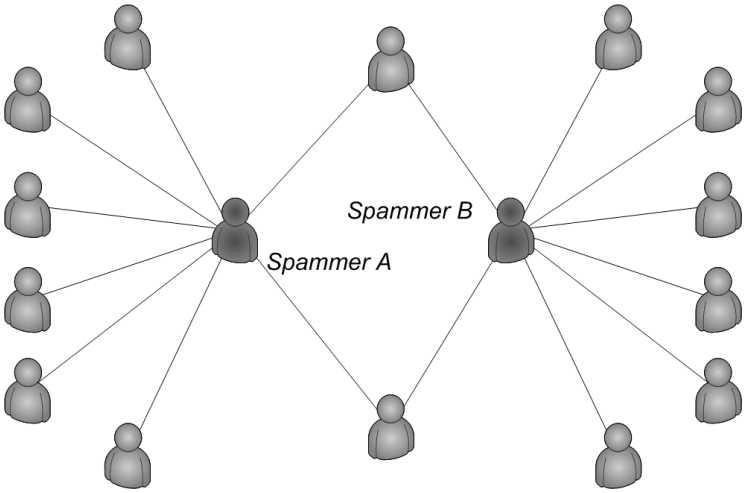
Analisando as relações sociais, geralmente um grupo de pessoas que se conhecem acaba trocando mensagens entre si, criando várias ligações entre os nós que representam cada pessoa. Já um *spammer* envia mensagens para uma grande quantidade de pessoas que não respondem ao *spammer* e, em geral, não possuem nenhuma relação entre si, fazendo com que existam várias ligações entre o nó que representa o *spammer* e os nós representantes dos usuários que receberam o *spam*, mas sem ligações entre diferentes nós. Analisando o grau de agrupamento de cada componente, se o mesmo for alto, pode-se classificar a componente como formada por uma relação social. Já no caso do coeficiente de agrupamento ter um valor baixo, a componente trata-se de uma componente anti-social que foi formada por um processo de envio de *spams*. Na Figura 5.16(a) é mostrado o exemplo de uma rede social representando a relação social entre as pessoas. Pode-se notar o grande número de arestas entre os nós. Já na Figura 5.16(b) é mostrado um exemplo de rede formada por dois *spammers* que enviam mensagens para um certo número de usuários em comum e para grupos distintos. Observando a figura, pode-se notar que não existe nenhuma ligação entre dois nós através de um terceiro, já que nenhum dos nós tem relações em comum. Já que os *spammers* enviam mensagens para um grande número de destinatários, podem acabar enviando uma mensagem para duas pessoas que possuem uma relação social, fazendo com que as duas componentes representando a rede social e a anti-social se unam em uma só componente. Essa união pode reduzir o grau de agrupamento dessa nova componente, dificultando a classificação como uma rede social ou anti-social. Para resolver esse possível problema, no processo de separação das componentes é utilizado um processo de remoção de arestas que fazem a ligação entre muitos nós, já que as arestas que ligam duas componentes que anteriormente eram separadas, são os únicos caminhos entre nós das duas componentes.

Após a classificação das componentes em redes sociais ou anti-sociais, os indivíduos pertencentes às redes sociais são colocados em uma lista branca, ajudando a reduzir a taxa de falsos positivos, que é o parâmetro mais importante de um sistema anti-*spam*. Já os *spammers* das componentes anti-sociais são colocados em uma lista negra, bloqueando suas mensagens. Uma contramedida que pode ser adotada pelos *spammers* é utilizar um endereço de origem diferente para cada *spam* enviado, impossibilitando dessa forma a formação de uma componente que pode ser classificada como sendo anti-social. É importante observar que os *spammers*, se servindo deste artifício, reduzem a eficácia das listas negras mas a rede social continua eficiente na formação das listas brancas, que reduzem os falsos positivos. Uma outra contramedida que pode ser adotada pelos *spammers* é a

obtenção da lista branca dos usuários. Neste caso, se os *spammers* enviarem os *spams* com remetentes forjados iguais aos endereços que se encontram na lista branca a eficácia do mecanismo fica totalmente anulada, pois o *spammer* passou a pertencer ao grupo social dos destinatários. Para a defesa contra tipo de ataque podem ser usados mecanismos de verificação da origem, como descrito na Seção 5.4.3. Se os *spammers* invadem e se servem de uma máquina de algum usuário pertencente a uma rede social, este mecanismo é ineficiente para identificar *spams* enviados desta máquina para outros usuários pertencentes a esta mesma rede social.



(a) Rede social.



(b) Rede anti-social.

Figura 5.16. Um exemplo de uma rede social e de uma rede anti-social.

5.4.3. Sistemas baseados na verificação da origem

O endereço de origem do remetente pode ser facilmente falsificado, dificultando o rastreamento da fonte de envio de *spams*, já que o protocolo SMTP não tem nenhum mecanismo de autenticação ou de verificação da origem. Os métodos baseados na verificação da origem geralmente buscam dois objetivos que podem ser concorrentes ou não.

Esses mecanismos podem buscar confirmar a autenticidade do endereço de origem e/ou determinar se o remetente não é um programa de envio automático de mensagens, que geralmente é utilizado por *spammers* para enviar uma grande quantidade de mensagens. Nesta seção serão apresentados os principais sistemas.

