Caracterização de Estratégias de Disseminação de Spams

Pedro H. Calais Guerra[1], **Dorgival Olavo Guedes**[1], **Wagner Meira Jr.**[1]
Cristine Hoepers[2], **Klaus Steding-Jessen**[2]

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

[2]CERT.br - Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil
NIC.br - Núcleo de Informação e Coordenação do Ponto br, São Paulo, SP

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

{cristine,jessen}@cert.br

Abstract. *To subsidize research on ways to identify and possibly block spam in its origin, avoiding network resources being consumed, we characterize some strategies that define spammers' behavior patterns. For that we use data collected from low-interaction honeypots, configured to emulate open relays and open proxies. After collecting data, we identified message groups that differ only due to text obfuscation, which correspond to a same original spam campaign. We then applied data mining techniques on those groups to find out how such groups use the network resources. The results show that it is possible to identify* spammers *with specific patterns on the way they abuse different ports in parallel and how they start* spam *campaigns from different origins at the same time.*

Resumo. *A fim de subsidiar estudos para identificar e possivelmente bloquear o* spam *em sua origem, evitando que recursos da rede sejam consumidos, caracterizamos algumas estratégias que definem padrões de comportamento de* spammers*. Para tal, utilizamos dados coletados em* honeypots *de baixa interatividade, emulando* proxies *e* relay *abertos. Em seguida, detectamos os grupos de mensagens que diferem apenas por ofuscação de texto, que correspondem a uma mesma campanha de* spam*. Aplicamos então técnicas de mineração de dados para verificar as formas com que os grupos exploram os recursos da rede. Os resultados mostram que é possível identificar* spammers *com padrões específicos na forma como abusam diferentes portas em paralelo, e como iniciam campanhas de* spam *de diferentes origens ao mesmo tempo.*

1. Introdução

O desenvolvimento e popularização da Internet, além de diversos benefícios, acentuou também o crescimento de alguns problemas, como por exemplo o *spam* [Hayes 2003, Messaging Anti-Abuse Working Group (MAAWG) 2007]. O fato do custo de envio de *e-mails* ser muito baixo, comparado ao da correspondência convencional, serve como incentivo ao uso do correio eletrônico para o envio de *e-mails* comerciais não-solicitados em grandes quantidades [Cerf 2005]. Além disso, o *spam* tem sido um meio usual para enviar mensagens relacionadas com a obtenção de dados pessoais com objetivos ilícitos (*phishing*) e para disseminação de códigos maliciosos [Milletary 2005]. Devido à proporção

atingida pela questão do *spam*, várias abordagens técnicas têm sido utilizadas para lidar com o problema como, por exemplo, a adoção de recomendações para configuração de sistemas de *e-mail* [Lindberg 1999, Killalea 2000], uso de filtros de conteúdo de mensagens, como o SpamAssassin [SpamAssassin 2007] e listas de bloqueio [Cook et al. 2006]. Ao mesmo tempo, observa-se um aumento da sofisticação dos *softwares* de envio de *spam*, o que torna as técnicas existentes de bloqueio menos eficientes e a rastreabilidade do *spammer* mais difícil [Cranor and LaMacchia 1998, Milletary 2005]. Um exemplo disso é o crescimento na utilização de máquinas infectadas por códigos maliciosos, como os *bots*, para o envio de *spam* e *phishing*, permitindo que o *spammer* permaneça no anonimato [Milletary 2005]. Esses fatores têm motivado pesquisas no sentido de desenvolver mecanismos para combater o *spam* na sua origem, antes que ele chegue aos servidores de *e-mail* dos destinatários [Messaging Anti-Abuse Working Group (MAAWG) 2005, Gellens and Klensin 2006]. Esse tipo de solução teria também a vantagem de evitar o consumo de recursos da rede com a transmissão de mensagens que seriam eventualmente descartadas pelo destinatário.

Para atingir esse objetivo, no entanto, é necessário entender com detalhes como os *spammers* agem para distribuir suas mensagens na rede. Nesse sentido, o objetivo deste artigo é caracterizar diversas estratégias empregadas por emissores de *spam* para enviar suas mensagens. Definimos como uma estratégia qualquer recurso utilizado pelo *spammer* para maximizar o alcance de suas mensagens, reduzindo a probabilidade de que a mensagem seja retida em filtros anti-spam e que ela seja identificada e rastreada.

Em geral, os *spammers* disfarçam e variam o conteúdo que enviam de maneira sistemática, inserindo trechos aleatórios no corpo da mensagem e nos links nela contidos [Sophos.com 2004]. O objetivo nesse caso é evitar ao máximo o envio de mensagens idênticas, pois isso facilitaria a detecção de sua atuação. Sem a capacidade de identificar grupos de mensagens que tiveram origem em um texto base comum (que definem uma campanha de *spam*), não é possível isolar o tráfego gerado por diferentes *spammers*. Sendo assim, para analisar o comportamento de *spammers* na rede é necessário identificar os grupos de mensagens que correspondam a uma mesma campanha, neutralizando o impacto da ofuscação das mensagens e das estratégias de disseminação, para então estudar as características de exploração da rede de cada um desses grupos. Neste artigo propomos uma metodologia de identificação de grupos de mensagens associadas à mesma campanha de *spam*.

A partir da identificação das mensagens de uma campanha, também propomos uma metodologia de caracterização das estratégias de disseminação de *spam*. Essa metodologia é baseada na detecção de invariantes e padrões de co-ocorrência das estratégias de disseminação em cada grupo de mensagens associadas a uma única campanha. Esses invariantes e padrões representam comportamentos dos *spammers* e podem servir para a definição de critérios para a detecção e a identificação de campanhas de *spam*.

