O uso de roteadores para proteger os recursos da rede do abuso de spams

**Banit Agrawal, Nitin Kumar e Mart Molle,
da Universidade da Califórnia (EUA)**

Diariamente, spams superlotam as caixas de entrada de e-mails. Várias técnicas já foram utilizadas na luta contra o abuso, principalmente no nível de usuário final. Neste artigo o autor propõe um sistema de bloqueio no fornecimento de spams através da limitação de taxa e controle feito pelo roteador, o que diminui o congestionamento da rede e o número de "falsos positivos".

A maioria dos trabalhos sobre controle de spams é dirigida ao nível do receptor. Com o uso de uma variedade de heurísticas essas técnicas têm bastante sucesso na identificação e/ou bloqueio da entrega de spam à caixa de entrada de e-mail do usuário. Entretanto, mesmo que os spammers não tenham sucesso em atingir o receptor, ou de ocupar espaço em seu servidor de correio, eles ainda assim estão livres para consumir grandes quantidades de largura de banda da rede no processo. Sem dúvida, alguns spammers usam o recurso como uma arma para levar ataques de negação de serviço: maciças quantidades de spam dirigidas para um único nó satisfazem o spammer ao superlotar os recursos de computação e/ou comunicação do alvo, mesmo que nenhuma daquelas mensagens seja jamais lida.

Com base numa análise da Brightmail, a porcentagem total de e-mails da Internet identificados como spam cresceu de 16%, em junho de 2002, para 50%, em agosto de 2003, e a probabilidade é de essa tendência continuar. As estatísticas da Excedent mostram que 45% do tráfego de e-mail em todo o mundo é de spam; várias empresas chegam a gastar cerca de US$ 30 bilhões por ano para controlar spams. Em 2003, a Ferris Research estimou que 44% dos custos gerais impostos por spams às empresas de todo o mundo correspondem ao desperdício de largura de banda e à conseqüente lentidão do tráfego e potencial interrupção do serviço. Imaginamos que essas questões sejam particularmente importantes para as organizações de TI de países em desenvolvimento, como o Brasil, onde a conectividade da Internet requer dispendiosos links internacionais. Portanto, o desafio aqui é de proteger os recursos da rede do abuso de spams, não apenas os usuários finais.

Neste artigo, propomos um mecanismo de controle de spam no nível do roteador. Nossa abordagem utiliza o fato de os spams serem geralmente enviados para múltiplos receptores com poucas alterações de conteúdo. Assim, nosso roteador segrega o tráfego de entrega de mensagem de outros tráfegos para maior processamento. Sempre que detecta tráfego de entrega de e-mail, invoca a primeira fase de nosso algoritmo e tenta fazer a correspondência do conteúdo da mensagem que entra em relação a um cache de mensagens candidatas recentemente vistas. Se há uma continuidade de envio de

mensagens similares, utilizamos a segunda fase de nosso algoritmo, que consiste na classificação Bayesiana da mensagem em massa (veremos neste artigo como funciona). Se o fluxo de mensagem é qualificado como spam, limitamos a taxa de sua entrega, reconfigurando a sessão TCP se o tempo passado entre as cópias consecutivas ficar abaixo de nosso limite mínimo de atraso. Portanto, essa técnica introduz custos e gastos extras para o spammer e minimiza o abuso do tráfego da rede Internet.

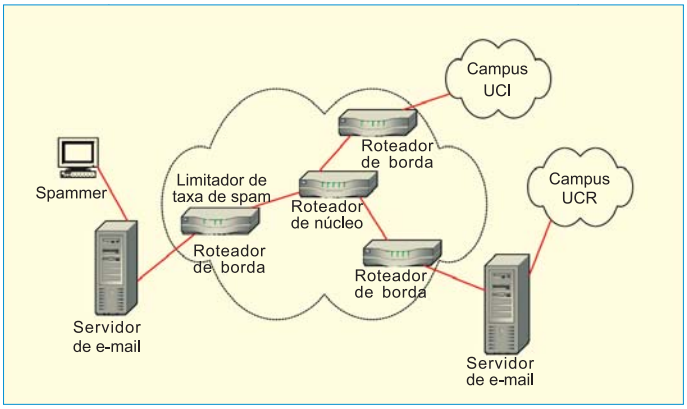

Fig. 1 – Roteamento de pacote de e-mail na Internet

Controle de spam de e-mail no roteador

Se um usuário (digamos, Alice) quer enviar uma mensagem de e-mail para outro usuário (Bob), então o correio de Alice procura o endereço do servidor de correio de Bob e abre uma conexão TCP direta para a porta 25 do servidor de Bob. Nesse ponto, os dois servidores de correio realizam a transação de entrega do e-mail, de acordo com RFC 2822 e com o SMTP - Protocolo de Transferência de Correio Simples, como especificado em RFC 2821. A figura 1 ilustra o completo processo de roteamento de forma simplificada. Durante a transação de entrega do e-mail, o diálogo inteiro entre os dois servidores de correio é visível para todos os roteadores ao longo da via entre Alice e Bob.

Se qualquer desses roteadores quiser bloquear a transação, pode simplesmente forçar a sessão TCP a fechar, por meio do envio de um segmento de recomposição a ambas as partes. Assim, temos a oportunidade de controlar spam no nível do roteador, *monitorando* todas as sessões SMTP que passam através de um roteador; *classificando* cada sessão SMTP como (não atraente) spam ou (saudável) boa; e, finalmente, *policiando* o tráfego de spam para limitar o seu consumo de recursos. Note-se que não estamos advogando a política de bloquear completamente a entrega de todos os e-mails que nossos algoritmos classificam como spam, o que seria de difícil defesa e, muito possivelmente, ilegal. Em vez disso, sugerimos a imposição de um limite do número de cópias de e-mails em massa que aceitamos por unidade de tempo.

