Caracterização Temporal de Estratégias de Disseminação de Spam*

Luam C. Totti[1], Rubens E. A. Moreira[1], Elverton Fazzion[1], Osvaldo Fonseca[1], Wagner Meira Jr.[1], Dorgival Guedes[1], Cristine Hoepers[2], Klaus Steding-Jessen[2], Marcelo H. P. Chaves[2]

[1] Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

[2]CERT.br - Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança
NIC.br - Núcleo de Informação e Coordenação do Ponto BR

{luamct,rubens,elverton,osvaldo.morais,meira,dorgival}@dcc.ufmg.br

{cristine,klaus,mhp}@cert.br

Abstract. *In this work we present a characterization of spam campaign dissemination strategies, emphasizing their temporal aspects. Our analysis is based on spam campaigns detected in 1.1 billion messages collected in eight countries using low interactivity honeypots during a three-month period. We present a temporal analysis of the messages grouped both by campaigns and IP addresses, observing aspects such as message inter-arrival times, usage bursts, and inactivity periods. We also discuss how the observed behaviors can be used in the design and evaluation of spam prevention systems.*

Resumo. *Neste trabalho apresentamos uma caracterização de estratégias empregadas na disseminação de campanhas de spam, com enfoque em aspectos temporais. Nossas análises se baseiam nas campanhas de spam detectadas a partir de 1,1 bilhão de mensagens de spam capturadas em 2011, durante 3 meses, por honeypots de baixa interatividade instalados em oito países. Apresentamos análises temporais das mensagens agrupadas, tanto com base nos atributos que definem o conteúdo das mensagens, quanto pelos IPs de origem destas, observando, por exemplo, intervalos de envio das mensagens, comportamento de rajadas e períodos de inatividade. Por fim, discutimos como os comportamentos observados podem auxiliar no projeto e avaliação de sistemas de prevenção de spam.*

1. Introdução

O impacto da atuação de *spammers* na Internet é um problema conhecido por usuários e administradores de rede. Durante as duas últimas décadas, essa atuação tem evoluído em volume, conteúdo e sofisticação no envio das mensagens. Inúmeros trabalhos estimam os altos custos associados a essa prática, predominantemente arcados pelos destinatários. Provedores de serviços de correio eletrônico estimam que entre 40% e 80% de todas mensagens eletrônicas trafegadas são consideradas *spam*.

*Este trabalho foi parcialmente financiado por CNPq, Fapemig e NIC.BR.

Na tentativa de reduzir o impacto dessas práticas, diversas técnicas de detecção e filtragem de conteúdo malicioso foram propostas. Entretanto a constante inovação, tanto do conteúdo, quanto das formas de disseminação de *spam*, exigem um contínuo esforço de compreensão e combate.

Entender como as campanhas de *spam* são organizadas a partir de suas origens, como os recursos de rede são abusados e como elas evoluem ao longo do tempo pode contribuir para o desenvolvimento de novas ferramentas e técnicas que combatam a disseminação dessas mensagens. Tal conhecimento pode permitir identificar padrões de transmissão de mensagens indesejadas ainda durante o envio, prevenindo a entrega do conteúdo e evitando que recursos sejam desperdiçados. Além disso, entender os padrões de transmissão envolvidos provê aos administradores da rede a capacidade de determinar e atuar sobre os pontos de origem da disseminação.

Ao analisar o comportamento dos *spammers* ao longo do tempo, é possível identificar uma série de padrões que possam ser usados como alertas para comportamentos duvidosos na rede, ou ainda detectar e impedir o desenvolvimento de campanhas de *spam*. Como a disseminação de mensagens frequentemente utiliza redes de máquinas comprometidas por código malicioso (as *botnets*), caracterizar temporalmente o envio de *spam* e analisar os padrões de tráfego pode permitir a identificação de tais redes [Xie et al. 2008]. Além disso, torna-se possível entender se campanhas de *spam* evoluem ao longo do tempo de acordo com algum padrão comportamental que possa ser derivado e utilizado para identificá-las e, possivelmente, neutralizá-las.

Tendo isso em mente, este trabalho analisa a evolução temporal de campanhas de *spam* e a distribuição das máquinas transmissoras durante esse processo. A identificação de campanhas é baseada no uso de FPCluster, uma técnica desenvolvida em trabalhos anteriores [Calais et al. 2008], porém evoluída em alguns aspectos para melhorar seu desempenho nas condições de carga atuais. Essa técnica permite a identificação de mensagens que compartilham características similares (formato interno, palavras chave, URLs similares), mesmo quando o autor da campanha utiliza recursos de ofuscação para evitar o bloqueio de mensagens por filtros anti-*spam*. Parte-se do princípio de que, para comunicar o sentido da mensagem, o transmissor deve manter algum padrão na estrutura dos *e-mails* gerados, tanto por serem gerados automaticamente, quanto para que o objeto alvo da divulgação não seja ofuscado.

De posse dos conjuntos de mensagens que compartilham elementos comuns em termos de estrutura e conteúdo, foram identificados os endereços de origem que atuaram na transmissão de cada campanha, determinando, assim, os transmissores. Foi então analisada a evolução das campanhas ao longo do tempo e a participação dos transmissores em cada campanha durante o período observado. Considerou-se para as análises um conjunto de mais de um bilhão de mensagens de *spam*, coletadas por *honeypots* de baixa interatividade, configurados para simular o comportamento de *proxies* e *mail relays* abertos [Steding-Jessen et al. 2008]. As mensagens, coletadas por três meses, foram processadas utilizando técnicas de mineração de dados para extrair as características principais e identificar os agrupamentos existentes.

Nossos resultados mostram que as mensagens de uma campanha de *spam* são distribuídas na forma de rajadas ao longo do tempo, com períodos de atividade relativa-

mente curtos, seguidos por longos períodos de inatividade. Observou-se que, apesar de uma mesma campanha se manifestar através de múltiplos endereços, a maioria das campanhas se restringe a um pequeno número de IP's distintos, que podem ser mapeados a uma mesma *botnet* e futuramente monitorados. O perfil em rajadas do envio de mensagens de spam respalda a idéia de que métodos baseados no monitoramento de aspectos da camada de transporte podem ser utilizadas com sucesso para identificar comportamentos maliciosos. Por fim, discute-se a viabilidade de se utilizar o conceito de campanhas no processo de detecção e prevenção de *spam*.