Demonstramos a efetividade da nossa estratégia aplicando-a a parte das mais de 500 milhões de mensagens de *spam* capturadas durante quinze meses por *honeypots* de baixa interatividade [Provos and Holz 2007], configurados de modo a simular computadores atuando como *proxies* e *relays* abertos [Steding-Jessen et al. 2008]. O processo de análise se inicia com a sumarização de características essenciais das mensagens coletadas. As mensagens sumarizadas são processadas para se obter agrupamentos contendo as

mensagens derivadas de uma mesma mensagem original por técnicas de ofuscação. As mensagens de cada agrupamento são então avaliadas em busca de correlações invariantes, na forma de características que co-ocorrem freqüentemente. Dados os grandes volumes de dados e a necessidade de automatização do processo de análise, técnicas de mineração de dados foram empregadas em cada etapa do processo.

Neste trabalho identificamos alguns padrões de comportamento de ofuscação de conteúdo e comportamento de rede que, em última análise, podem ser usados para subsidiar o estabelecimento de novas políticas que visem minimizar os efeitos negativos do *spam*. Consideramos ainda como contribuições a escolha de características a serem consideradas de cada mensagem e os processos de agrupamento e de extração de correlações entre essas características.

2. Trabalhos Relacionados

Diversos trabalhos recentes na literatura estudam as estratégias de abuso dos *spammers*. Um trabalho de 2006 [Ramachandran and Feamster 2006] estuda como os *spammers* exploram a infra-estrutura da Internet para enviar suas mensagens, incluindo as faixas de IP mais usadas para se enviar *spam* e tipos de abuso mais comuns (p.ex., *botnets*, BGP *hijacking*). Entre outras conclusões, os autores destacam o fato de que as mensagens de *spam* tendem a ser enviadas de faixas muito restritas de endereços IP. São mostradas também algumas estatísticas em relação à origem das mensagens, como os sistemas operacionais mais comuns e os sistemas autônomos (AS) que enviam mais mensagens.

O trabalho de Pu e Webb [Pu and Webb 2006] apresenta algumas análises acerca da evolução temporal dos *spammers* no que se refere ao uso de técnicas para construir suas mensagens. Essas técnicas foram computadas a partir das características identificadas nas mensagens pelo filtro *SpamAssassin* [SpamAssassin 2007]. Dessa forma, os autores mostraram que, ao longo do tempo, algumas técnicas de ofuscação deixam de ser usadas, muitas vezes em virtude de mudanças no ambiente, como a correção de alguma falha de segurança nos programas clientes de e-mail. Por outro lado, outras estratégias conseguem persistir por mais tempo.

O artigo de Li e Hsieh [Li and Hsieh 2006] é outro trabalho recente que analisa as estratégias de disseminação de *spam*. Os autores agruparam as mensagens pela URL contida no conteúdo e analisaram a estrutura do grafo que representa as relações entre IPs e URLs, ou seja, uma aresta entre um IP e uma URL significa que o IP enviou uma mensagem referenciando a URL. Eles analisam algumas propriedades desse grafo, com destaque para o surgimento de grandes grupos de IPs que enviam mensagens referenciando a mesma URL.

Por fim, o trabalho de Luiz Henrique Gomes [Gomes et al. 2007] apresentou uma extensa caracterização de cargas de trabalho de *spam*, apresentando comparações com cargas de trabalho de *e-mails* legítimos. Os autores derivaram modelos para representar a taxa de chegada de *spams* e o tamanho das mensagens.

Embora técnicas de mineração de dados estejam sendo extensivamente utilizadas para detecção de spam, não estamos a par de nenhum trabalho que explore a unificação das mensagens como estratégia para caracterizar *spammers*. Em geral os trabalhos avaliam as mensagens como um todo ou individualmente, não havendo uma abordagem sistemática de entendimento das estratégias de disseminação de *spam*.

3. Metodologia

Nesta seção apresentamos a metodologia proposta de caracterização das estratégias dos *spammers*. Como mencionado, essa metodologia se divide em três fases: a coleta dos dados, a sua unificação, com o objetivo de obter grupos coesos de mensagens com relação à sua finalidade, e a caracterização propriamente dita, quando os grupos de mensagens são analisados em busca de características invariantes nas estratégias dos *spammers*. Essas três fases são detalhadas nas seções a seguir.

3.1. Coleta de Dados

A captura das mensagens de *spam* analisadas foi realizada por 10 *honeypots* de baixa interatividade, instalados em redes brasileiras de banda larga de 5 operadoras diferentes (cabo e ADSL). Também fez parte da arquitetura um servidor central, configurado para realizar a coleta do *spam* capturado, bem como a monitoração periódica dos *honeypots* [Steding-Jessen et al. 2008].

Os *honeypots* foram configurados de modo a simular computadores com *proxies* e *mail relays* abertos, tradicionalmente abusados para o envio de *spam* e para a realização de outras atividades maliciosas [Krawetz 2004]. Um *spammer* que tentasse abusar de um desses *honeypots* para o envio de *spam* seria levado a acreditar que teve sucesso em enviar suas mensagens, embora nenhum *spam* fosse efetivamente entregue.

A captura de mensagens utilizou o *software* `Honeyd` [Provos and Holz 2007] em conjunto com subsistemas de emulação de SMTP e *proxies* HTTP e SOCKS. Todas as transações efetuadas pelos subsistemas do `Honeyd` foram armazenadas em *logs* com informações como data e hora, IP de origem, IP e porta pretendidos de destino, assim como versão do protocolo utilizado. Todos os *spams* capturados pelos *honeypots* foram coletados em intervalos regulares por um servidor central através de um túnel criptografado. Ao todo, foram coletados mais de 500 milhões de mensagens, durante um período de 15 meses [Steding-Jessen et al. 2008].

