Como o spam afeta a comodidade do Correio Eletrônico

Bianca Madoka Shimizu Oe
Gustavo Shinji Inoue
Rafael Umino Nakanishi

11 de junho de 2012

Resumo

O uso do correio eletrônico se tornou uma necessidade pessoal com o advento da tecnologia. Com esse novo meio de comunicação há maior praticidade e agilidade na troca de mensagens, de forma que não é necessário se locomover longas distâncias para conversar com outras pessoas.

Entretanto, a facilidade adquirida também permite o envio de milhares de mensagens eletrônicas em poucos segundos, que podem ser mensagens importantes, como um aviso de uma empresa de grande porte para seus funcionários, ou *spam*, um fenômeno que cresce dia após dia.

((Mudar isso)) Nosso objetivo, nesta monografia, é mostrar como o spam vem trazendo inconveniências os usuários de e-mail. Mostraremos a origem da palavra e seus usos nos dias atuais. ((colocar algo como em seguida aqui)) métodos utilizados para separar mensagens importantes de spam. Em seguida, alguns exemplos de como esse tipo de mensagem traz desconforto para quem o recebe. Por fim, um estudo de caso, analisando as violações dos códigos da Association for Computing Machinery (ACM) [**?**] e a conclusão.

Sumário

1 Introdução

Nesta seção será feita uma breve descrição de como o termo spam (para falar de mensagens indesejadas) foi originado, seguida de ((blablabla)).

1.1 Origem do termo

A aparição da palavra Spam que originou sua utilização atual ocorreu no episódio 12 da segunda temporada de uma série britânica de comédia chamada *Monty Python's Flying Circus*. Spam é uma mistura de carnes de porco apimentadas e enlatadas, vindo de *SPiced hAM* (Figura 1), criado pela empresa estadunidense *Hormel Foods Corporation*.

Durante o episódio, personagens discutem sobre o cardápio de um café. A carne enlatada é um ingrediente presente em todos os pratos do estabelecimento, sendo considerada algo indesejável por um dos personagens, e, durante a discussão, "Spam"foi dito mais de 50 vezes em menos de 4 minutos.

Apesar de não se saber ao certo quando o termo começou a ser utilizado para denotar mensagens indesejadas, atualmente ele é vastamente utilizado, principalmente para se referir a e-mails de propagandas inconvenientes.

Figura 1: Lata de presunto apimentado enlatado

1.2 Spam ((na atualidade))

A empresa especializada em segurança *M86 Security Labs* [**?**] separa spam em 13 categorias listadas abaixo:

Fraude Seu objetivo é fazer o destinatário acreditar que ganhou algo, como um prêmio.

Adulto Possui conteúdo pornográfico e oferece cadastro gratuito a *sites* adultos ou a serviços de acompanhantes.

Financeiro Relacionado a financiamentos e oferecimento de crédito falso.

Ações Faz propaganda de ações de empresas para causar o aumento do preço das mesmas.

Farmacêuticas Informa sobre vários tipos de drogas e remédios, geralmente promete uma pele melhor, mais energia, perda de peso, entre outros. Como um exemplo, tem-se o Viagra.

Phishing Tenta imitar e-mails legítimos enviados por empresas para conseguir dados e/ou credenciais de seus clientes. Seus alvos mais populares são bancos, *eBay* e *PayPal*.

Diplomas Anuncia qualificações como diplomas de universidades ou cursos de treinamento.

Réplicas Anuncia imitações baratas de produtos como bolsas, relógios e celulares.

Software Anuncia softwares baratos, usualmente prometendo venda de softwares .