2. Metodologia

Nesta seção são detalhados dois elementos principais da metodologia do trabalho, a estrutura de coleta das mensagens de *spam* e o algoritmo FPCluster de obtenção de campanhas.

2.1. Coleta dos dados

A captura das mensagens de *spam* consideradas neste trabalho foi realizada utilizando *honeypots* de baixa interatividade. Duas máquinas foram instaladas em redes brasileiras e outras sete nos seguintes países: Alemanha, Holanda, China, Estados Unidos, Colômbia, Coréia e Japão. A implantação de *honeypots* em diferentes países permite uma visão global do problema do *spam*, livre de distorções comuns em coletas restritas a uma única localidade. Os *honeypots* foram configurados de modo a simular *proxies* e *mail relays* abertos, componentes tipicamente abusados por *spammers* [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. Para fins de padronização, utilizou-se GMT como fuso-horário em todas as mensagens coletadas. Maiores detalhes sobre a infra-estrutura de coleta podem ser encontrados em trabalhos anteriories do grupo [Steding-Jessen et al. 2008].

Foram utilizados dados do dia 01 de agosto até o dia 31 de outubro de 2011. A Tabela 1 resume algumas características dos dados utilizados. Como será discutido nas seções seguintes, as campanhas foram inicialmente identificadas de forma isolada para cada dia coletado. Posteriormente, um algoritmo de fusão foi aplicado para se agrupar campanhas que se estenderam por vários dias. Nota-se que a estratégia de agrupamento de campanhas de dias distintos reduz consideralvemente o número de campanhas.

2.2. FPCluster

As análises apresentadas neste trabalho se baseiam no conceito de campanha de *spam*. Denomina-se campanha de *spam* um conjunto de mensagens que possuam um objetivo

Total de mensagens	1.083.699.945
Média de mensagens por dia	11.779.347
Volume de dados (TB)	1,36
IPs únicos	87.237
Campanhas antes do agrupamento	339.647
Campanhas após o agrupamento	107.656
Dias considerados	92

Tabela 1. Estatísticas gerais dos dados processados.

comum e adotem uma mesma estratégia de disseminação. Porém, devido ao uso de técnicas de ofuscação, como a inserção de trechos textuais aleatórios, agrupá-las pode se tornar impraticável utilizando métodos convencionais. Por essa razão utilizou-se neste trabalho o FPCluster, um algoritmo de agrupamento proposto e avaliado em trabalhos anteriores [Pires et al. 2011] que objetiva identificar tais estratégias.

O algoritmo organiza as mensagens e identifica os atributos invariantes das mesmas utilizando a estrutura hierárquica de uma árvore de padrões frequentes (FP-Tree) [Han et al. 2000]. O processo pode ser dividido em três etapas principais: a extração e pré-processamento dos atributos (*features*), a construção da árvore de padrões frequentes e a extração dos agrupamentos de mensagens.

Na primeira etapa extrai-se das mensagens os atributos considerados relevantes. Adotamos neste trabalho os seguintes atributos: *From, Organization, Received, Content-Transfer-Encoding, Content-Type, Date, Mime-Version, To, Subject, Reply-To, X-Mailer, X-Mail-From, Layout e Language*, sendo os dois últimos derivados do corpo da mensagem.

Cada mensagem é então representada por seus atributos, ordenados por sua frequência global na base de dados. Em seguida os atributos de cada mensagem são inseridos em uma FP-Tree. Uma mensagem é portanto representada por um caminhamento na árvore. Devido à ordem de inserção dos atributos, os mais frequentes se encontram nos níveis mais altos da árvore. A Figura 1 mostra uma parte de uma FP-Tree gerada. As cores simbolizam diferentes atributos e o diâmetro dos nodos a frequência do valor do atributo em questão. Os nodos com um grande número de filhos sugerem o uso de ofuscação de conteúdo pelos *spammers*, pois atributos com muitos valores distintos frequentemente são frutos de inserção de trechos aleatórios.

Como exemplo, considere um grande conjunto de mensagens com os atributos *From, Content-Type, To, Reply-To, X-Mailer, X-Mail-From, Layout e Language* possuindo valores idênticos entre as mensagens. Essas mensagens serão representadas por um único caminhamento na árvore, onde cada nó é um par $< Tipo, Valor >$ representando um tipo de atributo e seu valor. Porém se o campo *Subject* dessas mensagens é do formato *"Best Buy Viagra Online [random]"*, onde *random* é um trecho aleatório de texto único para cada mensagem, haverá um grande número de filhos nesse ponto, indicando que esse grupo de mensagens pertence a uma mesma campanha de *spam*.

Uma vez construída a árvore, executa-se o algoritmo de corte para se identificar as campanhas. O algoritmo baseia-se no tamanho das sub-árvores para definir o ponto de corte. Define-se a frequência f_n de um nodo n como o número de mensagens contidas na sub-árvore em que n é a raiz.

Como os atributos são inseridos em ordem decrescente de frequência, os caminhamentos mais próximos da raiz apresentam os atributos mais comuns, ou seja, que foram observados em um número maior de mensagens. Quanto mais se caminha na árvore na direção das suas folhas, menos frequentes são os atributos e mais específicas são as mensagens que passam por aqueles nodos. A etapa de corte basicamente estabelece um parâmetro de frequência f_c e define que qualquer nodo que apresentar somente sub-árvores com menos que f_c mensagens é um ponto de corte e portanto define uma campanha. O objetivo é maximizar o número de atributos comuns entre as mensagens