5.4.3.1. Verificação do endereço DNS reverso

A verificação do endereço de DNS reverso foi um dos primeiros mecanismos anti-*spam* baseados na verificação da origem que surgiram. O objetivo da verificação do endereço de DNS reverso é tornar mais difícil a falsificação do endereço de origem, que pode ser facilmente forjado no protocolo SMTP, como mostrado na Seção 5.3.1. O sistema DNS tem dois principais tipos de registro, os registros A e os registros PTR. Os registros A são utilizados para fazer o mapeamento entre nomes de domínio e endereços IP. Quando é feita uma consulta a um servidor de DNS para descobrir o endereço IP de um determinado domínio, o registro A é consultado. Por outro lado, pode ser feita uma requisição do registro PTR de um determinado endereço IP, para descobrir o nome de domínio registrado para ele. Esse tipo de consulta é chamado de consulta reversa, pois funciona de forma contrária ao mecanismo normal de resolução de nomes para endereços IP.

Uma maneira de verificar a autenticidade da origem é fazer uma consulta reversa ao endereço IP do servidor que está tentando enviar uma mensagem. Quando o servidor do destinatário recebe uma conexão para receber uma mensagem, ele em primeiro lugar faz uma consulta DNS reversa do endereço IP do servidor que se conectou a ele. Se o endereço IP não possuir um nome associado, a mensagem é então descartada. Caso exista um nome registrado, é feita uma consulta DNS desse nome, para verificar se o endereço IP desse nome realmente corresponde ao endereço IP original. Caso os endereços IP sejam correspondentes, diz-se que o endereço de DNS reverso é valido. Em seguida, o servidor espera pelos comandos `HELO` e `MAIL FROM` do protocolo SMTP e em seguida compara se os domínios informados nesses dois tipos de mensagem estão de acordo com o domínio que foi obtido pela verificação reversa do endereço IP. Caso os domínios sejam diferentes, a mensagem é recusada. Com essa medida, apenas os servidores cujo endereço IP tem como endereço reverso um nome que pertença ao domínio podem enviar mensagens do domínio. Na verificação do DNS reverso, muitas vezes esse segundo passo não é executado, pois podem acontecer situações onde o servidor de correio eletrônico está em um domínio diferente do qual está enviando as mensagens, como é o caso de servidores que apenas encaminham mensagens de outros domínios.

Essa verificação de DNS é feita, pois os *spammers* geralmente não configuram o endereço reverso de seus servidores, já que se forem configurados, pode-se obter o nome do domínio que o endereço IP pertence. A partir do nome do domínio, podem-se descobrir informações sobre a pessoa que registrou o domínio, aumentando as chances de rastreamento. Provedores de serviço da Internet, em geral, também não registram o endereço reverso de seus clientes, fazendo com que máquinas funcionando como zumbis também não passem no teste de verificação do DNS reverso.

O teste de DNS reverso tem uma baixa taxa de falsos negativos, já que ele elimina grande parte dos *spams* gerados por máquinas zumbis sem DNS reverso e por servidores

que não possuem registro de DNS reverso de forma proposital para dificultar o rastreamento. Em contrapartida, sua taxa de falsos positivos é geralmente bem alta, acima da média dos outros sistemas anti-*spam*, pois muitos provedores legítimos de correio eletrônico não configuram corretamente seus DNSs reversos, fazendo com que as mensagens de todos os seus usuários sejam descartadas por servidores que utilizam a verificação do DNS reverso.

5.4.3.2. *Sender Policy Framework* (SPF)

Esse mecanismo também tem como objetivo dificultar a falsificação do endereço de origem das mensagens. Seu funcionamento se baseia na publicação de informações sobre quais servidores tem permissão de enviar mensagens de um determinado domínio. Dessa forma, cada domínio fica sendo responsável por determinar quais máquinas podem enviar mensagens utilizando o domínio no endereço do remetente. Para determinar as máquinas autorizadas a enviar mensagens, o domínio determina uma série de testes que devem ser realizados por outros servidores que recebam uma mensagem com o domínio do remetente igual ao domínio em questão.

As informações sobre as máquinas autorizadas a enviar mensagens são publicadas em registros no servidor DNS do domínio, utilizando um registro de DNS de modo texto também chamado TXT. O registro do SPF publicado é composto de uma parte inicial, identificada pela seqüência "v=" que especifica a versão utilizada do SPF. Atualmente somente a versão 1 está definida, sendo utilizado o identificador spf1 para a mesma. Em seguida à informação de versão, são definidos os mecanismos que são conjuntos de testes que podem retornar um resultado positivo ou negativo. Para cada um dos mecanismos, podem ser atribuídos modificadores, que irão determinar a ação a ser tomada caso o teste feito pelo mecanismo forneça um resultado positivo. Os mecanismos são avaliados da esquerda para a direita e caso um deles forneça um resultado positivo é tomada a ação definida pelo modificador e os outros mecanismos não são testados.

Os mecanismos definidos na RFC4408 [Wong e Schlitt, 2006] para o SPF são apresentados na Tabela 5.1 e os modificadores utilizados com esses mecanismos são apresentados na Tabela 5.2.

Na Figura 5.17 é mostrado um exemplo de registro SPF permitindo que o servidor cujo endereço IP está associado no nome do domínio e os servidores de correio eletrônico do domínio enviem mensagens. Além disso, todas as regras utilizadas no domínio dominio.com.br também serão verificadas, permitindo que as máquinas autorizadas por ele também sejam aceitas. Todas as outras máquinas que não atendam às características anteriores são impedidas de enviar mensagens como sendo do domínio.

```
v=spf1 a mx include:dominio.com.br -all
```

Figura 5.17. Um exemplo de registro SPF.

O mecanismo SPF, embora não seja utilizado diretamente para filtrar *spams*, pode ajudar a reduzi-los a partir do momento que torna mais difícil o envio de mensagens com

Tabela 5.1. Mecanismos do SPF.