Qualquer proposta para aumentar a quantidade de processamento de um roteador da Internet tem de incluir uma avaliação do custo de seu suporte. Felizmente, os processadores de rede especiais estão se tornando comercialmente disponíveis para muitas aplicações em redes de alta velocidade, o que oferece um significativo poder de processamento, flexibilidade e facilidade de uso e de reutilização. Os processadores de rede também podem ser usados para fazer o processamento da camada de aplicação junto com outros processamentos de camadas mais baixas, para muitas e diversas aplicações em rede, como equilíbrio de carga baseado em conteúdo, md5, criptografia, *hashing*, etc. Muitos fornecedores oferecem memória endereçada por conteúdo (CAM, do inglês *content-adressable memory*) como co-processadores a serem usados com o processador da rede para obter uma busca de dados de alta velocidade.

A limitação de taxa de e-mail pode ser acrescentada aos roteadores de borda a um baixo custo com o uso de qualquer poderoso processador de rede e co-processadores CAM de alta velocidade, o que oferece a flexibilidade e a possibilidade de programação aos desenvolvedores de roteadores. Como outra alternativa, podemos simplesmente configurar o roteador para encaminhar o tráfego SMTP para um computador externo para processamento offline, onde aplicamos nosso algoritmo limitador de taxa e controlamos o spam. O resto do tráfego da rede não é perturbado e é enviado por meio da via de dados de alta velocidade.

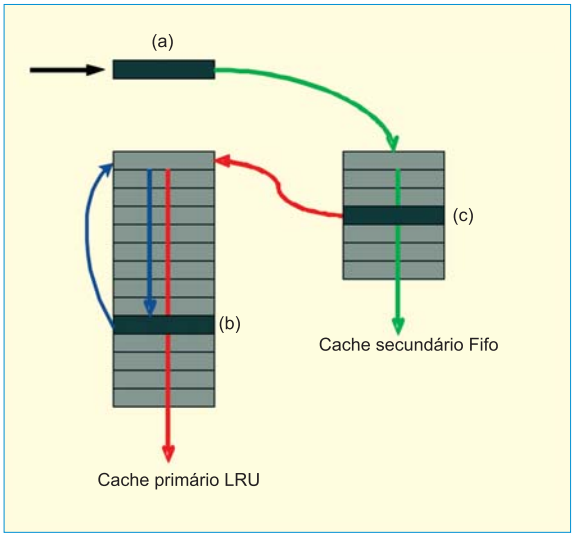

Fig. 2 – Atualização do sistema de cache de dois níveis durante a fase 1 de comparação de conteúdo. Observe que a mensagem que entra ou não é classificada como conteúdo exclusivo ou igual a algum outro padrão existente de mensagem como (b) ou (c), o algoritmo sempre deixa atrás um padrão de correspondência no topo de um dos caches

Fase de equiparação de conteúdo

Nessa primeira fase do processo de identificação,

tentamos encontrar uma correspondência entre cada nova mensagem de e-mail que entra e um cache de mensagens de e-mail previamente vistos.
O objetivo aqui é classificar corretamente cada mensagem como contendo conteúdo exclusivo ou repetido, e sem excessivos requisitos de computação ou armazenamento. Claramente, só podemos aumentar a probabilidade de identificar corretamente uma mensagem repetida comparando-a com mais amostras de conteúdo já conhecido.

No entanto, o padrão entre duas mensagens repetidas é caro por causa da informação do cabeçalho de comprimento variável específico de cada receptor. Assim, usamos uma amostragem (descrita abaixo), em vez de uma comparação do texto completo para detectar a correspondência, que pode levar a um problema de aumento do número de falsos positivos, se compararmos com muitas amostras de conteúdo exclusivo.

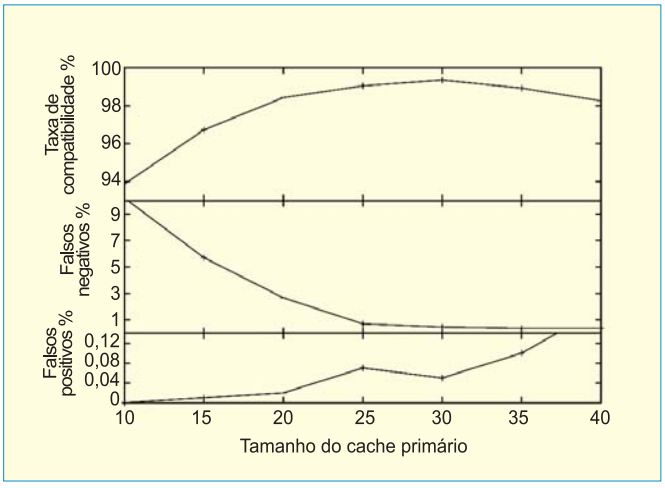

Fig. 3 – Ajuste do tamanho do cache primário de mensagens

Para minimizar esses problemas, usamos uma estrutura de cache de dois níveis, que explora a característica de "entrega de e-mails em massa em curto espaço de tempo" para separar o conteúdo de nossa estrutura de cache a mensagens repetidas. O cache de mensagem primária é usado para armazenar um protótipo de cada um dos *k* dos tipos de mensagens repetidas mais recentemente na ordem LRU – Least Recently Used, onde *k* é o tamanho do cache primário ajustável. Também mantemos um padrão de tempo de cada mensagem enviada armazenado no cache de mensagem primário para limitar a taxa dos spams. Em contrapartida, o cache de mensagem secundário é usado para armazenar e novos candidatos do tipo de mensagem repetida em ordem Fifo – First On, First Out, visto mais recentemente.