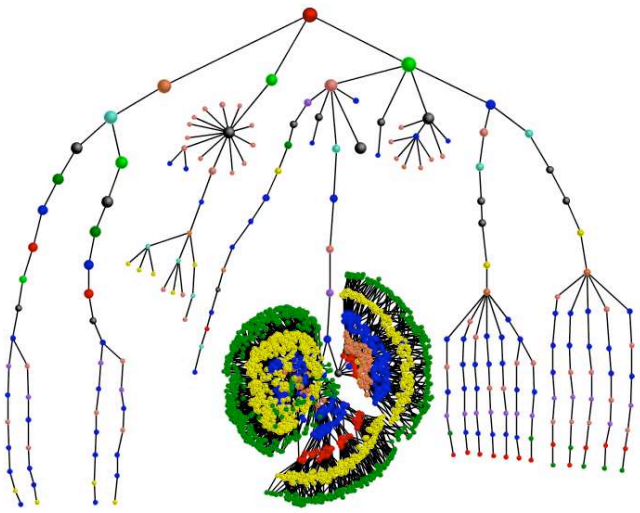

Figura 1. FPTree gerada com parte das mensagens processadas.

sem que agrupemos todas em somente uma campanha, e ao mesmo tempo queremos caminhar na direção das folhas apenas o suficiente de modo a não segmentar mensagens de uma mesma campanha.

O principal objetivo da estratégia adotada é minimizar o impacto das técnicas de ofuscação adotadas pelos *spammers*, que alteram sistematicamente o conteúdo de suas mensagens para torná-las únicas. Ao não exigir a definição de uma métrica de similaridade, o FPCluster se mostra mais flexível frente às frequentes evoluções no comportamento dos *spammers*. Métodos convencionais de agrupamento não levam em consideração tais estratégias de ofuscação e portanto não apresentam bom desempenho nesses cenários. Maiores detalhes sobre essa técnica podem ser encontrados em trabalhos anteriores [Calais et al. 2008, Pires et al. 2011].

2.3. Fusão de Campanhas

As mensagens capturadas são agregadas em um servidor de armazenamento a cada 24 horas. Elas são então disponibilizadas através de NFS em um servidor de processamento responsável por extrair estatísticas e outras informações relevantes dos dados. A fim de se ter informações atualizadas do comportamento dos *spammers*, sempre que novas mensagens são disponibilizadas, novas campanhas são geradas. Entretanto, tais campanhas se limitam ao dia sendo processado, devido à periodicidade adotada para a coleta.

Para se obter campanhas que se estendam por múltiplos dias foi definido um processo simples de fusão de campanhas. Todas campanhas definidas por um mesmo conjunto de atributos, independente da ordem dos mesmos, são agrupadas como sendo uma mesma campanha. Como mostrado na Tabela 1, esse procedimento reduziu o número de campanhas em 69%. Utilizou-se a granulação de um dia por esse ser o intervalo mais natural para o processo de captura estabelecido.

Uma abordagem alternativa seria gerar somente uma árvore de padrões frequentes utilizando os dados de todo o período desejado, porém tal abordagem apresenta uma série de desvantagens. Primeiramente o volume de dados de entrada seria da ordem de Teraby-

tes, o que pode tornar seu tratamento inviável, mesmo o algoritmo adotado possuindo uma estratégia *out-of-core*. Uma execução única também inviabiliza a oportunidade de paralelismo ao se executar diferentes dias de forma concorrente. Além disso, para cada intervalo de tempo de interesse uma árvore diferente deve ser gerada, enquanto a abordagem adotada exige apenas a computação de uma árvore por dia seguida de um processo eficiente de fusão de campanhas entre múltiplos dias. Por fim, a tarefa de se obter um parâmetro de corte ideal para a detecção de campanhas seria consideravelmente mais difícil, uma vez que as árvores de entrada apresentariam dimensões extremamente variadas, desde aproximadamente 12 milhões de mensagens (1 dia) até mais de 1 bilhão de mensagens (92 dias).

3. Resultados

Considerando a proposta deste trabalho, os resultados são apresentados a seguir, primeiramente para uma análise das campanhas de *spam* observadas, em seguida para as rajadas identificadas dentro de cada campanha e, por fim, para o comportamento dos transmissores (identificados por endereços IP).

3.1. Campanhas

Após introduzido o novo critério de detecção campanhas, é interessante entender o reflexo da variação do parâmetro de corte f nas campanhas geradas. Seja $N(f) = |C_f|$ onde C_f é o conjunto de campanhas geradas utilizando o parâmetro de corte f. Temos portanto que $N(f)$ descreve o número de campanhas geradas utilizando-se o parâmetro f. A Figura 2 mostra sua distribuição acumulada. Cada ponto é obtido pela média dos valores de $N(f)$ em 10 dias não sequenciais do mês de outubro, assim como o desvio padrão desses valores. Observa-se que à medida que f cresce, o número de campanhas tende a se estabilizar em valores baixos.

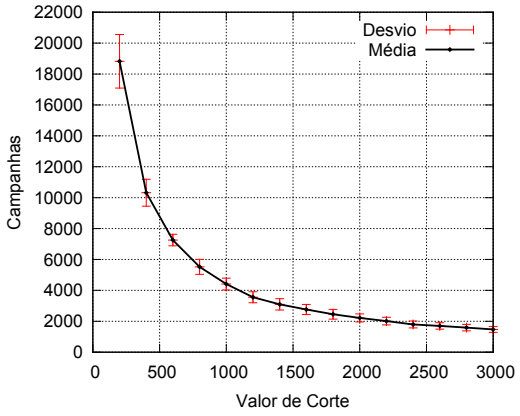

Figura 2. Número médio de campanhas geradas variando-se o parâmetro f de corte.

Seja $T(c)$ o tamanho em número de mensagens de uma campanha c e C_{1000} o conjunto das campanhas geradas utilizando-se $f = 1000$, as Figuras 3(a) e 3(b) mostram, respectivamente, a distribuição de frequência e a distribuição acumulada de $T(c_i)$ para $c_i \in C_{1000}$. Observa-se que o procedimento de corte não impede que as campanhas

tenham menos de f elementos. Cerca de 2% das campanhas possuem tamanhos inferiores a 1000, enquanto 10% possuem tamanhos maiores que 10000. Porém, a grande concentração de valores, cerca de 50%, reside no intervalo de 1000 até 10000. Apesar da concentração acentuada nesse intervalo, uma investigação subsequente mostrou que apenas 30% das mensagens totais pertencem a esse intervalo.