Mecanismo	Descrição
all	Esse teste sempre retorna verdadeiro e é geralmente utilizado como o último mecanismo a ser executado, para definir uma ação padrão caso nenhum dos mecanismos anteriores tenha retornado um resultado positivo.
include	Utilizado para incluir os mecanismos SPF definidos em um outro domínio especificado no parâmetro nome de domínio. Esse mecanismo é geralmente utilizado quando um servidor de um domínio também aceita que as suas mensagens sejam enviadas através de outros domínios, dessa forma incluindo também as restrições impostas pelo outro domínio. A sintaxe desse mecanismo é `include:<nome domínio>`.
a	Realiza uma consulta DNS ao nome do domínio para verificar se o endereço IP do servidor que está enviando a mensagem é um dos endereços IP associados ao nome do domínio.
mx	Através do protocolo DNS, consulta quais são os servidores de correio eletrônico do domínio, através dos registros MX do DNS. Após descobrir os servidores registrados de correio eletrônico, verifica se o endereço do servidor que está tentando enviar a mensagem corresponde a um dos endereços dos servidores registrados.
ptr	Esse mecanismo realiza o teste do DNS reverso, apresentado na Seção 5.4.3.1.
ip4	Define uma faixa de endereços IPv4 que estão autorizados a enviar mensagens. A sintaxe desse mecanismo é `ip4:<endereço de rede>/<máscara de sub-rede>`.
ip6	Similar ao mecanismo anterior só que utilizado para testar faixas de endereços IPv6. Sua sintaxe é a mesma do mecanismo ip4.
exists[2]	Esse mecanismo permite a utilização de macros para criar um determinado nome de domínio baseado em informações da mensagem, como o endereço IP, domínio do remetente e endereço utilizado no comando `HELO` do protocolo SMTP. Com base no nome de domínio que foi criado, verifica-se se o domínio possui um registro de DNS válido. Caso o registro DNS seja válido, o mecanismo retorna um resultado positivo.

endereços falsos, caso os provedores utilizem o SPF para determinar de forma precisa as máquinas autorizadas a enviar mensagens. O uso do SPF, entretanto, não impede que um *spammer* crie vários domínios e publique registros SPF permitindo que qualquer máquina envie mensagens utilizando como remetente esses domínios. Nessa situação os servidores que utilizem o mecanismo do SPF irão concluir que o servidor que está tentando enviar mensagem tem autorização, aceitando a mensagem. Os *spammers*, no entanto, não conseguirão falsificar endereços de provedores legítimos que usem o SPF, já que os mesmos

[2] A principal utilização desse mecanismo é em conjunto com as listas negras DNSBL, apresentadas na Seção 5.4.1.1. Com base em macros utilizando o endereço IP do remetente da mensagem, pode-se construir o nome de domínio necessário para verificar se o endereço IP que está tentando enviar a mensagem está na lista negra.

Tabela 5.2. Modificadores do SPF.

Modificador	Descrição
+	A mensagem é classificada como em conformidade com a política definida e o remetente é autorizado pelo domínio a enviar mensagens. Esse é o modificador padrão, tornando opcional sua definição explícita.
-	A mensagem não está de acordo com a política e o remetente não está autorizado a enviar mensagens como sendo do domínio.
?	O resultado da análise é neutro.
$\sim$	Define que o remetente provavelmente não está autorizado a enviar mensagens como sendo do domínio, mas não é feita uma afirmação da sua autenticidade ou não, permitindo que o servidor trate esse caso de forma diferente dos casos em que o servidor garante que o remetente está autorizado a enviar mensagens ou não.

podem definir políticas permitindo que somente seus servidores enviem mensagens como sendo do seu domínio. Embora um registro SPF devidamente configurado garanta que apenas os servidores de correio eletrônico do domínio especificados na política do SPF sejam aceitos como remetentes de mensagens do domínio, devem-se utilizar métodos para a autenticação dos clientes que enviam mensagens para os servidores do domínio. Se não for utilizado um mecanismo de autenticação, um *spammer* pode enviar uma mensagem com o endereço forjado para o servidor do domínio que ele está utilizando no endereço forjado e quando esse servidor reenviar a mensagem para o servidor do destinatário, ela será aceita já que está sendo enviada por uma máquina autorizada.

5.4.3.3. Desafio e resposta

A idéia desta proposta é limitar o envio de *spams*, usando como artifício o aumento do custo por mensagem enviada, fazendo com que esse custo seja maior que o lucro por mensagem. Estima-se que o lucro obtido por mensagem seja no mínimo de 0,01 centavos de dólar por mensagem [Detroit Free Press, 2002]. Uma das primeiras propostas nesse sentido é criar um custo inicial para a criação de uma nova conta no servidor de correio eletrônico. Esse custo pode ser imposto através de uma CAPTCHA (*Completely Automated Public Turing test to tell Computers and Humans Apart*), de desafios computacionais ou do pagamento de quantias financeiras.

Um exemplo de CAPTCHA é requisitar que o usuário digite letras que são dispostas aleatoriamente em uma imagem, como a imagem mostrada na Figura 5.18. Para uma pessoa, o processo de reconhecimento de caracteres é realizado de forma fácil. No entanto, para uma máquina reconhecer os caracteres da imagem, são necessários complexos algoritmos de processamento de imagens [Mori e Malik, 2003]. O reconhecimento pela máquina se torna ainda mais complicado quando são utilizados outros artifícios, como adicionar linhas e fundos de cores diferentes na imagem. Esse mecanismo, embora seja eficiente no sentido de diferenciar uma máquina de uma pessoa, pode causar problemas

para indivíduos com deficiência visual. Para resolver esse problema podem ser utilizados testes baseados em sons.