3.2. O Problema da Unificação de Mensagens

A unificação de mensagens tem por objetivo neutralizar o efeito de técnicas de ofuscação e de distribuição utilizadas pelos *spammers*, que alteram sistematicamente o conteúdo das mensagens, seja o seu corpo ou o próprio assunto do e-mail [Sophos.com 2004]. O objetivo é tornar cada mensagem enviada única, e para tal os *spammers* contam com ferramentas desenvolvidas para enviar *spam*, as quais oferecem os mais diversos recursos para que eles possam personalizar (e ofuscar) suas mensagens. Um exemplo típico de estratégia é a inserção de termos aleatórios no texto da mensagem, impedindo que seja gerada uma "assinatura" da mensagem de *spam* que permitiria identificá-la facilmente.

O problema da unificação de mensagens consiste em identificar grupos de mensagens que satisfaçam algum critério de equivalência semântico, isto é, agrupar as mensagens que, por exemplo, embora possuam URLs diferentes e conteúdos com algumas variações, tenham a mesma finalidade, isto é, anunciar o mesmo produto/serviço ou atingir um objetivo comum. Em outra palavras, o problema pode ser entendido como a identificação das campanhas de *spam*.

A Figura 1 ilustra graficamente o processo de unificação. No grafo à esquerda estão as mensagens coletadas, onde os vértices origem das arestas representam emissores

de *spam* e os vértices destino são as mensagens. Sem o processo de unificação, e dado o esforço de ofuscação dos *spammers*, cada mensagem se repete muito pouco (ou não se repete) e qualquer análise sobre o comportamento dos *spammers* é difícil ou inviável.

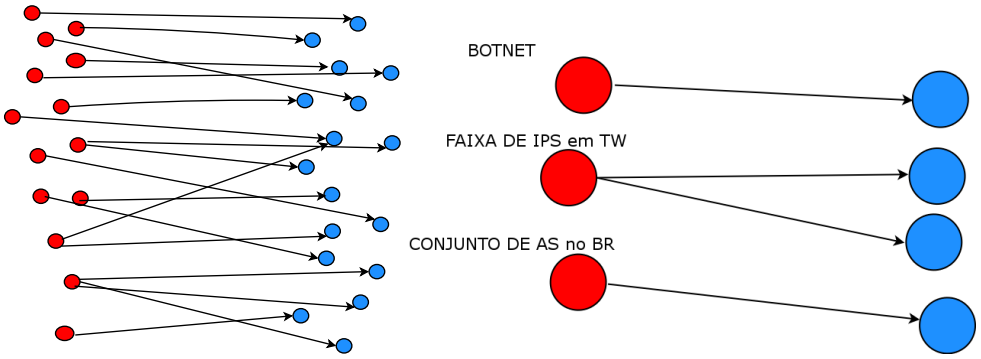

Figura 1. O processo de unificação de mensagens

O grafo à direita, por sua vez, exibe a visão dos dados após a unificação das mensagens que, embora possam ser diferentes, tratem do mesmo assunto e pretendam vender o mesmo produto ou serviço. Essa visão permite identificar grupos de máquinas que enviaram as mesmas mensagens e permite considerar esse grupo de máquinas como uma entidade única geradora de *spam*. Essa entidade pode ser um *botnet* [Cooke et al. 2005], que é um conjunto de máquinas de usuários comuns infectadas por um programa malicioso e programadas para enviar *spam* (máquinas "zumbis"), ou um conjunto de máquinas que situem-se na mesma faixa de IP, ou ainda um conjunto de sistemas autônomos (AS, ou *autonomous systems*).

Consideramos a unificação necessária para analisar a estratégia de disseminação de *spam* e entender como os *spammers* agem pelos seguintes motivos:

- A unificação neutraliza o efeito do volume variável de mensagens associado a cada campanha de *spam*. Para algumas análises, é mais interessante considerar apenas quais campanhas distintas que foram observadas no período, independente do número de mensagens em cada uma.
- A unificação cria novas dimensões que podem ser analisadas e correlacionadas, como o volume de mensagens enviadas durante uma campanha, ou a duração do período durante o qual tais mensagens foram enviadas.
- Como mencionado anteriormente, em pouco mais de um ano de coleta, o sistema armazenou centenas de milhões de mensagens. O processamento dessa quantidade de dados tem alto custo, mesmo com o emprego de técnicas de paralelização de algoritmos. A unificação provê uma sumarização que reduz o volume de dados a ser analisado.

3.3. Identificação de Padrões de Comportamento dos *Spammers*

Após identificarmos os grupos de mensagens componentes de cada campanha de spam, identificamos padrões de comportamento dos *spammers* a elas associados. Para isso procuramos por comportamentos semelhantes que podem identificar estratégias comuns entre os diversos grupos de mensagens. Invariantes identificados nesses grupos sintetizam as diferentes estratégias de *spammers*.

Nesse sentido, aplicamos duas técnicas de mineração de dados: agrupamento e identificação de padrões freqüentes [Tan et al. 2005]. Uma técnica de agrupamento objetiva segmentar um conjunto de entidades em grupos homogêneos que possuam alta semelhança entre seus membros e alta dessemelhança em relação a membros dos outros grupos. Neste trabalho, aplicamos o algoritmo (*k-means*) para identificar grupos de mensagens que exibam comportamento parecido no que se refere ao abuso dos recursos de rede. A mineração de padrões freqüentes procura encontrar atributos que co-ocorrem com freqüência em uma base de dados [Tan et al. 2005]. Aplicamos mineração de padrões freqüentes para encontrar combinações de características das mensagens que ocorrem comumente. Em particular, investigamos as correlações entre país de origem, país de destino e idioma das mensagens.