A figura 2 ilustra o processo online pelo qual o sistema de cache detecta o conteúdo de e-mail repetido. Quando o roteador recebe novas mensagens de e-mail (a), mostradas na parte superior da figura, ele compara seu conteúdo a todos os protótipos armazenados em ambos os caches. Se não forem encontradas correspondências, então o roteador classifica a mensagem (a) como sendo de conteúdo exclusivo e permite que a sessão SMTP entregue essa mensagem sem interferência. No entanto, a mensagem (a) ainda pode vir a ser o primeiro membro de um novo fluxo de mensagens repetidas

que chegarão no futuro. Assim, o roteador salva o protótipo da mensagem (a) na entrada do Fifo do cache secundário, ao mesmo tempo em que descarta os protótipos mais antigos da extremidade do cache, de modo a criar espaço.

Por outro lado, se um dos caches não contém uma equivalência, então o roteador classifica a mensagem (a) como conteúdo *repetido* e a encaminha para a fase 2 para processamento adicional de spam. Se a equivalência foi com o mesmo protótipo armazenado (b), digamos, do cache primário de LRU, então o protótipo (b) é promovido para o topo do cache LRU por meio de uma mudança circular. Do contrário, a equivalência deve ter sido com algum protótipo armazenado (c), digamos, do cache secundário Fifo. Por causa dessa equivalência, o protótipo (c) agora se qualifica para ser admitido no cache primário de LRU, descartando-se o protótipo menos recentemente usado da base do cache para criar espaço.

Nosso algoritmo de equivalência de conteúdo é otimizado para detectar spams, que são geralmente enviadas em massa a muitos usuários diferentes, em redes distintas, com uma pequena personalização do conteúdo comum. Para que sejam detectados, dividimos cada novo e-mail em múltiplas sublinhas de um comprimento fixo, chamadas padrões. O comprimento do padrão é um parâmetro ajustável, que precisa ser otimizado. O conjunto de padrões é então confrontado com todas as mensagens armazenadas no cache; se uma porcentagem suficientemente grande desses padrões for equivalente a uma dessas mensagens em cache, declaramos que contém conteúdo repetido.

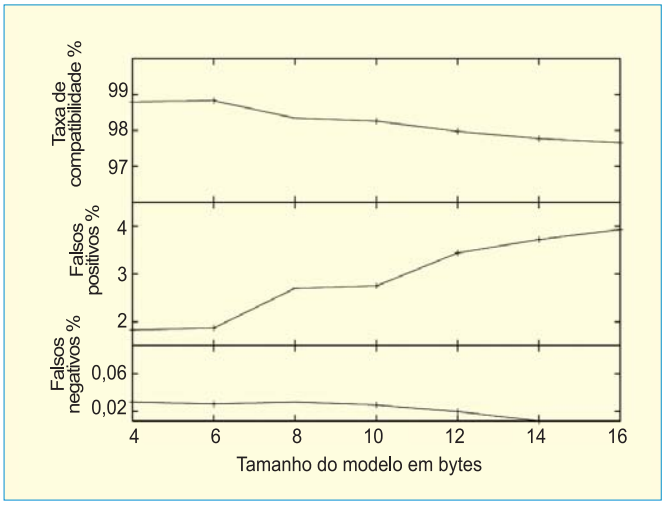

Fig. 4 – Ajuste no tamanho do padrão de mensagem

Algoritmo de Boyer Moore (BM)

Para encontrar um pequeno padrão em uma grande seqüência empregamos o algoritmo de Boyer Moore, que é muito mais eficiente em complexidade que outros algoritmos de força bruta. Ele resolve o problema de equivalência de padrão posicionando-o repetidamente sobre o texto e tentando fazer a correspondência. Para cada posicionamento, o algoritmo começa a comparar o padrão com texto a partir de sua extremidade direita. Se não ocorrem equivalências, então o padrão foi encontrado. Não sendo assim, o algoritmo faz uma mudança, que é a quantidade na qual o padrão será movido para a direita até que uma nova tentativa de equivalência seja realizada, em que o primeiro caractere do padrão e o caractere na liderança do texto se equivalham. A complexidade desse algoritmo é razoavelmente boa: o melhor tempo para se encontrar um padrão de comprimento M embutido em um texto de comprimento N é O (N/M).

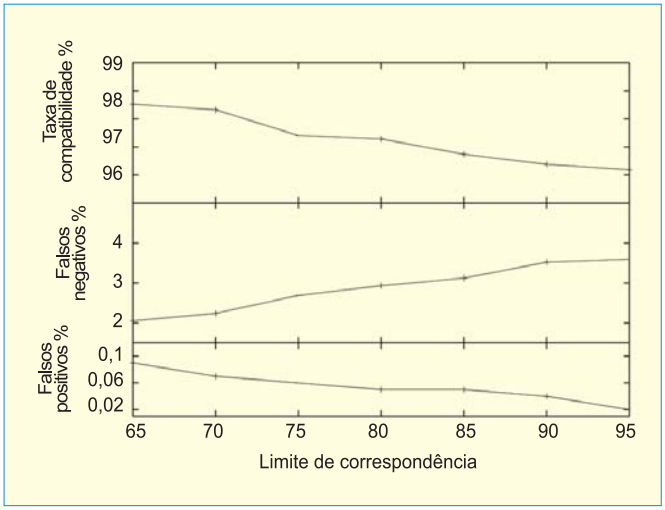

Fig. 5 – Ajuste do limite de correspondência

Fase de classificação bayesiana de spam

Aqui, descreveremos nosso método de classificação de spam. Usamos o classificador bayesiano para identificar spams. O método bayesiano é simples, autodidata, de múltiplas linguagens e leva em conta a mensagem inteira. Em geral, consiste de duas fases – treinamento e teste.

Fase de treinamento

Cada e-mail que entra é reduzido a um conjunto único de indícios para fazer parte de uma lista inicial de símbolos. A conta de bons e maus e-mails contendo esses indícios é mantida no conjunto inicial de indícios. Se um e-mail que entra possui um novo item ainda não reconhecido, então um novo sinal é acrescentado a esse conjunto e a conta é atualizada. Isso mantém o nosso conjunto sempre atualizado, com as novas palavras usadas pelos spammers, e ajuda a tornar o algoritmo mais forte contra as novas técnicas de spam.