Figura 5.18. Um exemplo de CAPTCHA.

Desafios computacionais são procedimentos ou algoritimos com um alto custo computacional. Eles são projetados de tal forma que um máquina gaste um tempo grande para sua resolução mas que possa ser resolvido precisamente, diferente das CAPTCHAs que são projetadas para não serem resolvidas por uma máquina. Como o poder de processamento das máquinas é em geral subutilizado, esses desafios computacionais podem ser resolvidos em segundo plano, fazendo com que o usuário praticamente não perceba que eles estão sendo resolvidos. A principal característica desses desafios é que eles devem ter um custo computacional elevado para sua resolução, mas um custo baixo para verificação. Dessa forma, o usuário irá gastar um tempo grande para resolvê-lo, mas o servidor pode verificar facilmente a resposta. Uma proposta de desafio computacional que atende essas características é um desafio onde deve-se encontrar uma determinada seqüência que adicionada ao começo do cabeçalho da mensagem faça com que o seu *hash* tenha um determinado número de zeros nos bits mais significativos. Para resolver esse desafio, são necessárias tentativas de um grande número de seqüências e, para cada uma, que seja calculado o *hash* da mensagem. Já para a verificação do resultado, só é necessário adicionar a seqüência informada ao cabeçalho e calcular o seu *hash*, para verificar se os bits iniciais realmente são zero. O número de tentativas para resolver o desafio vai aumentar conforme a restrição do número de bits iniciais que devem ser iguais a zero, de forma que ajustando esse valor, pode-se definir em média em quanto tempo o desafio será realizado.

Os dois últimos tipos de custo apresentados podem ser convertidos em custos financeiros, levando-se em conta que é necessária certa quantia financeira para manter um computador e uma pessoa e/ou uma pessoa que trabalhe resolvendo as CAPTCHAs. Para resolver as CAPTCHAs, um *spammer* terá que passar parte do seu tempo resolvendo os desafios, ou então contratar uma pessoa para resolvê-los. Nos dois casos, o tempo perdido ou o salário pago irão representar um custo financeiro, estimado da ordem de 2 centavos de dólar por cada teste que tenha que ser resolvido. Os desafios computacionais também representam um custo para o *spammer*, já que para resolvê-los, o *spammer* precisará utilizar a capacidade de processamento de um ou mais computadores, o que gera custos de aquisição e manutenção.

A proposta de realizar a cobrança de um determinado custo apenas inicialmente é a mais utilizada atualmente. Na maioria dos provedores, para um usuário criar uma nova conta ele precisa resolver uma CAPTCHA. Depois do desafio ter sido resolvido, ele não precisa resolver nenhum desafio e passa a ser limitado apenas pelo número de mensagens diárias que pode enviar. Goodman [Goodman e Rounthwaite, 2004] faz uma análise desse método levando-se em consideração o custo inicial (C), a probabilidade de uma pessoa que recebeu um *spam* notificar o provedor de correio eletrônico que originou

o mesmo (p), o número médio de dias entre o recebimento do *spam* e a reclamação (L) e, por fim, o limite diário de mensagens que podem ser enviadas (D). Nos primeiros L dias, o *spammer* conseguirá enviar D mensagens todos os dias, que é o limite máximo de mensagens, já que mesmo se a primeira pessoa que recebeu uma mensagem sua faça uma reclamação, ela só será feita, em média, após os L dias. Dessa forma, durante esse tempo, ele poderá continuar enviando *spams*. A partir do dia $L+1$, a probabilidade de reclamação para cada dia, será dada pela probabilidade de qualquer uma das D mensagens enviadas serem reportadas como *spam*. Como a probabilidade de cada mensagem não ser reportada é $(1-p)$, a probabilidade de nenhuma mensagem ser reportada no dia será $(1-p)^D$. Por fim, a probabilidade q de alguma mensagem ser reportada a cada dia será justamente o complemento da probabilidade de nenhuma mensagem ser reportada, o que leva a $q = 1-(1-p)^D$. Como nos primeiros L dias o *spammer* poderá enviar D mensagens diariamente e depois disso poderá continuar enviando em média por $1/q$ dias, o número total de mensagens que serão enviadas será $LD + D/q$. Para valores pequenos de D, q será aproximadamente pD, então o número de mensagens aproximado será $LD + 1/p$. A probabilidade de reclamação é, em geral, bem pequena, da ordem de $1/1000$. Já o termo L tem ordem de grandeza de dias e D de dezenas de mensagens, fazendo com que o termo $1/p$ seja maior que o termo LD. Esse fato leva a aproximação final de $1/p$ mensagens que podem ser enviadas, não tendo o parâmetro D grande importância. Como somente foi pago um custo C inicialmente e foram enviadas aproximadamente $1/p$ mensagens, o custo por mensagem será de C/p. Supondo o custo de dois centavos de dólar do desafio, e a probabilidade p sendo $1/1000$, o custo por mensagem será de 0,002 centavos, bem abaixo do lucro por mensagem. Essa é a razão pela qual esse método de cobrança inicial não conseguiu deter os *spammers*.

Uma evolução natural desse método seria impor um custo em cada mensagem. Essa abordagem, no entanto, afetaria de forma negativa os usuários legítimos, que poderiam acabar desistindo de utilizar o serviço. Como uma possível solução, Goodman propõe a cobrança de custos a cada *n* mensagens [Goodman e Rounthwaite, 2004], só que essa cobrança só é realizada *k* vezes, depois disso o usuário só é limitado pelo número de mensagens por dia que podem ser enviadas. A princípio, essa proposta não parece ser boa, mas os resultados mostram que o custo por mensagem utilizando-se o esquema de cobrança apenas nas *k* primeiras vezes tem um resultado bastante similar ao obtido caso a cobrança fosse feita indefinidamente, mostrando que esse método não diminui de forma considerável o custo para o *spammer*, mas é vantajoso para o usuário, que não tem que pagar para sempre os custos.

