Identificação e Caracterização de *Spammers* a partir de *Honeypots*[*]

Pedro H. Calais Guerra (aluno)
Wagner Meira Jr. (orientador), Dorgival Guedes (co-orientador)

[1]Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
Belo Horizonte, MG.

{pcalais,meira,dorgival}@dcc.ufmg.br

Abstract. *Despite current strategies to minimize the impact of spams, it is necessary a continuous effort to understand in detail how spammers generate and distribute their messages in the network, to maintain and even improve the effectiveness of anti-spam mechanisms. This work proposes a methodology for characterization of spamming strategies based on the identification of spam campaigns – groups of messages that share the same goal and are generated according to the same template. To identify spam campaigns, we designed a data mining technique that detect message invariants and is able to deal with spam evolution. We implemented our campaign detection technique in a system called Spam Miner, which is being used by the Brazilian Internet Steering Committee (CGI.br) and is helping the organization to better understand how the Brazilian network infrastructure is abused by spammers.*

Resumo. *Mesmo com as estratégias atuais que visam minizar os impactos do* spam*, um esforço contínuo para entender como* spammers *geram e distribuem suas mensagens na rede é necessário, para manter e mesmo melhorar a efetividade dos mecanismos de combate ao* spam*. Este trabalho propõe uma metodologia para caracterização de estratégias de disseminação de* spams *baseada na identificação de campanhas de* spam *– grupos de mensagens que têm o mesmo objetivo e são gerados por um mesmo* spammer*. Para identificar as campanhas, foi projetada uma técnica de mineração de dados que identifica os invariantes nas mensagens e que lida com a evolução inerente ao* spam*. O arcabouço de caracterização de campanhas foi instanciado em um sistema (*Spam Miner*) que tem sido utilizado pelo Comitê Gestor da Internet no Brasil (CGI.br) para compreender como a infraestrutura da Internet brasileira é abusada por* spammers*.*

1. Introdução

Simultaneamente ao desenvolvimento e popularização da Internet, o *spam* se tornou um dos maiores problemas de abuso da infraestrutura de redes da atualidade, acarretando prejuízos de bilhões de dólares às empresas e à sociedade em geral [Hayes 2003]. Alguns provedores reportam que entre 40% e 80% das mensagens de correio eletrônico recebidas em seus servidores são indesejadas [Sipior et al. 2004]. Motivados pelo alto impacto negativo do *spam*, pesquisadores de diferentes áreas tem trabalhado na detecção e mitigação

[*]A versão completa da dissertação de mestrado está disponível em www.dcc.ufmg.br/~pcalais/ctd.

do *spam* e do tráfego indesejado em geral. Técnicas de aprendizagem de máquina como filtros *bayesianos*, classificadores SVM e árvores de decisão tem sido aplicados para filtrar *spam* com relativo sucesso [Goodman et al. 2007].

No entanto, o *spam* é um problema de natureza evolutiva, em que *spammers* e *anti-spammers* mudam suas táticas tentando se sobrepor ao outro [Goodman et al. 2007]. Por isso, medir e monitorar as características do tráfego *spam* é um esforço contínuo, de modo a permitir que se reconheça as últimas estratégias utilizadas pelos *spammers* e se possa projetar contramedidas adequadas. Neste trabalho, agrupamos mensagens que correspondem a uma mesma *campanha de spam* – isto é, uma coleção de mensagens que são originadas por um único *template*. A identificação de campanhas é fundamental para caracterizar estratégias de disseminação de *spams* porque *spammers* ofuscam e inserem aleatoriedade no cabeçalho e corpo de suas mensagens [Stern 2008] e, por conta disso, o comportamento de cada *spammer* na rede fica fragmentado nessas mensagens isoladas. A campanha de *spam*, como critério de agrupamento de mensagens, desfaz esse esforço de ofuscação empreendido pelos *spammers* e permite analisar como cada um deles disseminou suas campanhas na rede.

A identificação de campanhas de *spam* é um problema difícil, que já foi tratado na literatura como um problema de detecção de duplicatas de texto, de detecção de imagens similares [Wang et al. 2007] e URLs similares [Xie et al. 2008]. No entanto, essas técnicas consideram características específicas das mensagens (como as imagens e as URLs). A técnica proposta neste trabalho considera diferentes características das mensagens simultaneamente e combina essas características heterogêneas em um arcabouço mais elegante e genérico. Além dsso, enxergamos o problema de identificação de campanhas como um genuíno problema de mineração de dados, em que os padrões de geração das campanhas são minerados, e não pré-determinados. Isso é diferente de outras técnicas que procuram por padrões fixos (como a ofuscação de parâmetros das URLs) e que não conseguem acompanhar a constante evolução dos *spammers*.