Fase de teste

Uma nova mensagem de e-mail é convertida em um

conjunto de indícios {t_1, t_2,...}. Em seguida, usa-se cada símbolo t_i (que está presente no conjunto inicial de indícios gerada na fase de treinamento) para calcular a probabilidade condicional de essa mensagem ser spam, dada a inclusão do indício t_i. Novos sinais, ainda não presentes no conjunto inicial de indícios, são acrescentados e sua contagem é atualizada. Uma vez que a probabilidade individual de cada indicativo tenha sido gerada, nós as combinamos usando o Teorema de Bayes para ter uma estimativa geral da qualidade de spam da mensagem de e-mail.

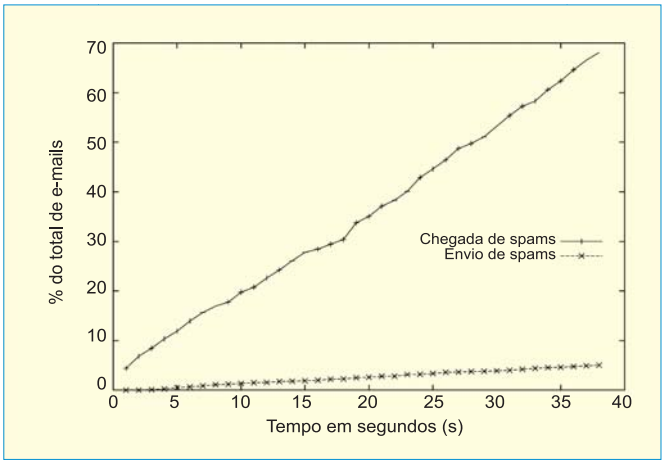

Fig. 6 – Limitação de taxa de spams

Notificação do lado do receptor

Fornecemos um mecanismo para informar o usuário final de spams, de forma que ações apropriadas possam ser tomadas no lado do receptor. Usamos uma etiqueta de cabeçalho de TCP para marcar o spam. Uma vez que a mensagem spam é encontrada no roteador, a etiqueta reservada é colocada. É prerrogativa do usuário final utilizar esse aspecto para filtrar os spams no lado do receptor. Em geral, ao se enviarem pacotes TCP/IP, todas as etiquetas reservadas de cabeçalhos TCP são removidas. Os spammers podem não querer colocar a etiqueta reservada para TCP, porque isso derruba seu objetivo. Portanto, ele enviará mensagens de e-mail com etiquetas reservadas para cabeçalho TCP removidas. A mensagem spam passa por vários roteadores na Internet. Nosso limitador de taxa no roteador da extremidade classificará a mensagem spam e colocará a classificação apropriada.

Em geral, essa classificação particular também pode ser usada para marcar um mau pacote na camada de transporte. Quando as portas da fonte e do destino são SMTP (25), o pacote TCP com a etiqueta pode ser classificado como spam. Se a porta TCP é 80, então pode ser classificado como pacote de detecção de intrusão ou pacote com vírus. Assim, podem-se fazer diferentes interferências com base nas portas TCP e nessa etiqueta.

Como essa etiqueta é colocada pelo roteador e os spammers não têm qualquer controle sobre o roteador, pode-se conseguir uma classificação eficiente.

Teste

A operação do algoritmo no roteador foi representada por simulação da execução do loop de

Scheduler de tarefa do roteador. Cada vez que o agendamento de tarefa recebe um pacote IP, invoca o nosso algoritmo. Criamos um fluxo de pacotes IP de arquivos disponíveis ao público de e-mails bons e spams.

No loop de despacho (*dispatch loop*) do roteador, acompanhamos um pacote em cada passo da simulação. Depois o pacote foi para o classificador de pacote. O classificador verifica para saber se é um pacote SMTP e então o envia para o analisador SMTP. Esse analisador determina se o pacote recebido é um pacote de mensagem ou não. Depois de verificada a identidade entre os dados, é chamado o algoritmo de correspondência de conteúdo para ver se a mensagem é uma mensagem em massa ou não. Depois da confrontação, a mensagem de e-mail é enviada para a segunda fase de classificação bayesiana.

O treinamento do classificador bayesiano foi feito usando-se um arquivo de 1513 e-mails bons e 2401 spams obtidos (outubro de 2002). O teste foi feito com o uso de um conjunto inteiramente diferente de 1000 e-mails (tanto bons quanto spams) retirados dos arquivos (fevereiro de 2003). A avaliação do algoritmo baseou-se em: precisão, porcentagem de falsos positivos e porcentagem de falsos negativos. A precisão é definida pela seguinte fórmula:

Precisão = ½ (total de e-mails bons ÷ total de e-mails bons classificados corretamente) + (total spams ÷ total de spams) X 100

Falso positivo é definido como a porcentagem de e-mails bons identificados como spams, enquanto falso negativo é definido como a porcentagem de spams identificados como e-mails bons. Em nossa análise usam-se valores típicos de 100 para o tamanho de cache primário, 20 para tamanho de cache secundário, 75% para limite de correspondência e 60 segundos para intervalo de tempo de limite de taxa.