Outro sistema baseado em desafios e repostas que é utilizado atualmente funciona de forma similar às listas cinzas, apresentadas na Seção 5.4.2.2. Diferentemente do método de cobrança de um determinado custo por mensagem ou por grupo de mensagens, nesses sistemas os usuários precisam resolver um desafio que é enviado pelo servidor do destinatário na primeira vez que enviarem uma mensagem ao destinatário. O servidor, ao receber uma mensagem para um dado par destinatário/origem que ainda não foi confirmado, guarda a mensagem e envia um desafio para o remetente da mensagem, geralmente uma CAPTCHA. Caso o remetente responda o desafio, a mensagem que tinha sido guardada é entregue ao destinatário e nas próximas vezes que o par se comunicar não será mais necessária essa verificação. Como os *spammers* utilizam mecanismos automatiza-

dos de envio de mensagens com endereços de origem forjados, acabarão não recebendo as mensagens contendo os desafios e, por conseqüência, suas mensagens não chegaram ao destino. Devido a essas características, esse método é bastante eficaz em relação à taxa de falsos negativos. Entretanto, usuários legítimos podem não responder aos desafios enviados. Dessa forma, mensagens legítimas acabarão sendo descartadas, gerando uma alta taxa de falsos positivos.

5.4.4. Perspectivas futuras

Não existe hoje nenhum indício que permita inferir que a atividade de enviar *spams* diminuirá nos próximos anos. Ao contrário, os *spammers* vêm se especializando e usando técnicas cada vez mais elaboradas para burlar os sistemas anti-*spam*. Vale ressaltar que os sistemas anti-*spam* estão em constante evolução já que para cada novo mecanismo criado novas técnicas são desenvolvidas pelos *spammers* para enganá-los e permitir a passagem das mensagens não solicitadas.

Uma nova forma de mensagem não solicitada que está surgindo é o *spam* através de serviços de voz sobre IP (VoIP) [MacIntosh e Vinokurov, 2005]. Os *spams* de VoIP, também chamados de SPITs (*Spam over Internet Telephony*), consistem de mensagens, em sua maioria de conteúdo publicitário, enviadas em difusão através de serviços de telefonia IP. Espera-se que, em um curto espaço de tempo, o volume de *spams* dessa natureza cresça em virtude do aumento do número de usuários dos sistemas de telefonia sobre IP. Assim como as mensagens não solicitadas de texto, os *spams* de VoIP têm como atrativo para o *spammer* um custo bem menor do que o custo do envio de mensagens através da rede telefônica convencional. O telemarketing e as gravações publicitárias são formas de *spam* usando a rede telefônica convencional. Pela própria característica da Internet, que utiliza a técnica de comutação de pacotes, um *spammer* pode enviar simultaneamente um grande número de mensagens utilizando apenas um acesso à Internet. Por outro lado, na rede convencional de telefonia isto seria inviável, pois para se enviar várias mensagens de telemarketing e gravações publicitárias ao mesmo tempo seriam necessárias diversas linhas telefônicas. Para combater os *spams* de VoIP, a maioria dos mecanismos anti-*spam* deve ser modificada, já que muitos desses mecanismos se baseiam no conteúdo do texto da mensagem para filtrá-las. A análise por conteúdo se torna muito difícil nos *spams* de VoIP, pois exige a execução de algoritmos de reconhecimento de voz. Atualmente, esses algoritmos demandam um alto custo computacional e não são muito eficientes. Além disso, ao contrário de uma mensagem de texto que é recebida por um servidor e depois encaminhada ao destinatário para que ele a leia quando quiser, uma chamada telefônica ocorre em tempo real, ou seja, é necessário que o usuário atenda a chamada para que ela se inicie. Sendo assim, a análise do conteúdo de um *spam* de VoIP tem que ser feita no momento da comunicação, o que torna o mecanismo ineficiente, uma vez que o usuário já atendeu à ligação e escutou um determinado trecho da propaganda. Por esse fato, os *spams* de VoIP tendem a se tornar ainda mais incômodos do que os *spams* de texto para os usuários, que precisam parar suas atividades para atender a chamada e só depois descobrir que era uma mensagem não solicitada.

As mensagens não solicitadas em forma de vídeo também estão surgindo e assim como os *spams* de VoIP são difíceis de ser classificadas, já que para analisar o conteúdo das mensagens de vídeo são necessárias técnicas de processamento e reconhecimento de

padrões de imagem. Os *spams* de vídeo estão começando a aparecer em sítios da Internet que, sem a permissão do usuário, carregam e mostram um determinado vídeo, que na maioria das vezes possui um conteúdo publicitário. Dessa forma, durante a exibição do vídeo, a atenção do usuário é desviada para o anúncio, devido aos sons e aos movimentos. A identificação dos *spams* de vídeo é complexa, pois em muitas ocasiões um sítio contém um vídeo que é do interesse do usuário e, portanto, não deve ser considerado como um *spam*.