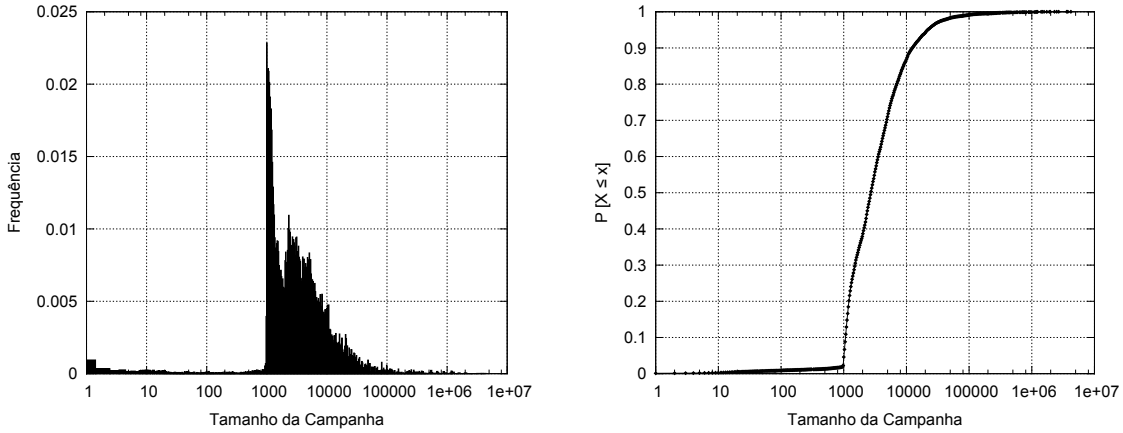

Figura 3. Distribuição de frequência dos valores de tamanhos das campanhas em números de mensagens para o valor de corte adotado $f = 1000$ (esquerda); distribuição acumulada de frequência de tamanhos das campanhas (direita).

Em relação aos aspectos temporais, seja C_i uma campanha definida pelas mensagens $m_0, m_1, ..., m_n$ e $min_t(C)$ e $max_t(C)$ funções que retornam respectivamente o momento de envio da primeira e da última mensagem em C. Definimos a duração $D(C)$ de uma campanha como $D(C) = max_t(C) - min_t(C)$, ou seja, a diferença entre a mensagem mais antiga e a mais recente presentes na campanha. A Figura 4(a) mostra a curva de distribuição acumulada da função D para as campanhas identificadas no período. A curva *Total* corresponde à distribuição das durações de todas as campanhas, enquanto as curvas tracejadas correspondem à distribuição dos grupos das 5% maiores campanhas de acordo com dois critérios distintos, número de mensagens e total de *bytes* transferidos.

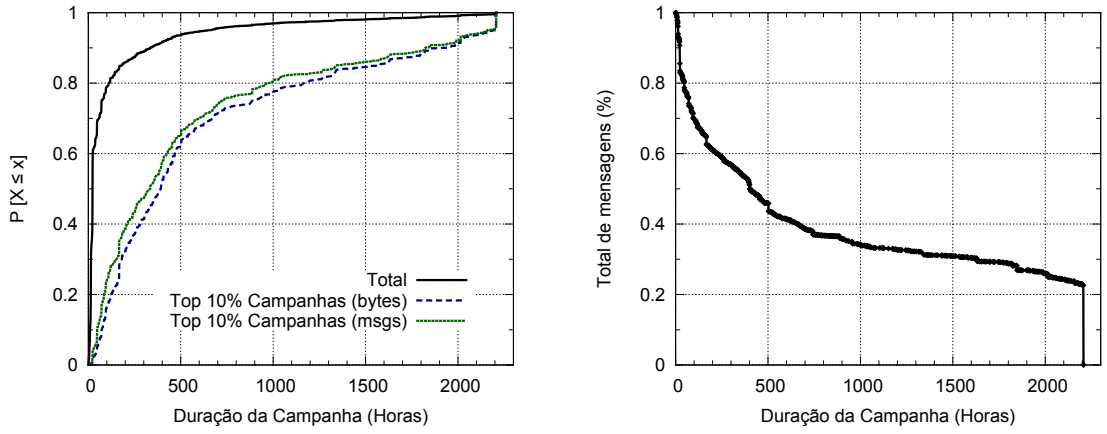

Figura 4. Distribuição acumulada das durações das campanhas

Observa-se que 80% das campanhas duraram menos de 100 horas. Nas curvas tracejadas observa-se um número grande de campanhas com o tamanho máximo possível

nesse intervalo. Isso é um indicativo de que essas campanhas possivelmente se estenderiam para além do intervalo avaliado. Verificou-se uma alta correlação positiva entre $T(c)$, o tamanho em número de mensagens de uma campanha, e $D(c)$, sua duração em horas, de acordo com o coeficiente de Pearson ($r = 0.993$). Esse comportamento é esperado, implicando que as campanhas que entregam um volume maior de mensagens permanecem ativas por mais tempo.

Seja M o número total de mensagens enviadas no período estudado, definimos $E(h) = \sum_{c_i \in C} |c_i|/M$ tal que $D(c_i) \geq h$, ou seja, $E(h)$ é a porcentagem de mensagens pertencentes a campanhas com duração igual ou maior que h horas. A Figura 4(b) mostra a curva definida por $E(h)$ para o intervalo analisado. Observa-se que apesar de 80% das campanhas durarem menos que 100 horas, mais de 70% das mensagens analisadas são entregues por campanhas com mais de 100 horas de duração. A redução abrupta no valor 2208 é consequência da grande concentração de mensagens pertencentes a campanhas com o tamanho máximo possível no intervalo, como mostrado nas curvas tracejadas da Figura 4(a).