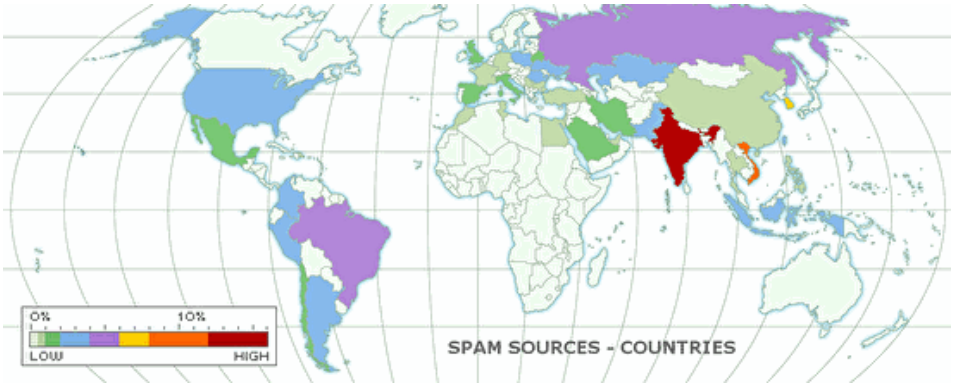
Malware Contem anexos maliciosos ou *links* que levam a *websites* com vírus.

Jogos de azar Promove cassinos ou *sites* de pôquer, que geralmente oferecem bônus por cadastramento.

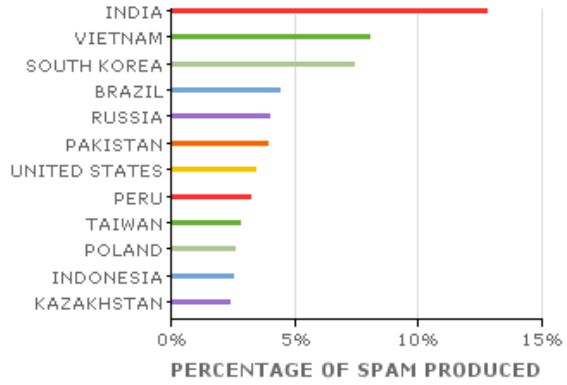
Relacionamento Podem ser incluídos na categoria de fraudes, em que mulheres ou homens que fingem ser mulher tentam criar um relacionamento para extorquir dinheiro.

Outros Não são classificados nas categorias supracitadas.

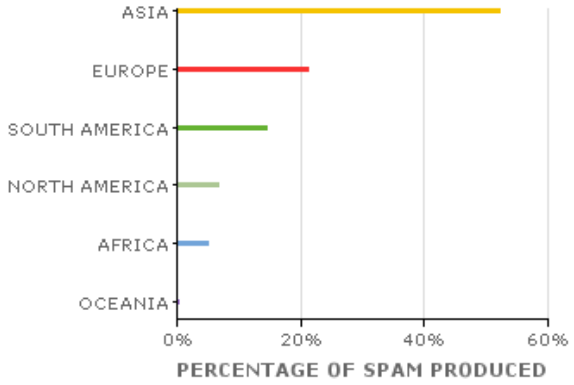
Segundo , 72.7% dos e-mails recebidos são spam e 56% dessas mensagens são farmacêuticas. A distribuição de emissores de spam pelo mundo pode ser vista na Figura 2.



(a) Mapa de emissores de spam((mude isso))



(b) 12 países que mais emitem spam



(c) Percentual de spam emitido por continente

Figura 2: Emissão de spam no mundo

2 Filtros de Spam

Para acompanhar o aumento do número e da variedade de spams, vem sendo criados vários métodos para separar mensagens importantes de propagandas. São inúmeras as abordagens utilizadas. Algumas se baseiam no cabeçalho da mensagem, outras na frequência das palavras utilizadas. Nas seções seguintes, serão brevemente discutidas algumas estratégias utilizadas no combate ao spam, e, como um exemplo, serão mostrados alguns dos filtros utilizados pelo *Gmail*((citar?)).

2.1 Filtro baseado na estrutura do texto

Este tipo de filtro se baseia em cadeias específicas do cabeçalho do e-mail, como, por exemplo, a língua na qual foi escrito e o tipo do conteúdo da mensagem.

Estes filtros podem ser facilmente criados pelos próprios usuários em grandes servidores de e-mail, e tem como objetivo não só detectar spam, mas também separar mensagens relevantes em categorias.

A vantagem desta estratégia é o alto nível de personalização permitido, já que cada usuário pode criar seu próprio filtro dependendo do tipo de spam recebido. Além disso, a probabilidade de haver falsos-positivos é menor, já que é o próprio usuário que escolhe os parâmetros de filtragem.