Outra forma indireta de *spam* corresponde à manipulação do resultado de mecanismos de busca na Internet [Gyongyi e Garcia-Molina, 2005], como o Google. A finalidade desses *spams* é fazer com que um determinado produto ou sítio apareça como uma das primeiras referências retornadas pelos mecanismos de busca, quando é realizada a busca de determinadas palavras. Para manipular os mecanismos de busca, os *spammers* se aproveitam do fato de que esses mecanismos, geralmente, dão mais importância a sítios que são referenciados por outros sítios da Internet [Brin e Page, 1998]. Dessa forma, um *spammer* cria vários sítios que contém apenas atalhos para um determinado sítio do seu interesse. Neste sítio, que na maioria das vezes tem conteúdo pornográfico ou comercial, são inseridas palavras com as quais se deseja manipular o resultado dos mecanismos de busca. Geralmente, estas palavras estão camufladas e não têm qualquer relação com o conteúdo do sítio. Assim, como o sítio do interesse do *spammer* contém a palavra buscada e é muito referenciado por outros sítios, ele acaba tendo um resultado de destaque na busca e, conseqüentemente, atrai um grande número de usuários que somente após acessar o sítio descobrem o seu real conteúdo.

Na luta contra os *spams*, deve-se levar em consideração que a maioria dos usuários da Internet não tem formação técnica em computação, possuindo uma capacidade limitada para gerenciar e configurar seus computadores. Portanto, é fundamental que se construam sistemas o mais independentemente possível da intervenção humana. Uma das propostas nesse sentido é a proposição de sistemas autônomos [Kephart e Chess, 2003] para combater os *spams*. A idéia é fazer com que os sistemas anti-*spam* sejam capazes de se adaptar, sem a intervenção humana, às novas técnicas que vão sendo criadas pelos *spammers*. Tais sistemas devem possuir, além da característica de auto-aprendizado já encontrada em alguns sistemas atuais, as propriedades de auto-gerenciamento, auto-manutenção, auto-configuração e auto-recuperação. Neste novo paradigma, já existem propostas que utilizam mecanismos bio-inspirados [Oda e White, 2003] criando um sistema imunológico artificial capaz de combater os *spams*. O sistema imunológico humano se baseia na identificação e na destruição de agentes que causem mal ao sistema, chamados de patógenos, que podem ser vírus, bactérias ou fungos. O sistema imunológico para detectar e neutralizar os patógenos utiliza linfócitos, que são células capazes de identificar os patógenos e destruí-los. A idéia do mecanismo bio-inspirado proposto por Oda e White é utilizar linfócitos virtuais, que de forma análoga aos linfócitos do sistema imunológico humano, possuem receptores capazes de se ligar a antígenos, que representam as características do *spam* [Oda e White, 2003]. Um antígeno é uma partícula capaz de iniciar a produção de um anticorpo específico. A proposta de adoção desse modelo se baseia no fato de que um *spam* é similar a um resfriado. Ele não causa efeitos graves, mas é um mal que incomoda as pessoas e está em constante evolução.

Está claro que o envio de *spams* se tornou uma atividade financeira atrativa para

os *spammers* tanto para anunciar produtos e serviços quanto pela possibilidade de enriquecimento ilícito através de fraudes. Para tanto, os *spammers* estão se especializando cada vez mais. Neste sentido, prevê-se uma batalha interminável entre os *spammers* e os desenvolvedores de sistemas anti-*spam* e a criação de medidas legais para punir os infratores. A solução mais eficaz passa pela destruição da base do modelo de negócios dos *spammers*, que se baseia na idéia de que se mesmo um percentual muito reduzido de usuários comprarem os produtos anunciados por eles, seus anunciantes irão obter lucro e continuaram os financiando. A verdadeira base do problema está na conscientização dos usuários em não comprarem produtos anunciados através de *spams* que, na prática, é muito difícil de ser atingida.

Agradecimentos

Este trabalho foi realizado com recursos da CAPES, CNPq, FAPERJ, FINEP, RNP e FUNTTEL.

Referências

[Agência Globo, 2005] Agência Globo (2005). Brasil é 5^o maior receptor de spam; spywares representam 22% das infecções. http://www.certisign.com.br/certinews/banconoticias/2005/agosto/agosto_15_Brasil_e_5_maior_receptor_de_spam.jsp.

[Andreolini et al., 2005] Andreolini, M., Bulgarelli, A., Colajanni, M. e Mazzoni, F. (2005). Honeyspam: Honeypots fighting spam at the source. Em *International Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI'05)*, páginas 77–83.

[Apache, 2006] Apache (2006). Spamassassin. http://spamassassin.apache.org/.

[Boykin e Roychowdhury, 2005] Boykin, P. O. e Roychowdhury, V. P. (2005). Leveraging social networks to fight spam. *IEEE Computer Magazine*, 38(4):61–68.

[Brin e Page, 1998] Brin, S. e Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Seventh International World-Wide Web Conference*.

[Canter e Siegel, 1994] Canter, L. A. e Siegel, M. S. (1994). Green card lottery- final one? http://www.bio.net/bionet/mm/dros/1994-April/000326.html.

[CGI.BR, 2006] CGI.BR (2006). Comitê gestor da Internet no Brasil - Antispam.br. http://www.antispam.br/.

[Commtouch, 2006] Commtouch (2006). Spam lab online statistics. http://www.commtouch.com/Site/ResearchLab/statistics.asp.

[Costales e Flynt, 2005] Costales, B. e Flynt, M. (2005). *sendmail Milters A Guide for Fighting Spam*. Addison Wesley Professional, 1^a edição.

[Cukier et al., 2006] Cukier, W. L., Cody, S. e Nesselroth, E. J. (2006). Genres of spam: Expectations and deceptions. Em *Hawaii International Conference on System Sciences (HICSS)*, páginas 1–10.

[Cullen, 2002] Cullen, L. T. (2002). Some more spam, please. *Time*, 160(20):58–59.

[Decreto-lei n^o 2.848, 1940] Decreto-lei n^o 2.848 (1940). Código penal. http://www.planalto.gov.br/ccivil_03/Decreto-Lei/Del2848compilado.htm.