4. O Processo de Unificação de Mensagens

Nesta seção, apresentaremos a técnica proposta para agrupar as mensagens de *spam* e, em seguida, os principais resultados que foram obtidos.

4.1. Árvore de Padrões Freqüentes

A abordagem proposta para tratar o problema da unificação de mensagens explora o fato de que os *spammers*, em geral, mantêm uma parte da mensagem fixa e variam de forma sistemática e automatizada alguns fragmentos bem definidos. Por exemplo, cada mensagem de uma mesma campanha pode ter um campo *assunto* ligeiramente diferente, embora mantenha algumas palavras-chave padrão; no corpo da mensagem, a saudação pode alternar entre “Hello” e “Hi”; as URLs mencionadas na mensagem podem conter fragmentos aleatórios, que não fazem sentido algum e se prestam apenas a tornar a URL única, a fim de se evitar que ela seja identificada e bloqueada.

O processo de unificação desenvolvido e que explora essas possibilidades é constituído de duas partes. Na primeira, extraímos características relevantes de cada mensagem, tais como idioma, *layout*, tipo da mensagem (HTML, texto, figura), URL e assunto. O idioma de cada mensagem foi extraído através de uma implementação da técnica baseada em cálculo de N-Gramas [Cavnar and Trenkle 1994]. O *layout* corresponde às características de formatação de cada mensagem e baseia-se na proposta de Claudiu Musat [Musat 2006]. Por exemplo, uma mensagem de texto que possua duas linhas em branco, seguida de uma URL e duas linhas de texto será mapeada para o *layout* BBUTT. *Tags* HTML também são consideradas. Como será mostrado, o *layout* é um invariante importante para permitir o agrupamento das mensagens da mesma campanha de *spam*. Isto é, mesmo que os *spammers* insiram textos e caracteres aleatórios, a aparência geral da mensagem não é alterada. A URL de cada mensagem foi quebrada em *tokens* para que cada componente da URL seja uma característica da mensagem a ser considerada.

A partir das características de cada mensagem, monta-se uma árvore de padrões freqüentes (*FP-Tree*) [Tan et al. 2005]. Essa árvore é uma representação em que as características de cada mensagem são inseridas de forma que as mensagens que possuam características em comum compartilhem o mesmo caminho. A árvore é construída de forma que as características mais freqüentes (globalmente) fiquem nos níveis mais altos e as características infreqüentes ou aleatórias fiquem nos níveis mais baixos (próximos às folhas). Dessa forma, duas mensagens que possuam muitas características freqüentes em comum (como o idioma, o tipo e o *layout*) e sejam diferentes apenas por uma

característica infreqüente tendem a ser filhas de um mesmo nodo. Em geral, essa característica infreqüente é um fragmento da URL gerado aleatoriamente pelo *spammer*. A metodologia proposta prevê que as mensagens que sejam filhas de um mesmo nodo sejam agrupadas, pois elas compartilham invariantes (todas as características dos níveis superiores da árvore) e diferem apenas por um padrão infreqüente.

4.2. Resultados da Unificação

A Figura 2 exibe a distribuição do número de mensagens encontradas em cada grupo gerado pela heurística da árvore de padrões freqüentes, em comparação com os agrupamentos que seriam obtidos por um agrupamento baseado exclusivamente em equivalência binária dos conteúdos, que pode ser utilizada como padrão de comparação. A escala é *log-log*. Vemos que a árvore de padrões freqüentes consegue agrupar mais mensagens do que o simples agrupamento binário. Algumas campanhas possuem cerca de 1 milhão de mensagens, enquanto o grupo mais popular do agrupamento binário possui apenas cerca de 100 mil. Um caso clássico de unificação em que mensagens únicas são agrupadas são aquelas utilizadas para validar *e-mails* das vítimas. Nelas o *spammer* insere o *e-mail* do destinatário em uma URL no corpo da mensagem, de forma a poder registrá-lo como válido quando o usuário seleciona o link. Na árvore de padrões freqüentes, cada *e-mail* encontrado em uma URL seria um padrão infreqüente e portanto todas as mensagens estariam no mesmo nível da árvore, compartilhando todas as outras características.

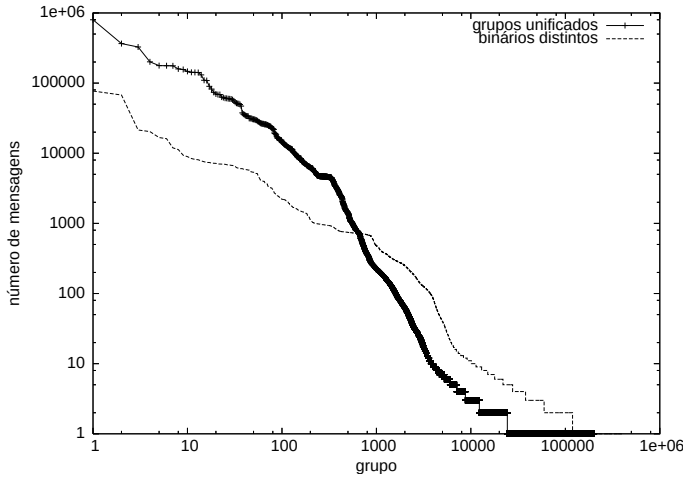

Figura 2. Número de mensagens por campanha de *spam*

A Figura 3 exibe o número de mensagens encontradas em cada nível da árvore de padrões freqüentes. Podemos notar que a árvore é muito desbalanceada e o número de características consideradas por mensagem varia significativamente. Percebe-se que os primeiros níveis contêm poucas mensagens, pois representam características que são muito freqüentes e não são úteis para discriminar as mensagens. Em particular, o primeiro nível contém o fragmento de URL "http://", já que 96,5% das mensagens possuem alguma URL. Características como o idioma da mensagem e o seu tipo também são colocadas na parte superior da árvore. As mensagens mais comuns são do tipo HTML, correspondendo a 86,61% do total. Mensagens de texto puro respondem por 13,35% e os 0,04% restantes são mensagens enviadas como imagens (GIF e JPEG). Após essa divisão de mensagens por idioma e tipo, a próxima característica mais útil para discriminar