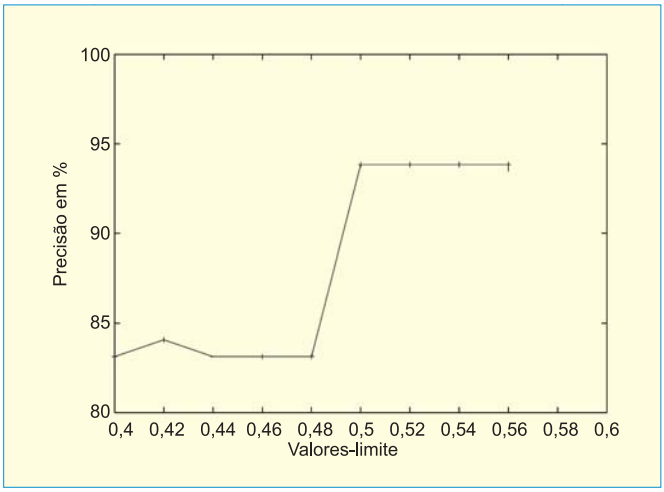

Fig. 7 – Precisão observada em vários valores-limite

Fase de correspondência de conteúdo

Nosso algoritmo de correspondência de conteúdo inclui vários parâmetros de controle, como o tamanho do cache primário, tamanho de cache secundário, tamanho do padrão, valor do limitador de taxa e limite de correspondência para um fino ajuste do desempenho de nosso algoritmo de equiparação. Em nossas análises usam-se valores típicos de 100 para o tamanho de cache primário, 20 para o tamanho de cache secundário, 75% para o limite de correspondência e 60 segundos de intervalo de tempo do limite de taxa. Os efeitos da alteração desses parâmetros são mostrados nas figuras 3 e 6.

Variação do tamanho dos caches de mensagens

A sensibilidade de três métricas de desempenho para o tamanho do cache primário de mensagem é mostrada na figura 3. Vemos que a precisão é geralmente alta, e sobe a aproximadamente 99%, para tamanhos de cache entre 25 e 35. Ao mesmo tempo, vemos que a porcentagem de falsos negativos cai significativamente (de aproximadamente 10% para quase 0) na medida em que aumentamos o tamanho do cache primário de mensagens. No entanto, essa melhoria é equilibrada por um aumento muito menor na porcentagem de falsos positivos (de quase zero para aproximadamente 0,15%).

Vemos muito pouca sensibilidade ao tamanho do cache secundário, o que sugere que é necessário fazer outros testes com o uso de traços de mensagem adicionais, antes de chegar a qualquer conclusão.

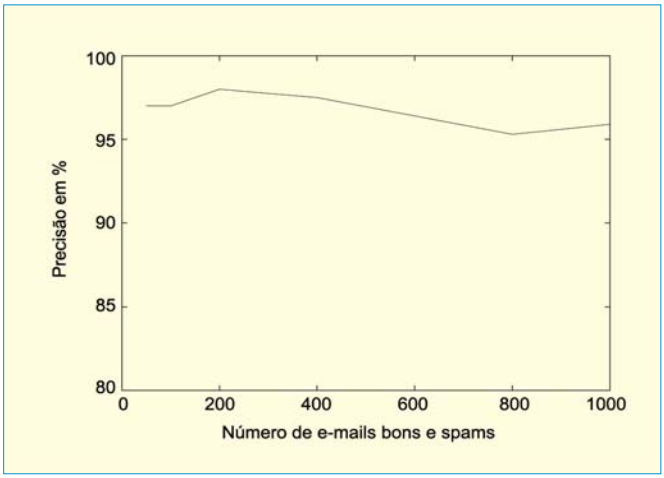

Fig. 8 – Precisão de correspondência bayesiana

Variação do tamanho do padrão

A figura 4 mostra a sensibilidade de nossas três métricas de desempenho ao tamanho do padrão usado para correspondência de conteúdo. Usualmente, um spammer envia múltiplas cópias de uma correspondência fazendo poucas alterações. Uma vez mais, vemos que a precisão é sempre muito

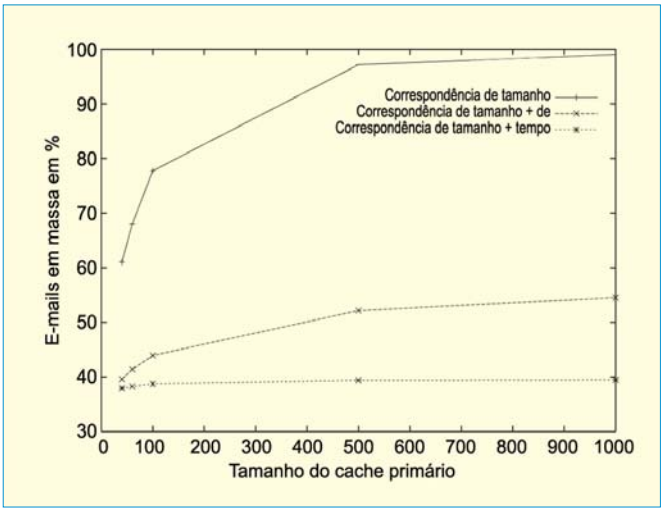

Fig. 9 – Identificação de e-mail em massa com crescente tamanho de cache primário. Vemos um aumento muito menor da porcentagem de e-mails em massa com o aumento do cache primário em todos os casos

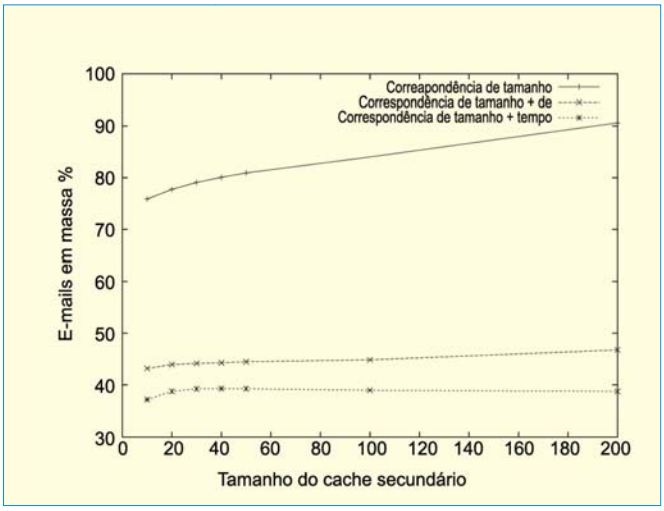

Fig. 10 – Identificação de e-mails em massa com um número crescente do cache secundário. Há um aumento marginal na identificação dos e-mails em massa com o aumento do tamanho do cache secundário

alta, enquanto a probabilidade de falsos positivos é sempre extremamente baixa e a probabilidade de falsos negativos é pequena, porém crescente.