[Detroit Free Press, 2002] Detroit Free Press (2002). Spam king lives large off others' e-mail troubles. http://www.freep.com.

[Emery, 2003] Emery, T. (2003). MIT conference takes aim at spam emails. *Associated Press*.

[FTC, 2005] FTC (2005). FTC - spam - home page. http://www.ftc.gov/spam/.

[Gomes et al., 2006] Gomes, L. H., Bettencourt, L. M. A., Almeida, V. A. F., Almeida, J. M. e Castro, F. D. O. (2006). Quantifying social vs. antisocial behavior in email networks. *ArXiv Physics e-prints*.

[Gomes et al., 2004] Gomes, L. H., Cazita, C., Almeida, J. M., Almeida, V. e Wagner Meira, J. (2004). Characterizing a spam traffic. Em *ACM SIGCOMM conference on Internet measurement (IMC'04)*, páginas 356–369. ACM Press.

[Goodman e Rounthwaite, 2004] Goodman, J. T. e Rounthwaite, R. (2004). Stopping outgoing spam. Em *ACM conference on Electronic commerce (EC'04)*, páginas 30–39. ACM Press.

[Gregory e Simon, 2005] Gregory, P. e Simon, M. A. (2005). *Blocking Spam & Spyware for Dummies*. Wiley Publishing, Inc.

[Grupo Brasil AntiSPAM, 2006a] Grupo Brasil AntiSPAM (2006a). Código de Ética AntiSPAM e melhores práticas de uso de mensagens eletrônicas. http://brasilantispam.locaweb.com.br/main/codigoopt.htm.

[Grupo Brasil AntiSPAM, 2006b] Grupo Brasil AntiSPAM (2006b). Página Brasil AntiSPAM.org. http://brasilantispam.locaweb.com.br.

[Gyongyi e Garcia-Molina, 2005] Gyongyi, Z. e Garcia-Molina, H. (2005). Web spam taxonomy. Em *First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*.

[Hambridge e Lunde, 1999] Hambridge, S. e Lunde, A. (1999). *DON'T SPEW: A Set of Guidelines for Mass Unsolicited Mailings and Postings (spam*)*. RFC 2635.

[Hoanca, 2006] Hoanca, B. (2006). How good are our weapons in the spam wars? *IEEE Technology and Society Magazine*, 25(1):22–30.

[Holmes, 2005] Holmes, N. (2005). In defense of spam. *IEEE Computer Magazine*, 38(4):86–88.

[Hormel Foods, 2000] Hormel Foods (2000). Your use of our trademark SPAM on your "Page-O-SPAM" website. http://www.rsi.com/spam/.

[Jung e Sit, 2004] Jung, J. e Sit, E. (2004). An empirical study of spam traffic and the use of DNS black lists. Em *ACM SIGCOMM conference on Internet measurement (IMC' 04)*, páginas 370–375. ACM Press.

[Kephart e Chess, 2003] Kephart, J. O. e Chess, D. M. (2003). The vision of autonomic computing. *IEEE Computer*, 36(1):41–52.

[Klensin, 2001] Klensin, J. (2001). *Simple Mail Transfer Protocol*. RFC 2821.

[Krim, 2003] Krim, J. (2003). Lawsuits by AOL escalates fight against junk e-mail. *The Washington Post*, 15:A1.

[Laufer et al., 2005] Laufer, R. P., Moraes, I. M., Velloso, P. B., Bicudo, M. D. D., Campista, M. E. M., de O. Cunha, D., Costa, L. H. M. K. e Duarte, O. C. M. B. (2005). *Livro Texto dos Mini-cursos do V Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais*, chapter Negação de Serviço: Ataques e Contramedidas, páginas 1–63. Sociedade Brasileira de Computação.

[Lei n^o 8.078, 1990] Lei n^o 8.078 (1990). Código de defesa do consumidor. http://www.planalto.gov.br/ccivil_03/LEIS/L8078compilado.htm.

[Levine et al., 2004] Levine, J. R., Young, M. L. e Everett-Church, R. (2004). *Fighting Spam For Dummies*. John Wiley & Sons, 1^a edição.

[MacIntosh e Vinokurov, 2005] MacIntosh, R. e Vinokurov, D. (2005). Detection and mitigation of spam in ip telephony networks using signaling protocol analysis. *IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication*, páginas 49–52.

[Mori e Malik, 2003] Mori, G. e Malik, J. (2003). Recognizing objects in adversarial clutter: breaking a visual CAPTCHA. Em *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, páginas 134–141.

[Myers, 1999] Myers, J. (1999). *SMTP Service Extention for Authentication*. RFC 2554.

[Oda e White, 2003] Oda, T. e White, T. (2003). Developing an immunity to spam.

[Pearl, 1988] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc.

[Pfleeger e Bloom, 2005] Pfleeger, S. L. e Bloom, G. (2005). Canning spam: Proposed solutions to unwanted email. *IEEE Security & Privacy Magazine*, 3(2):40–47.

[Postel, 1982] Postel, J. B. (1982). *Simple Mail Transfer Protocol*. RFC 821.

[Project, 2006] Project, A. S. (2006). Spamassassin tests performed: v3.1.x. http://spamassassin.apache.org/tests_3_1_x.html.

[Spammer-X et al., 2004] Spammer-X, Posluns, J. e Sjouwerman, S. (2004). *Inside the SPAM Cartel: Trade Secrets from the Dark Side*. Syngress Publishing, 1^a edição.

[Walker, 2005] Walker, A. (2005). *Absolute Beginner's Guide to: Security, Spam, Spyware & Viruses*. Que Publishing.