mensagens é o *layout*. No período analisado, foram encontrados apenas 2.234 *layouts* de mensagem distintos. Descendo mais na árvore, encontramos os fragmentos de URL que os *spammers* fixam ao enviar suas mensagens. Em geral, os domínios das URL são fixos; entretanto, quando os *spammers* são donos de sub-domínios, eles conseguem inserir aleatoriedade no restante do endereço usado na URL. O pico no nível 6 acontece porque nesse nível concentraram-se as características aleatórias das mensagens, em especial, fragmentos aleatórios das URLs, e há um grande número de caminhos da árvore (e portanto mensagens) que terminam nesse nível. Um outro pico é notado no nível 9, isto é, existem muitas mensagens que compartilham 8 características (idioma, tipo, *layout* e fragmentos da URL) e diferem apenas por uma característica menos representativa. Cerca de 60% das mensagens foram agrupadas nos níveis 6 e 9 da árvore (Figura 3). Por outro lado, algumas mensagens chegam a compartilhar até cerca de 40 características, o que acontece quando a mensagem contém várias URLs e, portanto, o número de fragmentos compartilhados aumenta.

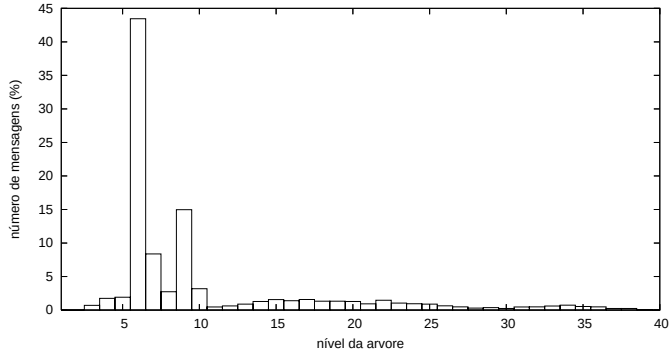

Figura 3. número de mensagens em cada nível da árvore de padrões freqüentes

O trabalho de Yeh e Lin [Yeh and Lin 2006] propõe uma metodologia para agrupamento de mensagens para fins de filtragem de *spam*. Os autores utilizam primordialmente a informação das URLs e comparam a semelhança entre as mesmas para decidir se duas mensagens são duplicatas, caso a similaridade seja superior a um certo limiar. Consideramos nossa abordagem mais robusta no sentido de exigir que as mensagens compartilhem outras características, de não necessitar de parâmetros e de não gerar grupos para mensagens que possuam URLs similares mas as mensagens sejam totalmente distintas, o que é comum em caso de redirecionadores, como *tinyurl.com*. Além disso, os padrões não são pré-definidos; eles surgem naturalmente na montagem da árvore. Em nenhum momento, definimos regras em que as mensagens que serão agrupadas precisem compartlhar certas características e diferir-se por outras. Isso torna a técnica mais resistente ao processo constante de mudança e evolução inerente ao *spamming*. Cabe ressaltar, ainda, que à medida que novas características sejam detectadas e se mostrem relevantes, o resultado da unificação pode ser aperfeiçoado.

5. Estratégias de Disseminação de Spam

Com base na unificação das mensagens discutida, apresentamos a seguir os resultados mais relevantes encontrados para os padrões de comportamento de *spammers*. O período analisado é de um mês, de 01/06/2007 e 01/07/2007. Nesse período, 9.301.182 mensagens foram coletadas e 150.912.121 destinatários distintos foram registrados. Apresen-

tamos três tipos de análise, relativas à forma como os *spammers* utilizaram as portas das máquinas com falhas de segurança, ao relacionamentos entre origem e destino das mensagens e o idioma de cada uma e à caracterização da infra-estrutura usada pelos *spammers* para enviar suas mensagens (a origem de cada campanha).

5.1. Portas Abusadas

Ao se aproveitar de uma máquina como *open proxy* ou *mail relay*, um *spammer* se conecta a uma porta na máquina-alvo referente a algum serviço que usualmente ofereça a opção de repasse de conexões. Dessa forma o *spammer* estabelece uma conexão para uma outra máquina a fim de injetar as mensagens que deseja enviar ocultando sua origem real. A Tabela 1 exibe a proporção de mensagens recebidas por cada porta abusada, obtida a partir da análise dos *logs* dos *honeypots*.

Embora a Tabela 1 mostre que as portas 1080 (SOCKS) e 8080 (HTTP) são as mais abusadas e que o uso de *mail relays* abertos para o envio de mensagens (porta 25) é pouco significativo (menos de 1% das mensagens), não há informações suficientes para se determinar as estratégias que podem estar relacionadas com o abuso daquelas portas. Isso se deve ao fato desse resultado ser apenas um agregado de todas as mensagens e representar um comportamento médio que não é necessariamente representativo.