As técnicas de de mineração de dados desenvolvidas foram instanciadas em uma ferramenta, denominada *Spam Miner*, capaz de monitorar e caracterizar grandes volumes de tráfego *spam* e que atualmente é utilizado pelo Comitê Gestor da Internet no Brasil (CGI.br) para monitorar o tráfego de *spam* que circula nas redes brasileiras.

Os resultados da dissertação de mestrado do aluno foram publicados em duas conferências nacionais [Guerra et al. 2008a, Guerra et al. 2009a] e duas conferências internacionais [Guerra et al. 2008b, Guerra et al. 2009b], além de apresentados em um *demo* em um evento internacional [Guerra et al. 2009c]. Nas próximas seções, apresentaremos a metodologia do trabalho, bem como os principais resultados obtidos.

2. Coleta de Dados

Os dados considerados no trabalho foram coletados a partir de um conjunto de *honeypots* configurados para coletar *spams*. *Honeypots* são recursos computacionais dedicados a serem sondados, atacados ou comprometidos, em um ambiente que permita o registro e controle dessas atividades [Spitzner 2003]. Para estudar o problema do *spam*, foram implantados 10 *honeypots* desenvolvidos pelo CGI.br [Steding-Jessen et al. 2008] em 5 redes de banda larga brasileiras, emulando *proxies* abertos e *relays* abertos, que são vulnerabilidades comumente abusadas por *spammers* para disseminação de *spams*.

Um *proxy* aberto é uma máquina que permite que conexões sejam estabelecidas de qualquer origem para qualquer endereço e porta de destino. *Relays* abertos, por sua vez, são servidores de correio eletrônico mal-configurados que permitem a entrega de mensagens de qualquer origem para qualquer destinatário. O objetivo dos *spammers* em abusar essas máquinas é ganhar anonimidade na rede.

Os *honeypots* capturaram cerca de 350 milhões de *spams* em um período de 15 meses. A Tabela 1 exibe uma visão geral dos dados coletados. O número de mensagens únicas (cerca de 32 milhões) e de URLs únicas (mais de 6 milhões) também é expressivo. Essas mensagens originaram-se de 160.291 endereços IP, associados a 165 países distintos. Entre esses países, Taiwan foi responsável por 72,28% das mensagens enviadas e poucos abusos se originam de máquinas instaladas na própria rede brasileira (0,16%).

Tabela 1. Visão geral dos dados analisados

Característica	Valor
Período	08/07/2007 à 23/06/2008
Endereços IP	160.291
Sistemas Autônomos	2.557
Mensagens	350.565.583
Mensagens únicas	32.111.981
Spams com URLs	318.881.218 (91%)
URLs únicas	6.413.148

Esses números globais, embora forneçam uma ideia geral dos *spams* que trafegam na rede brasileira, não são suficientes para explicar e determinar o comportamento dos *spammers*, pois os dados são agregados. Em decorrência do grande volume de dados coletados, existem correlações e padrões implícitos entre as grandezas consideradas que não são visíveis ao se tratar a coleção de mensagens como um todo, o que motivou o projeto da metodologia de caracterização de estratégias de disseminação de *spams* proposta neste trabalho baseada em técnicas de mineração de dados.

3. Identificação de Campanhas de *Spam* utilizando uma Árvore de Padrões

A partir das mensagens de *spam* coletadas pelos 10 *honeypots* implantados na Internet brasileira, determinamos grupos de mensagens que possuam o mesmo objetivo e uma mesma estratégia de disseminação, compondo uma *campanha de spam*.