Variação do limite de correspondência

O limite de correspondência é o limite mais baixo da porcentagem obtida para que uma mensagem que entra seja qualificada como uma mensagem de e-mail. Se mantivermos o valor do limite muito baixo, então as chances de falsos positivos aumentam. O esquema das três métricas com variação de limite de equiparação é mostrado na figura 5.

E-mails spam com limite de taxa no roteador de borda

Esquematizamos a porcentagem de mensagens recebidas no roteador de borda e a porcentagem de spams enviados pelo roteador. O esquema é mostrado na figura 6.

A partir da figura, podemos ver que 60% dos spams têm limite de taxa no final da simulação.

Fase de classificação Bayesiana de spam

Uma mensagem é declarada como spam se a estimativa da qualidade de spam do e-mail é maior do que o

valor-limite; não sendo assim, é declarada como e-mail bom.

Variação do número de spams e e-mails bons

A precisão foi calculada variando-se o número de spam e e-mails bons, mostrados no gráfico da figura 8. A precisão encontrada foi, em média, de 97%.

Variação do valor do limite

É usada para a classificação de um e-mail como bom ou spam com base no indicador de qualidade de spam. A figura 7 mostra a precisão obtida, considerando-se vários valores de limite. Esse experimento usa o valor-limite de 0,5.

Número de falsos positivos e falsos negativos

Obtivemos falsos positivos quase desprezíveis pelo classificador bayesiano. Os resultados experimentais detalhados encontram-se na tabela I.

Tab. I – Precisão do algoritmo bayesiano

Caso	Número de e-mail bom não visto	Número de e-mail spam não visto	Precisão (%)	Número de falsos positivos	Número de falsos negativos
1	50	50	97	0	0
2	100	100	97	0	2
3	200	200	98	1	6
4	400	400	97,5	1	17
5	800	800	95,3	0	71
6	1000	1000	95,8	3	121

Análise de registros de e-mails em tempo real

Aqui fornecemos uma análise detalhada de registros de e-mails em tempo real. Usamos, por um mês, um registro de e-mails de nosso departamento para a análise experimental. O tamanho do e-mail, campo “de” do e-mail e a marcação de tempo do e-mail são os únicos atributos usados em nossa análise. Fornecemos a justificativa do curto espaço de tempo para e-mails em massa no domínio temporal, selecionando diferentes valores de intervalo de tempo e de limitação de taxa. Também fornecemos vários parâmetros, como tamanho do cache primário e secundário,

porcentagem de correspondência na identificação de e-mails em massa. Usamos diferentes formas de classificação de volume, como:

• Quando o tamanho do e-mail é compatível ("tamanho"), usar apenas o tamanho da correspondência do e-mail não é suficiente. Achamos que o falso positivo, nesse caso, é muito mais alto e propomos usar outros aspectos do e-mail para identificar e-mails em massa.

• Quando o tamanho e o campo "de" do e-mail equiparam-se ("tamanho + de"). O campo extra de endereço ajuda-nos a reduzir os falsos positivos.

• Quando os tamanhos equiparam-se e o novo e-mail não passa no tempo de limite de taxa imposto pela marcação de tempo do e-mail equiparado e no espaço de tempo do limite de taxa ("tamanho + tempo"). Usam-se valores típicos de 100 para tamanho de cache primário, 20 para tamanho de cache secundário, 1% para porcentagem de equiparação de tamanho e 60 segundos para intervalo de tempo do limite de taxa para encontrar os resultados de desempenho com a variação de diferentes parâmetros.

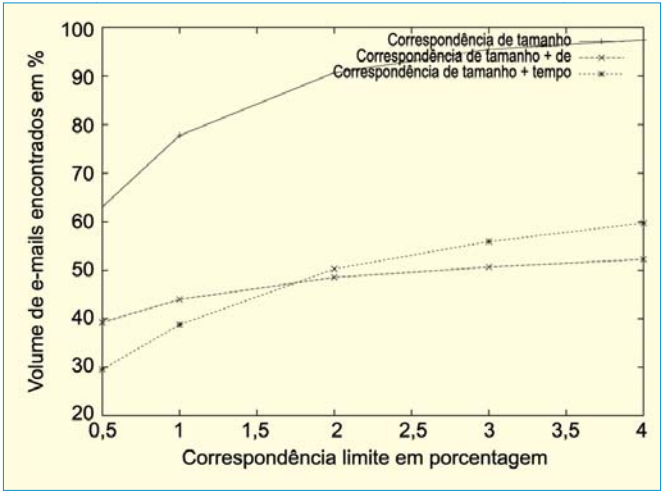

Fig. 11 – Identificação de e-mails em massa com crescente porcentagem de limite de correspondência

Tamanho de cache primário

Variamos o tamanho do cache primário de 40 para 1000 a fim de encontrarmos seus efeitos sobre a identificação de e-mails em massa. Na figura 9, podemos ver que há um leve crescimento na porcentagem de e-mails em massa quando o tamanho do cache primário é aumentado de 100 para 1000. Quando estamos interessados somente no tamanho da mensagem, então vemos que cerca de 90% dos e-mails são classificados como e-mails em massa ao se usar um tamanho de cache primário maior que 500. Isso nos diz que usar apenas o aspecto do tamanho levará a um grande número de falsos positivos para um tamanho maior de cache primário. Porém, no caso de o tamanho da mensagem ser combinado ao aspecto do endereço ou do espaço de tempo, vemos que apenas 30% a 50% dos e-mails totais são classificados como e-mails em massa. O aumento do tamanho do cache primário além de um limite particular (≥ 500) não aumenta em grande extensão a porcentagem de e-mails em massa.