Essa evidência sugere que abordagens *anti-spam* baseadas em detecção de campanhas podem ser bem efetivas. Se um método é capaz de detectar qualquer campanha d horas a partir de sua primeira manifestação, de acordo com $E(h)$ o método poderá evitar a entrega de um volume considerável de conteúdo. Se $d = 24$, o método terá identificado campanhas contendo aproximadamente 83% das mensagens no nosso conjunto de dados, podendo bloquear todas as mensagens ainda não entregues. Para estimar com maior precisão a porcentagem de mensagens que podem ser bloqueadas é preciso entender o comportamento de envio de mensagens dentro de cada campanha. Se as mensagens são enviadas preferencialmente no início da vida da campanha, a prevenção será menos efetiva, pois um volume maior de conteúdo malicioso já terá sido entregue quando a campanha for identificada. Por outro lado, se a entrega se concentra no final da vida da campanha, pode ocorrer do método não possuir mensagens suficientes para identificar uma campanha.

A Figura 5(a) mostra a relação entre a porcentagem média de mensagens enviadas por uma campanha e cada instante da vida dessa campanha. Ou seja, observando-se um grupo suficientemente grande de campanhas, no meio da vida de cada campanha, elas terão enviado metade de todo seu conteúdo. A distribuição observada é predominantemente homogênea, sugerindo que não existe concentração de mensagens em nenhum instante específico das campanhas.

Entretanto, isso não significa que as mensagens sejam transmitidas a uma taxa constante durante a vida da campanha, mas sim que não há momentos preferenciais para esse envio. A Figura 5(b) ilustra esse ponto ao exibir o perfil de envio em uma granulação de mensagem de 20 campanhas aleatoriamente escolhidas. Cada faixa horizontal do gráfico corresponde a uma campanha e cada traço vertical ao longo de uma faixa simboliza uma mensagem da campanha enviada naquele instante. O comportamento de diversas campanhas, como C_2, C_3, C_{10} e C_{20}, mostram que as transmissões podem ocorrer em momentos bem separados, de duração variada. Em casos como o da campanha C_{10}, entretanto, o padrão pode ser mais variado, com intervalos mais intensos e outros com mensagens mais espaçadas, mas ainda relativamente próximas.

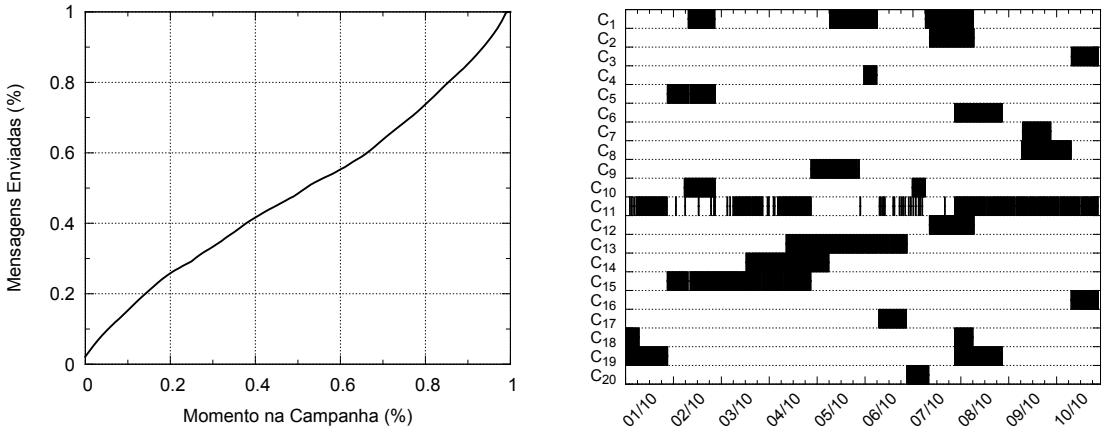

Figura 5. Perfil de envio das mensagens dentro das campanhas.

3.2. Rajadas

A Figura 5(b) sugere fortemente que o envio de mensagens por *spammers* é concentrado em sessões bem definidas. Denominaremos essas sessões rajadas. Estabelecendo-se um valor IAT_{max} como o intervalo máximo entre duas mensagens de uma mesma rajada, podemos identificar cada rajada dentro de uma campanha.

A Figura 6(a) mostra a distribuição acumulada dos valores dos intervalos entre mensagens (IAT). O comportamento da curva indica que valores na cauda longa são possivelmente bons parâmetros para detecção de rajadas. A Figura 6(b) exibe a relação entre o intervalo máximo entre mensagens de uma mesma rajada e o número de rajadas identificadas utilizando-se esse valor. A partir de valores de IAT próximos a 1000 nota-se que a variação no número de campanhas é muito baixa, portanto adotaremos $IAT_{max} = 2000$ nas análises seguintes.

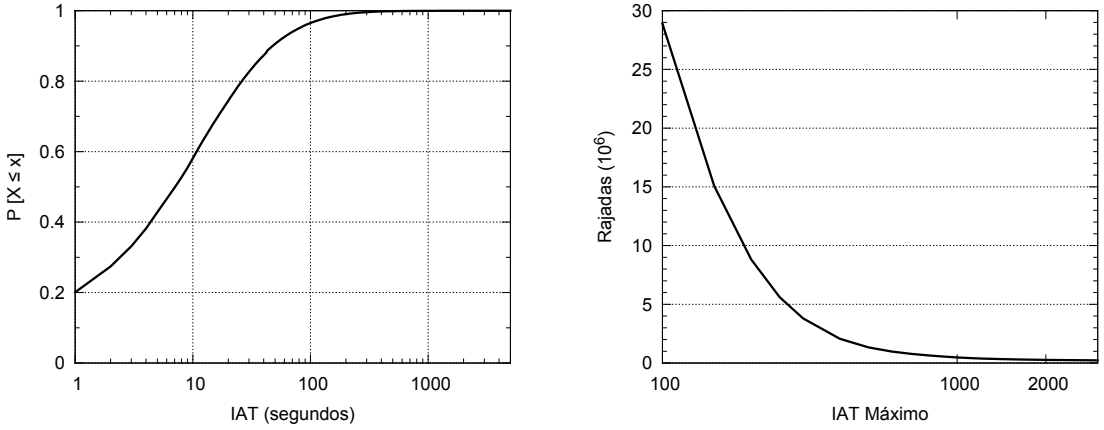

Figura 6. Perfil de envio das mensagens dentro das campanhas.