A premissa básica da técnica para identificar campanhas proposta neste trabalho é a de que um *spammer*, ao disseminar uma campanha, mantém estáticas algumas partes e fragmentos da mensagem e altera outras seções do conteúdo, de forma sistemática e automatizada, a partir das ferramentas de *Bulk Mail* [Stern 2008]. Essas ferramentas oferecem recursos para personalização e ofuscação das mensagens, como a inclusão de textos aleatórios no cabeçalho e corpo do *e-mail*. Por exemplo, cada mensagem de uma campanha pode conter termos ligeiramente diferentes no campo *assunto*, embora outras palavras estejam sempre presentes, já que o *spammer* precisa manter o assunto legível e um nível excessivo de ofuscação pode reduzir a probabilidade da mensagem ser lida. Outras formas de geração automatizada de campanhas com conteúdo diferente incluem a inserção do nome da vítima no texto e a inserção de fragmentos aleatórios nas URLs contidas no corpo da mensagem, a fim de prevenir que elas sejam identificadas e bloqueadas.

A técnica desenvolvida neste trabalho para identificação de campanhas de *spam* explora essas propriedades das ferramentas de *Bulk Mail* e é composta de duas etapas. Na primeira etapa, são extraídas características relevantes de cada mensagem de *spam*. Essas características são:

1. idioma (obtido a partir de um classificador de texto)
2. *layout* (sequência de características que captura a formatação visual da mensagem)
3. composição (tipo) da mensagem (HTML, texto, imagem e combinações)
4. URLs contidas no corpo da mensagem
5. assunto

A partir dos atributos extraídos de cada mensagem de *spam*, determinamos invariantes das mensagens a partir de uma estrutura de dados conhecida como Árvore de Padrões Frequentes, introduzida pelo algoritmo *FP-Growth* [Tan et al. 2005]. No nosso caso, cada nó da árvore (após a raiz) representa uma característica extraída dos *spams* que é compartilhada por todas as sub-árvores abaixo. Cada caminho na árvore representa um conjunto de características que co-ocorrem nas mensagens de *spam*, em ordem descrescente de frequência. Dessa forma, as mensagens que compartilham muitas características frequentes em comum (como o idioma, o assunto e o *layout* da mensagem) e diferem apenas por características infrequentes vão compartilhar um caminho comum na árvore. As campanhas são delimitadas pela ocorrência de características infrequentes e aleatórias encontradas em cada caminho. Esses fragmentos infrequentes são identificados porque apresentam uma frequencia significativamente menor que seu ancestral – ou, analogamente, o ancestral possui um grande número de filhos.

Em outras palavras, todas as mensagens presentes em uma sub-árvore que apresenta um aumento significativo no número de filhos em algum ponto da árvore são agrupados na mesma campanha de *spam*. Na prática, a árvore agrupa mensagens que compartilham características invariantes (frequentes) e diferem apenas por características ofuscadas (infrequentes). O poder da técnica reside no fato de que não pré-definimos nenhum padrão de ofuscação; os padrões são detectados naturalmente pela maneira como as características são organizadas na árvore e, portanto, nos torna capazes de identificar diferentes estratégias de ofuscação, mesmo aquelas que ainda não são conhecidas.

A Figura 1 ilustra uma pequena porção da árvore de padrões, mostrando três grandes campanhas ao centro (diferenciadas pelos diferentes padrões de ofuscação, determinados pela sequência distinta de cores em cada nível da árvore). Na Figura 2, vemos uma outra porção da árvore que representa uma única campanha. Podemos notar a sequência de invariantes, que são as características mantidas estáticas pelo *spammer*, seguida por uma sequência de características ofuscadas. O aspecto regular desta sub-árvore, na verdade, define a *assinatura* dessa campanha; as diferentes assinaturas encontradas representam diferentes estratégias de geração de conteúdo dos *spams*.

As nossas análises mostraram que as características ofuscadas mais comumemente encontradas na árvore são fragmentos de URLs gerados aleatoriamente pelos *spammers*, enquanto o *layout*, a composição e o idioma emergem como importantes invariantes de conteúdo. Novamente, é importante salientar que essas diferenças foram naturalmente encontradas pelas nossas técnica, e que se um *spammer* começar a ofuscar um novo atributo,

o novo padrão será automaticamente detectado pela Árvore de Padrões. Dessa forma, lidamos com a constante evolução inerente ao problema do *spam*. A título de exemplo, fomos capaz de identificar campanhas cujos *spammers* conseguem manipular os domínios das URLs e ofuscar o fragmento do domínio das URLs, além dos seus parâmetros. Algumas técnicas assumem que o domínio não será ofuscado e procuram variações no parâmetro das URLs [Xie et al. 2008]; na Árvore de Padrões, o atributo ofuscado não é considerado na determinação dos grupos.