Tabela 1. número de mensagens recebidas por cada porta abusada

Porta	Porcentagem	Porta	Porcentagem
1080	43,96%	4480	4,23%
8080	17,15%	81	4,10%
8000	10,09%	3382	1,93%
6588	6,63%	25	0,67%
3128	5,74%	3127	0,02%
80	5,48%		

Para melhor identificar as diferentes estratégias de abuso das portas, utilizamos o algoritmo de agrupamento *k-means* [Tan et al. 2005], bastante conhecido e por lidar bem com dados numéricos e poucas dimensões, para encontrar os grupos de campanhas de *spam* que exploram os mesmos conjuntos de portas. Após a execução do algoritmo, encontramos 8 grupos significativos que denotam comportamentos distintos dos *spammers* em relação ao abuso de portas.

Os resultados apresentados na Tabela 2 mostram que cerca de um terço das campanhas de *spam* identificadas (32%) foram enviadas para um mesmo conjunto específico de portas (grupo 1: 8080, 8000, 4480, 3127 e 1080). Por outro lado, muitas campanhas acabam abusando apenas uma porta, caso dos grupos 2 a 5. Uma parcela menor abusa outros conjuntos de portas, em que a porta 6588 é a mais freqüente (grupo 8).

A Figura 4 mostra como as campanhas do grupo 1 distribuíram suas tentativas de abuso das portas do conjunto durante o período amostrado. Pode-se ver que as portas são exploradas em paralelo durante todo o período. Há uma preponderância das portas 8080 e 8000 na maior parte do tempo, mas em em outros momentos todas as portas têm proporções semelhantes.

O conhecimento desses comportamentos distintos pode favorecer políticas de detecção de atividade de *spammers*, já que as proporções de abuso das portas podem ser

Tabela 2. grupos de portas, por frequência relativa de uso das principais portas

No.	**freqüência relativa - uso das portas**	***% de grupos***
1	8080 (30%), 8000 (18%), 4480 (14%), 6588 (14%), 1080 (13%), 3128 (11%)	32%
2	8000 (100%)	20%
3	8080 (100%)	15%
4	1080 (100%)	12%
5	4480 (100%)	7%
6	3128 (74%), 8080 (6%), 8000 (6%), 4480 (5%)	5%
7	80 (80%), 8080 (6%), 8000 (4%)	5%
8	6588 (95%), 8080 (3%), 3128 (2%)	4%

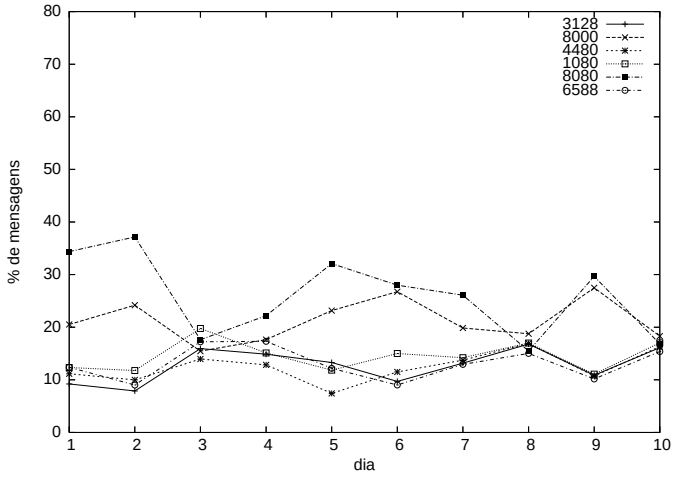

Figura 4. freqüência relativa de mensagens por porta para uma campanha

monitoradas e freqüências similares às descritas podem ser usadas como um forte indício da atividade de *spammers*.

5.2. Correlações entre origem, destino e idioma

A fim de entender melhor as estratégias utilizadas por *spammers* para disseminar suas mensagens, correlacionamos três características: o país de origem da mensagem, o país de destino e o idioma em que a mensagem está redigida. O país de origem foi obtido através da informação de alocação do IP de origem registrado para a mensagem, informação esta obtida através das tabelas de alocação dos cinco *Registries* Regionais [CYMRU 2007]. O país de destino da mensagem foi obtido através do domínio do *e-mail* dos destinatários (o domínio foi convertido para o IP correspondente e então o mesmo processo usado para obter o país de origem foi aplicado). Para essa análise, desconsideramos domínios genéricos como *hotmail.com* e *gmail.com*.

Como o número de mensagens é muito grande e as combinações entre origem, destino e idioma são potencialmente inúmeras, aplicamos mineração de padrões freqüentes [Tan et al. 2005] e analisamos os padrões e as regras de associação que podem ser derivadas deles. A Tabela 3 exibe as combinações entre país de destino, país de origem e idioma mais freqüentemente encontradas nas mensagens. Os resultados mostram que grande parte do *spam* que circulam pela Internet brasileira vêm da Ásia, em especial da China e de Taiwan. No período analisado, mais de 95% de todo o *spam* observado nos emuladores de *open proxy* e *open relay* dos *honeypots*, no Brasil, estavam redigidas em

chinês, o que pode indicar que o Brasil seja abusado por emissores de *spam* Asiáticos para ocultar a origem das mensagens.

Tabela 3. combinações de país de origem, país de destino e idioma

país de origem	país de destino	idioma	% de mensagens
Taiwan	Taiwan	Chinês	81,32%
China	Taiwan	Chinês	10,51%
Taiwan	EUA	Inglês	2,07%
EUA	Taiwan	Chinês	1,56%
Hong Kong	Taiwan	Chinês	1,41%

Aplicando regras de associação sobre esses atributos de cada mensagem, conseguimos revelar padrões e sutilezas que não são imediatamente visíveis. Algumas dessas regras estão listadas na Tabela 4, juntamente com algumas métricas de interesse. O suporte é a freqüência de ocorrência da regra no universo de mensagens.