Sua desvantagem é a possível necessidade de criação de vários filtros para se obter uma separação eficaz.

2.2 *Whitelist*/Verificação

Uma abordagem mais agressiva para a filtragem de spam é a utilização de *Whitelist* em conjunto com a verificação automática.

Qualquer endereço que esteja nesta *whitelist* tem sua mensagem enviada sem maiores problemas. Caso o endereço não esteja contido na lista, o *MTA* envia uma mensagem de volta para o remetente, com instruções que, ao serem seguidas, adicionam o endereço do remetente na *whitelist*.

Como a maioria das mensagens de spam possui endereços de resposta falsos, as instruções não seriam seguidas e o e-mail não chegaria à caixa de entrada do destinatário. Caso o *spammer* decida fazer o que lhe foi dito, ele será adicionado à lista, porém isso o torna mais facilmente rastreável.

Apesar de ser uma maneira eficaz de diminuir os spams, ela pode prejudicar usuários legítimos que não podem ou querem atender a essa exigência, já que isso implicaria no não recebimento de sua mensagem.

Como um exemplo deste tipo de abordagem, tem-se o *Corlive.com* [2], que é um servidor de e-mail que utiliza *Captcha* [4] para validar os e-mails enviados e não permitir que *bots* enviem spam.

2.3 Distribuição adaptativa de *blacklist*

A *blacklist* é formada por mensagens que foram classificadas como spam por usuários ou mesmo por endereços especiamente criados por servidores para atrair spam.

Ao receber uma mensagem, o MTA utiliza um filtro por *blacklist* para determinar se mesma é um spam conhecido, e ela só é enviada à caixa de entrada do destinatário caso ela não seja classificada como ilegítima.

Esta estratégia se baseia no uso de técnicas estatísticas para sumarizar o conteúdo de um mensagem de forma que pequenas mutações no spam não impeçam seu reconhecimento.

A probabilidade de haver falsos-positivos é pequena, já que é necessária uma marcação na mensagem para que ela entre na *blacklist*. Além disso, quando ocorre uma marcação errônea de mensagens vastamente enviadas, como informativos, o gerenciador da lista pode desmarcá-las.

Como é necessária a verificação da mensagem em um servidor, a performance deste método comparada a outros é baixa, sendo ele bastante lento.

Um exemplo de software que implementa esta abordagem é o *Pyzor* [3], um software implementado em *Python* sob a licença *GPL*.

2.4 Ranking baseado em regras

Este filtro possui regras de ranking, principalmente expressões regulares, e tenta fazer correspondências entre o padrão e a mensagem. Cada equivalência adiciona ou diminui pontos da mensagem.

Se a quantidade de pontos do e-mail exceder um determinado threshold, é classificado como spam, caso contrário, é classificado como uma mensagem legítima.

A dificuldade deste tipo de filtro é que, apesar de existirem regras que são contantes com o decorrer do tempo, como endereço de resposta falso ou áudio como tipo de conteúdo, há outras que mudam com o tempo, como os produtos dos quais os spams fazem propaganda, e isso causa a necessidade constante de atualização das regras.

Um software que implementa este tipo de filtro é o *SpamAssassin* [5], um projeto open-source da *Apache*.

2.5 Filtro Bayesiano

A abordagem do filtro Bayesiano, criado por Paul Graham, é a utilização de modelos Bayesianos de probabilidade para determinar se uma mensagem é legítima ou não.

É criado um dicionário de palavras contendo a probabilidade de ela estar em um spam e a probabilidade de estar em uma mensagem válida. Esse dicionário é utilizado para calcular a probabilidade geral de a mensagem ser um spam, com base na teoria da probabilidade condicional de Bayes.

Este filtro possui traz vários benefícios como ser automatizável, ou seja, não são necessárias pessoas para definir os valores de probabilidade, já que ele pode aprender, isto é, adicionar novas palavras ou modificar a probabilidade de uma palavra de acordo com a marcação de um e-mail como spam ou válido.

Além disso, sua implementação e a teoria em que se baseia são extremamente simples e sua performance é melhor que a de filtros baseados em regras.