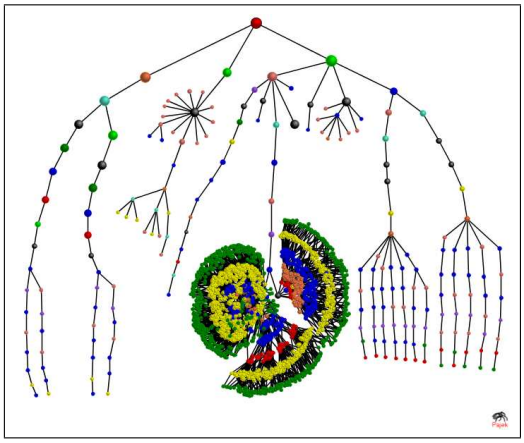

Figura 1. Árvore de Padrões mostrando diferentes campanhas de *spam*

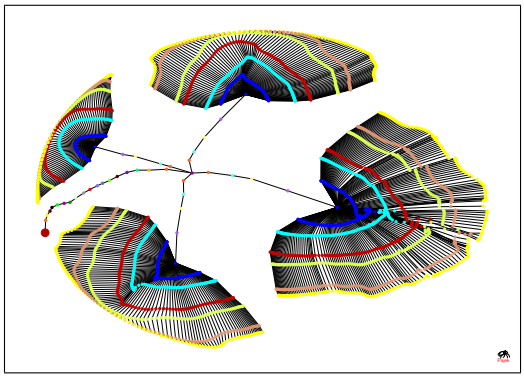

Figura 2. Ramos da Árvore de Padrões agrupados em uma única campanha de *spam*

A Árvore de Padrões foi capaz de agrupar 350 milhões de mensagens (com 6 milhões de URLs únicas) em cerca de 60 mil campanhas de *spam* distintas, o que corresponde a uma redução de aproximadamente duas ordens de grandeza. Esta redução sumariza o conjunto de dados e habilita o emprego de outras técnicas de mineração de dados que não seriam capazes de processar centenas de milhões de mensagens. A determinação das campanhas de *spam* também cria novas dimensões associadas ao tráfego de cada campanha, como o volume e duração dos abusos, que podem ser correlacionados e analisados, além de permitir análises mais complexas, como a coleta e análises das páginas *Web* apontadas pelas URLs presentes nas campanhas.

3.1. O Sistema *Spam Miner*

Os algoritmos desenvolvidos foram integrados em um sistema Web, o qual chamamos *Spam Miner* [Guerra et al. 2009c], que permite que administradores de rede e profissionais de segurança monitorem o tráfego *spam* por meio de campanhas, o que é uma tarefa mais fácil que a verificação individual de cada mensagem. O sistema tem sido utilizado pelo Comitê Gestor da Internet no Brasil (CGI.br) para monitorar o tráfego *spam* na Internet brasileira, e um protótipo está disponível em `http://spammining.speed.dcc.ufmg.br`. O sistema apresenta, na forma de um mapa de calor, a proporção de campanhas de *spam* originadas em cada país abusando a rede brasileira (lado esquerdo da Figura 3) e ícones que, quando clicados, exibem as características principais da campanha sendo originada naquele local (Figura 4), mapeado a partir de cada endereço IP por meio de um serviço de geolocalização. Cada cor representa uma campanha de *spam*, sendo possível avaliar a dispersão geográfica dos abusos.

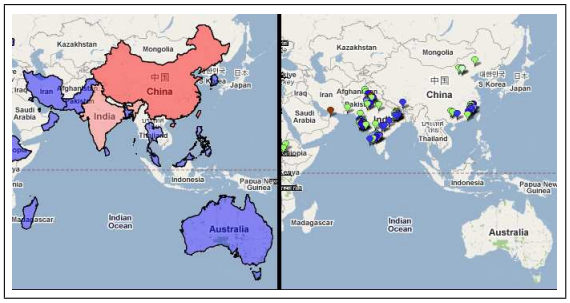

Figura 3. Diferentes Visões Oferecidas pelo Sistema *Spam Miner*: mapa de calor e dispersão de campanhas

Figura 4. Pop-Up exibindo sumário das características das campanhas

As campanhas isolam o tráfego agregado coletado pelos *honeypots* em grupos e as diferentes estratégias adotadas por cada grupo de mensagens esclarece como os *spammers* atuam. A partir das mensagens agregadas, não é possível saber, por exemplo, se há campanhas que abusam ao mesmo tempo *proxies* abertos e *relays* abertos, se um mesmo *spammer* envia mensagens de poucos ou muitos países ou se os abusos de um determinado *spammer* duram minutos, horas ou dias.