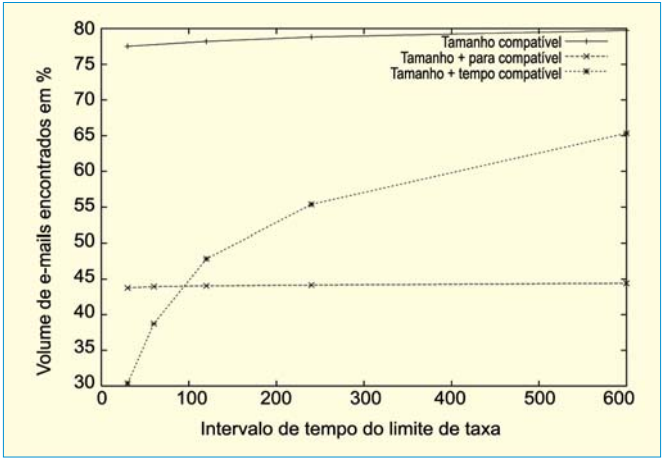

Fig. 12 – Identificação de e-mails em massa com intervalo de tempo crescente. No caso de equiparação "tamanho + tempo", a porcentagem de e-mail em massa aumenta na medida em que aumenta o intervalo de tempo

Tamanho de cache secundário

Fornecemos o efeito do aumento do tamanho do cache secundário na identificação de e-mails em massa. Variamos o tamanho do cache secundário de 10 para 200 e os resultados são mostrados na figura 10. Nesse caso, a inclinação da curva é muito menos intensa em comparação com o cache primário.

Vemos um aumento muito mais baixo da classificação de e-mail em massa com o aumento do tamanho do cache secundário, em comparação com o do tamanho do cache primário. Embora o falso positivo seja evidente no caso de correspondência de tamanho, a porcentagem de e-mails totais classificada como de e-mails em massa permanece na faixa de 30% a 50%, tanto para correspondência de "tamanho + de" quanto pra "tamanho + tempo". Vemos que a introdução das restrições de tempo com a correspondência de "tamanho" reduz os falsos positivos e identifica corretamente a maioria dos e-mails em massa.

Porcentagem de correspondência

Variamos a porcentagem de tamanho de 0,5% para 4% e o resultado é mostrado na figura 11. Se aumentarmos a porcentagem de correspondência de tamanho, as chances de ter mais igualdade torna-se alta, porque o tamanho da nova mensagem pode se assemelhar com o tamanho das mensagens armazenadas nos caches. Como podemos ver na figura, com o aumento da porcentagem de correspondência, no caso do "tamanho", os e-mails

em massa encontrados podem variar de 63% a 98%. Assim, precisamos manter a porcentagem de correspondência de tamanho para um valor razoavelmente baixo, que reduza os falsos positivos. Para se avaliar outros parâmetros da configuração, mantivemos o valor típico conservador de 1% para essa porcentagem de limite. Como na correspondência "tamanho + de" o campo de endereço deve também se equiparar, então não vemos muito aumento na porcentagem de e-mails em massa. Mas, no caso da equiparação "tamanho + tempo", como o intervalo de tempo permanece constante, o aumento da porcentagem de equiparação aumenta a porcentagem de e-mails em massa a um valor mais alto do que em comparação com a equiparação "tamanho + de". Portanto, vemos um ponto *crossover* na porcentagem do limite de equiparação de cerca de 2%.

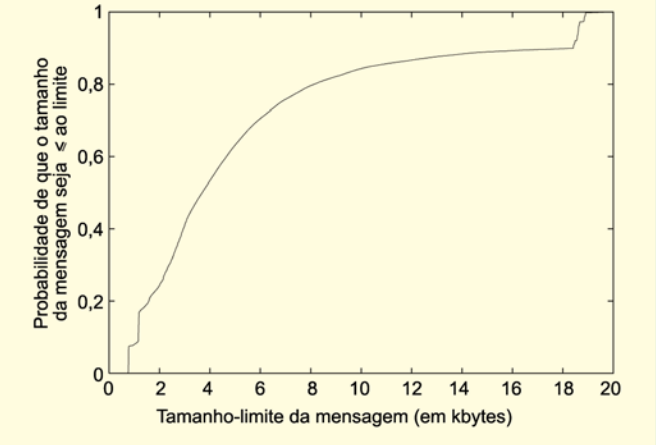

Fig. 13 – Distribuição de tamanho de e-mails em massa

Intervalo de tempo de limitação de taxa

Esse parâmetro afeta bastante os e-mails em massa "tamanho + tempo". Variamos o intervalo do limite da taxa de 30 para 600 segundos. O resultado correspondente é mostrado na figura 12. Como esperado, não vemos muitas mudanças no caso de "tamanho" e "tamanho + de". Mas, no caso de equiparação de "tamanho + tempo", como aumentamos o intervalo de tempo para identificar os e-mails em massa, vemos um maior aumento na porcentagem de e-mails em massa. Para um intervalo de tempo maior que 100 segundos, vemos que a porcentagem de e-mail em massa é maior que 40% e continua crescendo com os aumentos do intervalo de tempo do limite de taxa.

Assim, vemos muitos falsos positivos quando aumentamos o espaço de tempo do limite de taxa acima de 100 segundos. Portanto, mantemos o valor clássico de 100 segundos como o intervalo de tempo em outras avaliações.