Um software que implementa o filtro Bayesiano é o *SpamBayes* [6], disponível para os serviços de e-mail *Gmail*, *MSN Hotmail*, *Yahoo! Mail*, entre outros.

2.6 Filtro Markoviano

O filtro Markoviano é um filtro estatístico baseado na teoria de Cadeias de Markov [7], na qual o próximo estado depende unicamente do estado atual, e leva em consideração a probabilidade de transição entre uma palavra e outra, ou seja, dada uma palavra, ele tenta predizer qual é a próxima.

Diferentemente do filtro Bayesiano, que é baseado em palavras independentes, o filtro Markoviano trabalha em cima de frases, e por isso, seu desempenho tende a ser maior, já que utiliza uma abordagem holística do texto.

Esta estratégia é utilizada no filtro de spam *CRM114*, juntamente com outras abordagens que não serão discutidas neste trabalho. Seu desempenho foi testado para diferenciar documentos japoneses confidenciais de não confidenciais, e a acurácia obtida foi maior que 99% e taxa de falsos-positivos foi menor que 5.3% [1].

2.7 Gmail

A técnica mais utilizada pelo provedor de serviços *Gmail* é o aprendizado baseado em instâncias, chamado pelo provedor de *Community clicks*, ou seja, a cada mensagem que o usuário marca como spam ou como mensagem legítima, seus filtros são aprimorados.

Outra técnica é o aprendizado de máquina usado para combinar e indexar os conjuntos de busca do buscador Google, que permite a junção de vários fatores para identificar mensagens semelhantes e classificá-las como spam. Além disso, são aplicadas ferramentas como *OCR* de outro serviço da empresa, o *Google Book Search*, serviço de busca em livros (podem ser imagens), para identificar spams em forma de imagens.

A filtragem também é feita por endereço, ou seja, se um usuário marca a mensagem de um emissor como spam várias vezes, próximas mensagens enviadas por esse endereço irão para a caixa de spam automaticamente, como um filtro por distribuição de blacklist.

A eficácia do filtro pode ser vista na Figura 3.

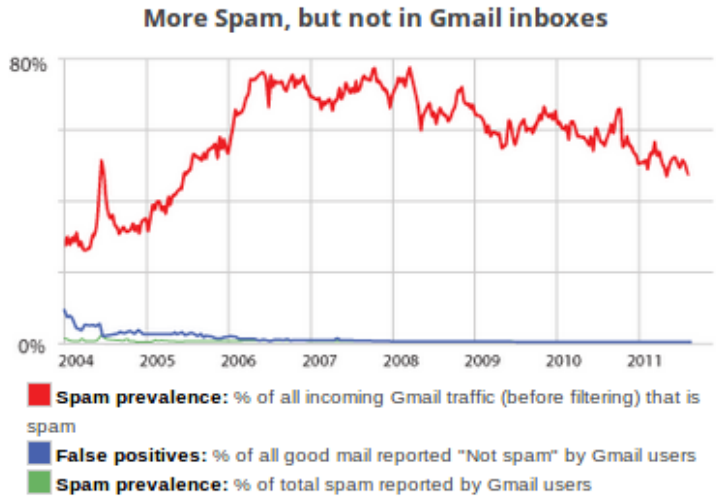


Figura 3: Percentual de filtragem ao longo dos anos

Glossário

cabeçalho Parte do e-mail que contem informações suplementares de transmissão. Entre seus campos, são encontrados endereço do emissor, endereço do receptor, endereço de resposta, data de emissão, tipo do conteúdo e assunto. 4

e-mail Correio eletrônico, termo usualmente utilizado para denotar a mensagem eviada por este meio. 1

MTA Mail Transfer Agent. Software que transfere mensagens de correio eletrônico de um cliente para outro, baseado em uma arquitetura cliente-servidor. 5

Acronyms

ACM Association for Computing Machinery. 1

Referências

[1]

[2] Corlive.com.

[3] Pyzor.

[4] recaptcha.

[5] Spamassassin.

[6] Spambayes.

[7] Weisstein, E. W. Markov chain.