Para responder a perguntas como essas, aplicamos outra técnica de mineração de dados – mineração de regras de associação – sobre os atributos de cada campanha, a fim de identificar relações que revelassem diferentes estratégias de disseminação de *spams*. Em particular, determinamos relações entre o país de origem das campanhas de *spam*, o idioma em que as mensagens das campanhas de *spam* estavam redigidas e o alvo das campanhas que esclarecem como a infraestrutura da Internet brasileira é abusada por *spammers*. As técnicas de mineração de dados que aplicamos aos dados coletados nos indicaram que as mensagens de *spam* trafegam no Brasil de duas formas bastante distintas [Guerra et al. 2008a, Guerra et al. 2008b]:

1. A partir de *spammers* que abusam *proxies* abertos no Brasil, enviando mensagens em grande quantidade. A origem das mensagens está associada a seu idioma e domínios de destino. Esses abusos são originados, primordialmente, de máquinas com endereços IP alocados aos *Country-Codes* TW e CN e com o sistema operacional *Windows* instalado, contendo mensagens escritas em chinês e inglês e são direcionadas a destinatários da Ásia;
2. A partir de *spammers* que abusam *relays* abertos no Brasil, enviando mensagens em quantidade bastante reduzida. A origem das mensagens não está associada ao idioma nem aos domínios de destino. Essas mensagens são originadas de toda a parte do mundo. A maior parte dos abusos originados de máquinas configurados com sistemas operacionais *Unix* estavam associados a esse tipo de abuso.

Uma outra característica do comportamento dos *spammers* na rede que as campanhas permitiram investigar é o *encadeamento de máquinas* para entrega dos *spams*: observamos que uma mesma campanha tenta abusar simultaneamente nossos *proxies* abertos e *relays* abertos, e tentam se conectar com máquinas de usuários comprometidas após estabelecer conexões com *proxies* abertos. A análise mostrou como *spammers* criam cadeias

de máquinas para se manterem no anonimato, intercalando *proxies* abertos com vários *relays* abertos e máquinas de usuários finais e como isso impacta a efetividade de listas de bloqueio [Guerra et al. 2009a, Guerra et al. 2009b]. Um dos resultados práticos obtidos é a observação de que combater *proxies* abertos ainda é necessário, mesmo com o crescimento no uso de *botnets* para a disseminação de *spams*, porque tais *proxies* ficam "escondidos"na cadeia de disseminação de *spam* e não são observados diretamente pelos servidores de correio. Esse trabalho foi inovador no sentido de que, até então, olhava-se apenas para um determinado ponto específico da cadeia de entrega de *spams*.

4. Conclusões

Neste trabalho, apresentamos uma nova metodologia para caracterização de estratégias de disseminação de *spams*. O processo de análise se inicia com a extração de características essenciais das mensagens coletadas. A partir dessas características, as mensagens sumarizadas são processadas para se obter agrupamentos contendo as mensagens derivadas de uma mesma mensagem original por técnicas de ofuscação. Esses agrupamentos determinam diferentes *campanhas de spam*, que oferecem abstrações de mais alto nível com as quais trabalhamos, como campanhas de *spam* e estratégias de geração dessas campanhas, ao invés de considerar cada mensagem individualmente. O agrupamento de mensagens em campanhas minimiza os efeitos das técnicas de ofuscação empregadas por *spammers*, que sistematicamente alteram o conteúdo das mensagens enviadas.

As mensagens de cada campanha são, então, avaliadas em busca de correlações invariantes, na forma de características que co-ocorrem frequentemente. Dados os grandes volumes de dados e a necessidade de automação do processo de análise, técnicas de mineração de dados foram empregadas em cada etapa do processo, que foi aplicado a um conjunto de dados de aproximadamente 350 milhões de mensagens de *spam* coletadas por *honeypots* e foi capaz de agrupar essas mensagens em cerca de 60 mil campanhas, que foram então caracterizadas em termos das estratégias empregadas para gerar seus conteúdos e disseminá-las na rede.