Tabela 4. regras de associação

regra	antecedente	conseqüente	suporte	confiança	lift
1	Chinês	Taiwan (Origem)	27,8%	88,5%	1,02
2	Taiwan (Origem)	Chinês	26,5%	80,8%	1,02
3	China (Origem)	Chinês	9,1%	83,6%	1,05
4	EUA (Origem)	Chinês	0,7%	74,0%	0,93
5	Brasil (Origem)	Chinês	4,13%	25,1%	0,31
6	Brasil (Origem)	Inglês	2,83%	40,2%	1,12
7	Brasil (Origem)	Português	0,21%	14,8%	3987,94
8	Brasil (Origem)	Alemão	0,001%	0,001%	18,90

A regra 1 indica que 88%[1] das mensagens redigidas em chinês são enviadas a partir de IPs alocados para Taiwan. Olhando por outro ângulo, se a mensagem vem de Taiwan, a probabilidade dela estar redigida em chinês é de mais de 80% (regra 2). Se a origem é a China, a proporção é um pouco maior: 83% (regra 3). As regras de associação estão aqui sugerem a idéia de que o spam na Ásia, apesar de passar pela Internet brasileira, é produzido para "consumo interno".

A regra 4 revela que 74% das mensagens oriundas dos Estados Unidos analisadas durante o período de coleta estão escritas em chinês. Há diferentes explicações para esse fato: pode haver um encadeamento de *proxies*, de forma que a origem das mensagens pode ser simplesmente uma outra máquina abusada, ou pode ser um indício de *IP Hijacking* [Ramachandran and Feamster 2006], ou mesmo que empresas da Ásia usem serviços de *hosting* ou de *spammers* localizados em redes norte-americanas. Mais informações seriam necessárias para uma definição nesse caso.

A regra 5 permite levantar hipóteses semelhantes para o Brasil: o idioma de uma em cada quatro mensagens de *spam* cuja origem é o próprio Brasil é o chinês. O *lift*[2] é baixo (0,31), indicando que o caso mais comum é a origem ser mesmo Taiwan ou China e que o Brasil também encaminha outros tipos de mensagens (em especial, inglês e português). As regras 6 e 7 representam esses casos. Em particular, o *lift* para mensagens

[1] confiança: probabilidade do conseqüente dado o antecedente.

[2] lift: razão entre a probabilidade obtida e a esperada, se os eventos que compõem a regra fossem independentes.

enviadas em português a partir do Brasil é bastante alto, o que indica que o IP de origem brasileiro aumenta muito a chance da mensagem estar em português e ser destinada a brasileiros. Isso é um reflexo do fato de que no período analisado não foram encontradas mensagens de *spam* em português oriundas de IPs na Ásia ou EUA.

Por fim, a regra 8 indica que mensagens em alemão também circulam pela rede brasileira. Embora em quantidade pequena, se comparado, principalmente, com o volume de mensagens em chinês, essas mensagens superam sensivelmente o que seria esperado por uma simples amostragem estatística dos dados (valor alto de *lift*).

Essas regras indicam que técnicas que levam em conta aspectos de idioma podem ser úteis para sinalizar comportamentos inesperados em termos da linguagem detectada nas mensagens observadas na rede.

5.3. Infra-Estrutura

A partir dos grupos de campanhas de *spam*, verificamos de quantos endereços IP de origem e *autonomous systems* (AS) distintos as mensagens em cada campanha foram enviadas. O número de IPs e AS em cada caso são mostrados na Figura 5. Os resultados indicam padrões de comportamento distintos. Alguns *spammers* contam com mais de 200 máquinas para enviar suas mensagens e algumas campanhas se originam de mais de 50 sistemas autônomos, mas a cauda longa indica que uma grande parte das campanhas são enviadas de um grupo limitado de IPs. Esses resultados parecem indicar dois tipos de infra-estrutura: a centralizada, em que um pequeno conjunto de máquinas envia as mensagens (usualmente devido à contratação de serviços de *hosting* em um AS) e a descentralizada, em que máquinas residenciais podem estar sendo abusadas através de exploração de *proxies* abertos ou da implantação de *botnets*.

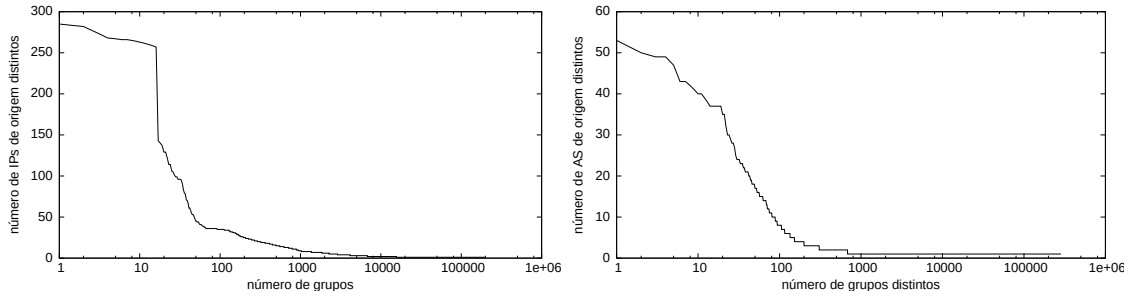

Figura 5. número de IPs e AS distintos para cada campanha de *spam*

6. Conclusão e Trabalhos Futuros

Embora filtros de *spam* consigam bloquear com relativa eficácia os *spams* que chegam à caixa-postal de usuários de correio eletrônico, eles não impedem que os preciosos recursos de banda e armazenamento da infra-estrutura da Internet sejam consumidos. A fim de subsidiar estudos que pretendam identificar e possivelmente bloquear o *spam* enquanto ele ainda trafega pela rede, caracterizamos algumas estratégias que definem padrões de comportamento de *spammers*. Para tal, utilizamos dados coletados em *honeypots* de baixa interatividade na Internet brasileira. Em seguida, detectamos os grupos de mensagens que diferem apenas por ofuscações de texto aleatórias e que correspondem ao mesmo *spam*. Para isso, características relevantes de cada mensagem foram extraídas e inseridas em uma

árvore de padrões freqüentes, que agrupa as mensagens que compartilham características freqüentes e se diferenciam apenas por uma característica infreqüente. A heurística permite detectar as diferentes campanhas de *spam*. Para essas campanhas, aplicamos técnicas de mineração de dados (agrupamento e regras de associação) para verificar as diferentes formas com que os grupos de *spam* exploram os recursos da rede.