As Figuras 7(a) e 7(b) exibem respectivamente a distribuição do número de mensagens em cada rajada e o número de rajadas por campanha. A mudança na taxa de crescimento da curva na Figura 7(a) ocorre devido ao maior número de campanhas com número de mensagens superior a 1000, fato esse provocado pela escolha do parâmetro de

corte $f = 1000$. A Figura 7(b) mostra que campanhas são tipicamente compostas por poucas rajadas, sendo que aproximadamente 67% concentram suas mensagens em apenas uma rajada e apenas cerca de 6% das campanhas têm mais que 5 rajadas.

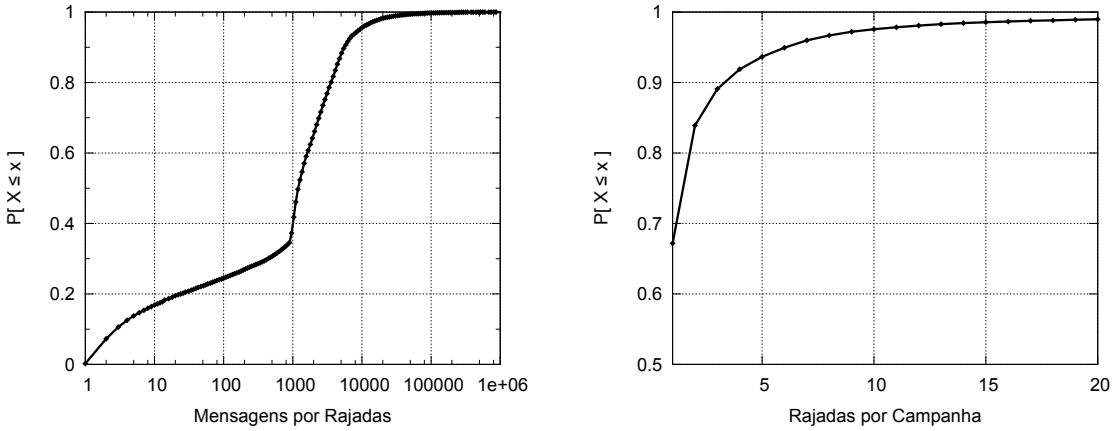

Figura 7. Distribuição acumulada dos números de mensagens em cada rajada e o número de rajadas por campanha.

As Figuras 9(a) e 8(b) exibem a distribuição dos períodos de inatividade das campanhas, que correspondem basicamente aos intervalos entre uma rajada e a subsequente. Aproximadamente 47% dos períodos de inatividade em campanhas duram menos de 100 minutos, mas há claramente campanhas com intervalos de inatividade mais longos. Os gráficos indicam que há um comportamento periódico associado à noção de dias: cerca de 8% dos períodos de inatividade se concentram ao redor de 1440 minutos (um dia). Os picos subsequentes correspondem a múltiplos dias.

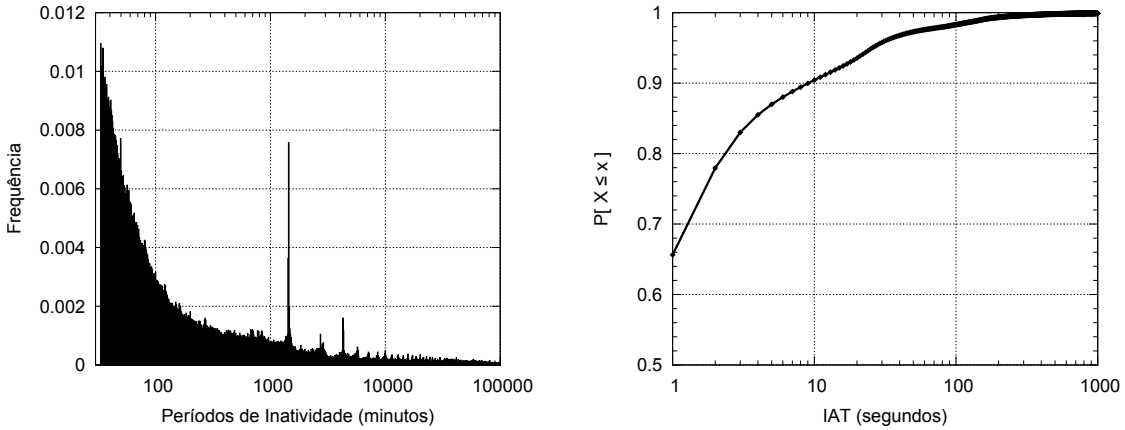

Figura 8. Distribuição dos períodos de inatividade entre as sessões de envio de um mesmo IP e distribuição acumulada dos intervalos entre mensagens (IAT) para cada endereço IP.

3.3. IP's de Origem

Ao longo do período de coleta observamos que as mensagens originaram de 87.237 IP's distintos. Nessa seção apresentamos análises das relações desses IP's com as campanhas identificadas e outros aspectos temporais relevantes.

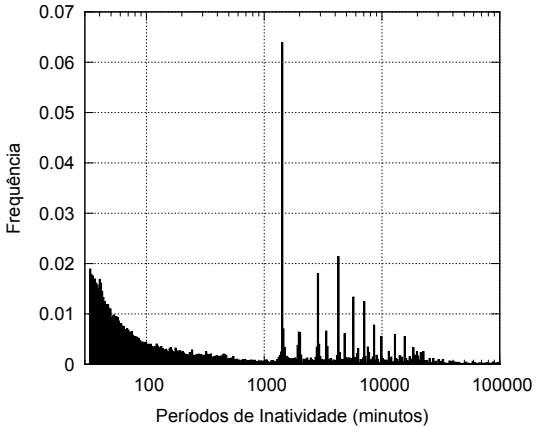

Figura 9. Distribuição dos períodos de inatividade das campanhas.

Analogamente ao caso das campanhas, a Figura 8(b) mostra a distribuição acumulada de IAT's relativo a cada IP, apresentando valores consideralvemente menores. Quase 66% dos intervalos entre mensagens enviadas por um mesmo endereço são menores que um segundo, o que confirma a intuição de que máquinas infectadas ou contratadas com o intuito de entregar *spam* mantêm altas taxas de envio de mensagens.