As principais contribuições do trabalho são:

- o emprego do conceito de *campanha de spam* para isolar o tráfego associado a diferentes *spammers*;
- a proposição de uma técnica (árvore de padrões) que consegue identificar campanhas de forma elegante, eficiente e sem pré-definir padrões, sendo capaz de lidar com a evolução do *spam*;
- a análise do tráfego *spam* que circula na Internet brasileira revelou vários aspectos que descrevem como a infraestrutura das nossas redes é abusada por *spammers*.

Os algoritmos desenvolvidos foram implementados em um sistema – *Spam Miner* – que tem sido utilizado pelo Comitê Gestor da Internet no Brasil (CGI.br) e tem contribuído para que a organização entenda melhor o problema do *spam* no Brasil e como a infraestrutura da Internet brasileira é abusada por *spammers* [Guerra et al. 2009c]. Por meio de técnicas de mineração de dados aplicadas às campanhas de *spam*, o sistema revelou que a maior parte dos abusos a rede brasileira são originados de fora do País, e que esses abusos tem relação com o idioma e destino dos *spams* quando o abuso é direcionado a *proxies* abertos. Por outro lado, abusos a *relays abertos* nas redes brasileiras são mais dispersos e se originam de muitas fontes simultaneamente, além de não guardarem relação com o idioma e destino do *spam*.

Referências

Goodman, J., Cormack, G. V., and Heckerman, D. (2007). Spam and the ongoing battle for the inbox. *Comm. ACM*, 50(2):24–33.

Guerra, P. H. C., Guedes, D., Jr., W. M., Hoepers, C., and Steding-Jessen, K. (2008a). Caracterização de estratégias de disseminação de spams. In *26o Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, Rio de Janeiro, RJ.

Guerra, P. H. C., Guedes, D., Jr., W. M., Hoepers, C., Steding-Jessen, K., and Chaves, M. H. (2009a). Caracterização de encadeamento de conexões para envio de spams. In *27o Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, Recife, PE.

Guerra, P. H. C., Guedes, D., Wagner Meira, J., Hoepers, C., Chaves, M. H. P. C., and Steding-Jessen, K. (2009b). Spamming chains: A new way of understanding spammer behavior. In *Proceedings of the 6th Conference on e-mail and anti-spam (CEAS)*, Mountain View, CA.

Guerra, P. H. C., Pires, D., Guedes, D., Wagner Meira, J., Hoepers, C., and Steding-Jessen, K. (2008b). A campaign-based characterization of spamming strategies. In *Proceedings of the 5th Conference on e-mail and anti-spam (CEAS)*, Mountain View, CA.

Guerra, P. H. C., Pires, D., Ribeiro, M. T., Guedes, D., Jr., W. M., Hoepers, C., Chaves, M. H. P. C., and Steding-Jessen, K. (2009c). Spam Miner: A platform for detecting and characterizing spam campaigns (demo paper). in: International conference on knowledge discovery and data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, Paris, França.

Hayes, B. (2003). Spam, spam, spam, lovely spam. *American Scientist*, 91(3):200–204.

Sipior, J. C., Ward, B. T., and Bonner, P. G. (2004). Should spam be on the menu? *Commun. ACM*, 47(6):59–63.

Spitzner, L. (2003). Honeypots: Catching the insider threat. In *ACSAC '03: Proceedings of the 19th Annual Computer Security Applications Conference*, page 170, Washington, DC, USA. IEEE Computer Society.

Steding-Jessen, K., Vijaykumar, N. L., and Montes, A. (2008). Using low-interaction honeypots to study the abuse of open proxies to send spam. *INFOCOMP Journal of Computer Science*.

Stern, H. (2008). A survey of modern spam tools. *Proceedings of the 5th Conference on Email and Anti-Spam (CEAS). Mountain View, CA.*

Tan, P., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co.

Wang, Z., Josephson, W., Lv, Q., Charikar, M., and Li, K. (2007). Filtering image spam with near-duplicate detection. In *Proc. of the Fourth Conference on Email and Anti-Spam (CEAS). Mountain View, CA.*

Xie, Y., Yu, F., Achan, K., Panigrahy, R., Hulten, G., and Osipkov, I. (2008). Spamming botnets: signatures and characteristics. *SIGCOMM Comput. Commun. Rev.*, 38(4):171–182.