[Wiki-Spam, 2006] Wiki-Spam (2006). http://en.wikipedia.org/wiki/E-mail_spam.

[Wong e Schlitt, 2006] Wong, M. e Schlitt, W. (2006). *Sender Policy Framework (SPF) for Authorizing Use of Domains in E-MAIL, Version 1*. RFC 4408.

[Zdziarski, 2004] Zdziarski, J. A. (2004). Bayesian noise reduction: Contextual symmetry logic utilizing pattern consistency analysis. http://bnr.nuclearelephant.com/BNR%20LNCS.pdf.

A. Artigos do código penal brasileiro

TÍTULO II
DOS CRIMES CONTRA O PATRIMÔNIO

CAPÍTULO III
DA USURPAÇÃO

Alteração de limites

Art. 161 - Suprimir ou deslocar tapume, marco, ou qualquer outro sinal indicativo de linha divisória, para apropriar-se, no todo ou em parte, de coisa imóvel alheia:

Pena - detenção, de um a seis meses, e multa.

§ 1º - Na mesma pena incorre quem:

Usurpação de águas

I - desvia ou represa, em proveito próprio ou de outrem, águas alheias;

Esbulho possessório

II - invade, com violência a pessoa ou grave ameaça, ou mediante concurso de mais de duas pessoas, terreno ou edifício alheio, para o fim de esbulho possessório.

§ 2º - Se o agente usa de violência, incorre também na pena a esta cominada.

§ 3º - Se a propriedade é particular, e não há emprego de violência, somente se procede mediante queixa.

CAPÍTULO IV
DO DANO

Dano

Art. 163 - Destruir, inutilizar ou deteriorar coisa alheia:

Pena - detenção, de um a seis meses, ou multa.

Dano qualificado

Parágrafo único - Se o crime é cometido:

I - com violência à pessoa ou grave ameaça;

II - com emprego de substância inflamável ou explosiva, se o fato não constitui crime mais grave

III - contra o patrimônio da União, Estado, Município, empresa concessionária de serviços públicos ou sociedade de economia mista; (Redação dada pela Lei nº 5.346, de 3.11.1967)

IV - por motivo egoístico ou com prejuízo considerável para a vítima:

Pena - detenção, de seis meses a três anos, e multa, além da pena correspondente à violência.

CAPÍTULO VI
DO ESTELIONATO E OUTRAS FRAUDES

Estelionato

Art. 171 - Obter, para si ou para outrem, vantagem ilícita, em prejuízo alheio, induzindo ou mantendo alguém em erro, mediante artifício, ardil, ou qualquer outro meio fraudulento:

Pena - reclusão, de um a cinco anos, e multa.

§ 1º - Se o criminoso é primário, e é de pequeno valor o prejuízo, o juiz pode aplicar a pena conforme o disposto no art. 155, § 2º.

§ 2º - Nas mesmas penas incorre quem:

Disposição de coisa alheia como própria

I - vende, permuta, dá em pagamento, em locação ou em garantia coisa alheia como própria;

Alienação ou oneração fraudulenta de coisa própria

II - vende, permuta, dá em pagamento ou em garantia coisa própria inalienável, gravada de ônus ou litigiosa, ou imóvel que prometeu vender a terceiro, mediante pagamento em prestações, silenciando sobre qualquer dessas circunstâncias;

Defraudação de penhor

III - defrauda, mediante alienação não consentida pelo credor ou por outro modo, a garantia pignoratícia, quando tem a posse do objeto empenhado;

Fraude na entrega de coisa

IV - defrauda substância, qualidade ou quantidade de coisa que deve entregar a alguém;

Fraude para recebimento de indenização ou valor de seguro

V - destrói, total ou parcialmente, ou oculta coisa própria, ou lesa o próprio corpo ou a saúde, ou agrava as conseqüências da lesão ou doença, com o intuito de haver indenização ou valor de seguro;

Fraude no pagamento por meio de cheque

VI - emite cheque, sem suficiente provisão de fundos em poder do sacado, ou lhe frustra o pagamento.

§ 3º - A pena aumenta-se de um terço, se o crime é cometido em detrimento de entidade de direito público ou de instituto de economia popular, assistência social ou beneficência.

B. Artigos do código de defesa do consumidor

TÍTULO I
Dos Direitos do Consumidor

CAPÍTULO V
Das Práticas Comerciais

SEÇÃO III
Da Publicidade

Art. 36. A publicidade deve ser veiculada de tal forma que o consumidor, fácil e imediatamente, a identifique como tal.

Parágrafo único. O fornecedor, na publicidade de seus produtos ou serviços, manterá, em seu poder, para informação dos legítimos interessados, os dados fáticos, técnicos e científicos que dão sustentação à mensagem.

Art. 37. É proibida toda publicidade enganosa ou abusiva.

§ 1° É enganosa qualquer modalidade de informação ou comunicação de caráter publicitário, inteira ou parcialmente falsa, ou, por qualquer outro modo, mesmo por omissão, capaz de induzir em erro o consumidor a respeito da natureza, características, qualidade, quantidade, propriedades, origem, preço e quaisquer outros dados sobre produtos e serviços.

§ 2° É abusiva, dentre outras a publicidade discriminatória de qualquer natureza, a que incite à violência, explore o medo ou a superstição, se aproveite da deficiência de julgamento e experiência da criança, desrespeita valores ambientais, ou que seja capaz de induzir o consumidor a se comportar de forma prejudicial ou perigosa à sua saúde ou segurança.

§ 3° Para os efeitos deste código, a publicidade é enganosa por omissão quando deixar de informar sobre dado essencial do produto ou serviço.

§ 4° (Vetado).