Entre alguns dos padrões de comportamento encontrados, destacamos: (i) existem oito perfis distintos de abuso das portas das máquinas alvo, e, em cada um, diferentes portas são exploradas em diferentes freqüências; (ii) os relacionamentos entre país de origem, país de destino e idioma das mensagens indicaram a predominância de *spam* oriundo da China e Taiwan, escrito em chinês, mas ainda, as regras de associação geradas também revelam mensagens provenientes de IPs brasileiros escritas em idiomas como chinês e alemão; (iii) existem diferentes estratégias em relação à quantidade de máquinas utilizadas para disseminar as mensagens.

Como trabalhos futuros, pretendemos aperfeiçoar ainda mais o processo de unificação para agrupar mais mensagens e apresentar uma validação formal. Em seguida, pretendemos aplicar novas minerações de dados sobre os grupos correlacionando os padrões já encontrados, por exemplo, para verificar se o tipo de comportamento em relação às portas abusadas está relacionado com a origem do spam.

7. Agradecimentos

Este trabalho foi parcialmente financiado por CNPq, CAPES, FAPEMIG, FINEP e NIC.BR.

Referências

Cavnar, W. B. and Trenkle, J. M. (1994). N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, pages 161–175, Las Vegas, US.

Cerf, V. G. (2005). Spam, spim, and spit. *Commun. ACM*, 48(4):39–43.

Cook, D., Hartnett, J., Manderson, K., and Scanlan, J. (2006). Catching spam before it arrives: domain specific dynamic blacklists. In *ACSW Frontiers '06: Proceedings of the 2006 Australasian workshops on Grid computing and e-research*, pages 193–202, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

Cooke, E., Jahanian, F., and McPherson, D. (2005). The zombie roundup: understanding, detecting, and disrupting botnets. In *SRUTI'05: Proceedings of the Steps to Reducing Unwanted Traffic on the Internet on Steps to Reducing Unwanted Traffic on the Internet Workshop*, pages 6–6, Berkeley, CA, USA. USENIX Association.

Cranor, L. F. and LaMacchia, B. A. (1998). Spam! *Commun. ACM*, 41(8):74–83.

CYMRU (2007). `http://www.cymru.com/BGP/asnlookup.html`.

Gellens, R. and Klensin, J. (2006). RFC 4409: Message Submission for Mail. `http://www.ietf.org/rfc/rfc4409.txt`.

Gomes, L. H., Cazita, C., Almeida, J. M., Almeida, V., and Wagner Meira, J. (2007). Workload models of spam and legitimate e-mails. *Perform. Eval.*, 64(7-8):690–714.

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

Killalea, T. (2000). RFC 3013: Recommended Internet Service Provider Security Services and Procedures. http://www.ietf.org/rfc/rfc3013.txt.

Krawetz, N. (2004). Anti-honeypot technology. *IEEE Security & Privacy*, 2(1):76–79.

Li, F. and Hsieh, M.-H. (2006). An empirical study of clustering behavior of spammers and group-based anti-spam strategies. *Proceedings of the Third Conference on Email and Anti-Spam (CEAS). Mountain View, CA*.

Lindberg, G. (1999). RFC 2505: Anti-Spam Recommendations for SMTP MTAs. http://www.ietf.org/rfc/rfc2505.txt.

Messaging Anti-Abuse Working Group (MAAWG) (2005). Managing Port 25 for Residential or Dynamic IP Space. http://www.maawg.org/port25/MAAWG_Port25rec0511.pdf.

Messaging Anti-Abuse Working Group (MAAWG) (2007). Email Metrics Program: Report #5 – First Quarter 2007. http://www.maawg.org/about/MAAWG20071Q_Metrics_Report.pdf.

Milletary, J. (2005). Technical trends in phishing attacks. Technical report, CERT Coordination Center, Carnegie Mellon University. http://www.cert.org/archive/pdf/Phishing_trends.pdf.

Musat, C. N. (2006). Layout based spam filtering. *Transaction on Engineering, Computing and Technology*.

Provos, N. and Holz, T. (2007). *Virtual Honeypots: From Botnet Tracking to Intrusion Detection*. Addison-Wesley Professional, 1st edition. ISBN-13: 978-0321336323.

Pu, C. and Webb, S. (2006). Observed trends in spam construction techniques: A case study of spam evolution. *Proceedings of the Third Conference on Email and Anti-Spam (CEAS). Mountain View, CA*.

Ramachandran, A. and Feamster, N. (2006). Understanding the network-level behavior of spammers. In *SIGCOMM '06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 291–302, New York, NY, USA. ACM.

Sophos.com (2004). The Spam Economy: The Convergent Spam and Virus Threats. August 2004. http://www.sophos.com/whitepapers/Sophos_spam-economy_wpus.pdf.

SpamAssassin (2007). http://spamassassin.apache.org.

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

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

Yeh, C.-C. and Lin, C.-H. (2006). Near-duplicate mail detection based on url information for spam filtering. In *Information Networking. Advances in Data Communications and Wireless Networks*, pages 842–851. Springer Berlin / Heidelberg.