Observa-se pela Figura 10(a) que a maioria das campanhas utiliza poucos endereços IP para entregar suas mensagens. Cerca de 50% das campanhas utiliza 1 ou 2 endereços e 90% utiliza menos de 10 endereços distintos. Como será melhor discutido na seção 4, isso também é evidenciado pela curva da Figura 10(b), que mostra a distribuição do número de mensagens enviadas por cada IP.

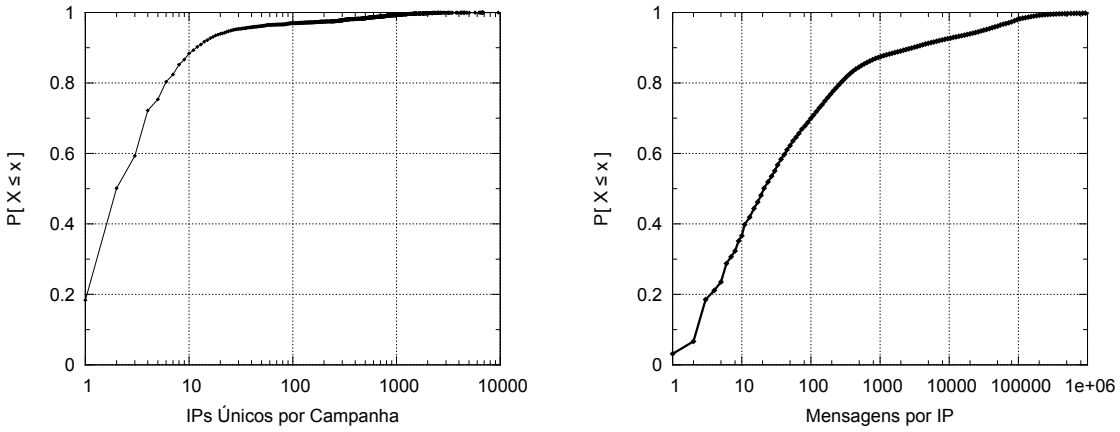

Figura 10. Distribuição acumulada dos endereços IP únicos por campanha e do número de mensagens enviadas por cada endereço IP.

A Figura 8(b) mostra a distribuição acumulada dos intervalos de entrega entre mensagens de um mesmo endereço IP e respalda a idéia de que as mensagens estão mais concentradas em poucas rajadas, pois a distribuição exibe valores menores de IAT's do que os observados para as campanhas. Definindo-se por período de inatividade qualquer intervalo entre mensagens maior que T_{max}, a Figura 9(a) também reforça essa hipótese mostrando a distribuição dos intervalos de inatividade. Aproximadamente 90%

dos períodos são menores do que 1000 minutos, enquanto para campanhas, apenas 65% dos períodos são menores do que esse mesmo valor. Isso sugere que um mesmo IP transmita mensagens de mais de uma campanha ao longo do tempo, o que justificaria os IAT's menores nesse caso.

Além disso, os intervalos de inatividade de um mesmo endereço IP não apresentam discontinuidades tão marcadas por um padrão de dias, ao contrário do que ocorre com campanhas (o pico em 1440 minutos é bem menos pronunciado na fig. 8(a) que na Figura 9(a)). Isso confirma que o comportamento das campanhas é marcado por fatores externos, já que os transmissores (endereços IP) não apresentam o mesmo padrão de inatividade.

4. Discussão

A análise do gráfico da Figura 4(a), com foco na curva que representa 10% das campanhas que mais enviam bytes, mostra que a maior parte destas (aproximadamente 98%), duram mais que 24 horas. Já a análise do gráfico da Figura 6(a), referente ao total de mensagens analisadas, mostra uma relação linear com o tempo de existência das campanhas.

Suponha que, em um pior caso, se todas as campanhas tivessem o mesmo tamanho em total de bytes enviados, 10% das mensagens representariam 10% do total de bytes enviados. Assim, neste caso, garantindo a identificação e interrupção das campanhas em até 24 horas, aproximadamente 9,8% dos bytes não seriam propagados. Por outro lado, se consideramos que a distribuição de bytes está concentrada em 10% das campanhas que geram mais volume de dados, bloqueá-las em até 24 horas seria o equivalente a evitar o envio da maior parte dos bytes envolvidos nesta análise.

5. Trabalhos Relacionados

O uso de *honeypots* se estabeleceu como um método eficaz para estudo e prevenção de ataques em rede [Provos and Holz 2007]. O *framework* para coleta de *spam* utilizado tem origem no agrupamento de diversos conceitos e metodologias que foram sendo aperfeiçoados na literatura e no desenvolvimento interno ao próprio grupo de pesquisa[1]. O coletor usado na pesquisa é uma atualização do projeto apresentado por Steding-Jessen *et al.* [Steding-Jessen et al. 2008], mantido pelo CERT.br, com vários aperfeiçoamentos na técnica e atualizações da estrutura.

Em relação ao estudo do comportamento de *spammers*, inúmeros trabalhos de caracterização e análise de conteúdo malicioso foram propostos na literatura, constituindo um conhecimento agregado valioso no projeto e aperfeiçoamento de técnicas de controle dessas práticas.

Em [Bertolotti and Calzarossa 2001] são apresentadas caracterizações de múltiplos conjuntos de emails, incluindo análises dos tamanhos das mensagens, número de recipientes e também um estudo focado no comportamento dos usuários. Entretanto os aspectos temporais explorados são limitados e os dados utilizados incluem ambos conteúdo *spam* e legítimo. O trabalho apresentado busca identificar assinaturas comportamentais maliciosas ao analisar conteúdo restritamente *spam*, garantido pelo processo de captura. Gomes *et al.* [Gomes et al. 2004] por sua vez fornece uma caracterização em

[1]O projeto *Spammining*, parceria entre o CERT.br e o Departamento de Ciência da Computação da UFMG.

conteúdo exclusivamente malicioso, porém limita suas análises temporais aos perfis de tráfego observáveis diariamente e à localidade temporal dos destinatários. Nesse trabalho observou-se como campanhas de *spam* podem se propagar por períodos longos e como esse conhecimento pode ser usado efetivamente em seu combate.

Inúmeros trabalhos exploram unicamente características de rede do tráfego de *spam*, parcialmente devido à dificuldade inerente de se tratar as ofuscações presentes nas mensagens. Ramachandran *et al.* investigam características de tráfego, coletadas da camada de rede, tais como a persistência de endereços IP e de rotas e características específicas de *botnets*, que sejam comuns a *spammers* [Ramachandran and Feamster 2006], enquanto Guerra *et al.* analisam os padrões de comunicação presentes em uma campanha de *spam* [Guerra et al. 2009]. Através da análise de características de fluxos de pacotes SMTP, Sperotto *et al.* propõem um algoritmo para detecção de *spams* utilizando apenas informações da camada de rede, como tempo de inatividade e quantidade de picos no fluxos de pacotes [Sperotto et al. 2009]. Os autores utilizam dados da rede de uma universidade holandesa em conjunto com informações de lista de bloqueio de DNS (*blacklists*) para validar o algoritmo proposto.

Diferente dos trabalhos citados, foram utilizados neste trabalho principalmente atributos associados ao conteúdo das mensagens para se obter campanhas, e em seguida os comportamentos observados nas campanhas foram contrastados com aqueles obtidos por agrupamentos por IP, que por sua vez é um atributo de rede. O estudo de mensagens *spams* agrupados por campanhas permite a observação de padrões comportamentais relevantes, muitas vezes imperceptíveis olhando-se todo o contigente de mensagens.

Análises de mensagens eletrônicas maliciosas na verdade constituem apenas uma pequena parcela dos trabalhos de caracterização disponíveis na literatura. Dentre os diversos comportamentos estudados incluiu-se *spam* na *Web*, spam em conteúdo multimídia e, mais recentemente, *spam* e *phishing* em redes sociais. Neste trabalho inclusive explorase o conceito de sessão de usuários proposto por Veloso *et al.*, utilizado originalmente no estudo de acesso de conteúdo multimídia, adaptado ao contexto de identificação de rajadas [Veloso et al. 2006].

6. Conclusão e trabalhos futuros

Entender o comportamento de *spammers* ao longo do tempo pode auxiliar no desenvolvimento de técnicas para melhor identificá-los e bloquear esse tipo de tráfego. Neste trabalho utilizamos o algoritmo FPCluster para agrupar mensagens que podem ser identificadas como uma única campanha de *spam* e assim avaliar a sua evolução temporal, bem como o comportamento dos endereços IP vistos em associação com cada campanha.

Nossos resultados indicam que o comportamento das campanhas ao longo do tempo é marcado por transmissões em rajadas, muitas vezes separadas por dias. Por outro lado, apesar dos transmissores (endereços IP) também operarem em rajadas, elas são mais intensas e com períodos de inatividade menos associados à noção de dias. Isso sugere que os transmissores (endereços IP) transmitem mensagens de várias campanhas durante um certo período e que a duração de campanhas é determinada por fatores externos que observam a gradação de dias. Sendo assim, técnicas para previsão de duração de campanhas não podem ser baseadas na observação dos seus padrões de tráfego apenas, mas talvez os padrões de comportamento de transmissores possam ser melhor modela-

dos. Aprofundar nossa análise nesses padrões é uma das linhas de trabalhos futuros que pretendemos seguir.

Referências

Bertolotti, L. and Calzarossa, M. C. (2001). Models of mail server workloads. *Performance Evaluation*, 46:65 – 76. Advanced Performance Modeling.

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

Gomes, L. H., Cazita, C., Almeida, J. M., Almeida, V., and Meira, Jr., W. (2004). Characterizing a spam traffic. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, IMC '04, pages 356–369, New York, NY, USA. ACM.

Guerra, P. H. C., Pires, D. E. V., Guedes, D., Jr., W. M., Hoepers, C., Steding-Jessen, K., and Chaves, M. (2009). Caracterização de Encadeamento de Conexões para Envio de Spams. In *Anais do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, Recife, Brasil.

Han, J., Pei, J., and Yin, Y. (2000). Mining frequent patterns without candidate generation. In Chen, W., Naughton, J. F., and Bernstein, P. A., editors, *SIGMOD Conference*, pages 1–12. ACM.

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

Pires, D. E. V., Totti, L. C., Moreira, R. E. A., Fazzion, E., Fonseca, O., Jr., W. M., de Melo-Minardi, R. C., and Guedes, D. (2011). Fpcluster: Uma estratégia eficiente de agrupamento out-of-core sem medida de similaridade. In *Anais do Simpósio Brasileiro de Bancos de Dados (SBBD)*, pages 1–6. SBC.

Provos, N. and Holz, T. (2007). *Virtual Honeypots: From Botnet Tracking to Intrusion Detection*. Addison-Wesley Professional.

Ramachandran, A. and Feamster, N. (2006). Understanding the Network-Level Behavior of Spammers. *SIGCOMM Comput. Commun. Rev.*, 36(4):291–302.

Sperotto, A., Vliek, G., Sadre, R., and Pras, A. (2009). Detecting Spam at the Network Level. In *EUNICE'09: 15th Open European Summer School and IFIP TC6.6 Workshop on The Internet of the Future*, Barcelona, Spain.

Steding-Jessen, K., Vijaykumar, N., and Montes, A. (2008). Using low-interaction honeypots to study the abuse of open proxies to send spam. *INFOCOMP Journal of Computer Science*, 7(1):45–53.

Veloso, E., Almeida, V., Meira, W., J., Bestavros, A., and Jin, S. (2006). A hierarchical characterization of a live streaming media workload. *Networking, IEEE/ACM Transactions on*, 14(1):133 – 146.

Xie, Y., Yu, F., Achan, K., Panigrahy, R., Hulten, G., and Osipkov, I. (2008). Spamming botnets: signatures and characteristics. In Bahl, V., Wetherall, D., Savage, S., and Stoica, I., editors, *SIGCOMM*, pages 171–182. ACM.