Distribuição de e-mails de tamanho cumulativo

Também analisamos a distribuição de tamanho cumulativo de e-mails em massa (tamanho + de) e e-mails temporais em massa (tamanho + tempo). Os gráficos da distribuição de tamanho cumulativo de e-mails em massa e de e-mails temporais estão nas figuras 13 e 14, respectivamente. O eixo x mostra o limite de tamanho da mensagem,

enquanto o eixo y mostra a probabilidade de o tamanho, na distribuição, ser menor que o limite. A partir das figuras 13 e 14 fica evidente que a maioria dos e-mails em massa está na faixa de 2000 a 8000 bytes de tamanho.

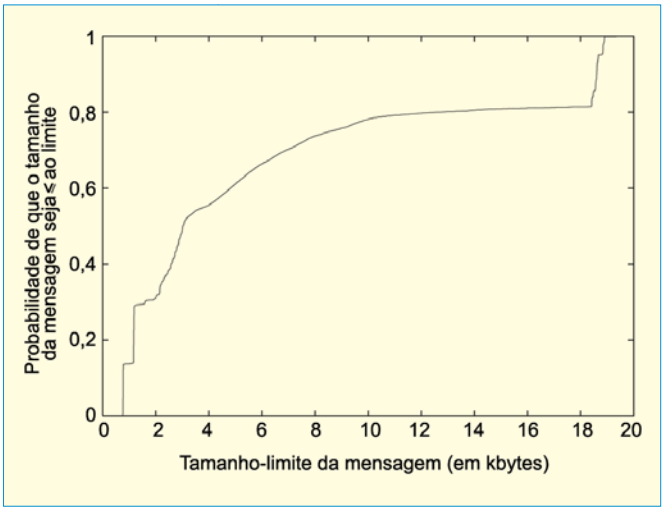

Fig. 14 – Distribuição de tamanho cumulativo de e-mails em massa temporais (tamanho + tempo)

Distribuição de chegada de e-mails

A distribuição de chegada de e-mails mostra seu verdadeiro comportamento no domínio temporal. Fizemos um gráfico do tamanho do e-mail versus a marcação de tempo do e-mail. Os gráficos da distribuição de chegada de e-mail em massa e e-mail em massa temporal estão nas figuras 15 e 16, respectivamente. Podemos ver que há muitas posições horizontais na figura, o que demonstra o aspecto do curto espaço de tempo dos e-mails em massa.

Trabalhos correlatos

O spam é um problema cada vez maior para usuários de e-mail e muitas soluções já foram propostas. Essas soluções variam de uma taxa de postagem até simplesmente não aceitar e-mails de pessoas que não se conhece. A filtragem do spam é uma forma de se reduzir o impacto do problema de um usuário individual (embora nada seja feito para reduzir o efeito que o spam produz na rede). Em sua forma mais simples, a filtragem de spam é um mecanismo para identificar e filtrar os spams. Não bloqueia os spams que entram, mas evita que um autêntico servidor de mensagens faça spam de outros.

Bloqueio por sub-rede ou número IP, o nome do domínio, nomes de domínio não resolvidos, filtragem de cabeçalho acionado por cabeçalhos inválidos, verificação do "para:" do destinatário no cabeçalho, são algumas das técnicas usadas hoje.

Como a precisão dessas técnicas não é 100%, alguns raros e-mails legítimos podem não ser entregues.

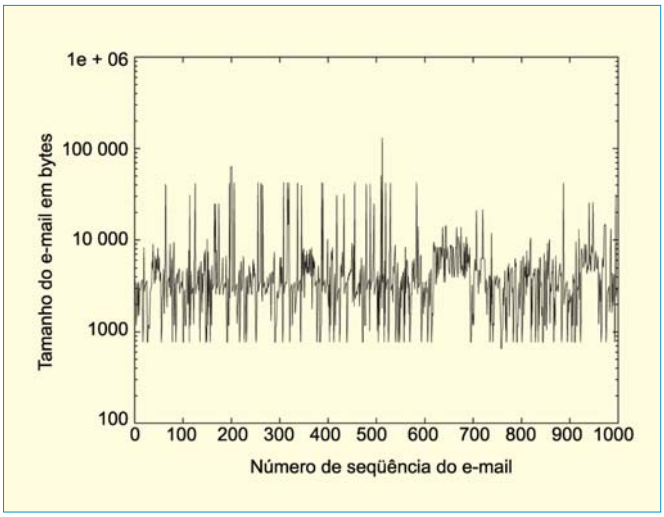

Fig. 15 – Distribuição de chegada de e-mails em massa

O *TarProxy* é um método de limitação (*throttling)* de conexões entre spammers e um servidor SMTP, que diminui a velocidade da taxa em que o spammer pode enviar spam. A técnica de filtragem de spam, baseada no conteúdo do e-mail, utiliza a freqüência de palavras, a freqüência das seqüências, valores pesados para palavras particulares, pesagem de vários aspectos, tais como parte "DE" suspeita e muitas outras perspectivas relativas ao cabeçalho da mensagem. Essas técnicas amplamente usadas incluem várias formas de filtragem bayesiana como a classificação *Bayesiana Naive*, a abordagem baseada em memória, cadeias *Markov* e máquinas de suporte de vetor (SVMs). A técnica de classificação bayesiana de spam propagandeia um nível de precisão de 99%, mas tem o ponto fraco de considerar a independência dos aspectos. O método de *Vipul* é um bom exemplo de filtragem colaborativa. É uma rede distribuída de detecção e filtragem de spam que estabelece um catálogo de atualização permanente de spams em propagação. Nesse esquema, usam-se assinaturas aleatórias e estatísticas para apontar conteúdo mutante de spam. Nossa abordagem é uma idéia razoavelmente nova de detecção e limitação de taxa de spams, usando um eficiente algoritmo de correspondência de conteúdo em tamanho de padrão ajustável e classificador bayesiano. Essa técnica é empregada no nível do roteador, que efetivamente utiliza a largura de banda, nunca bloqueia e-mails legítimos e limita a taxa de spams. Também notifica o usuário dos spams, identificados no roteador, estabelecendo uma classificação reservada no cabeçalho TCP.

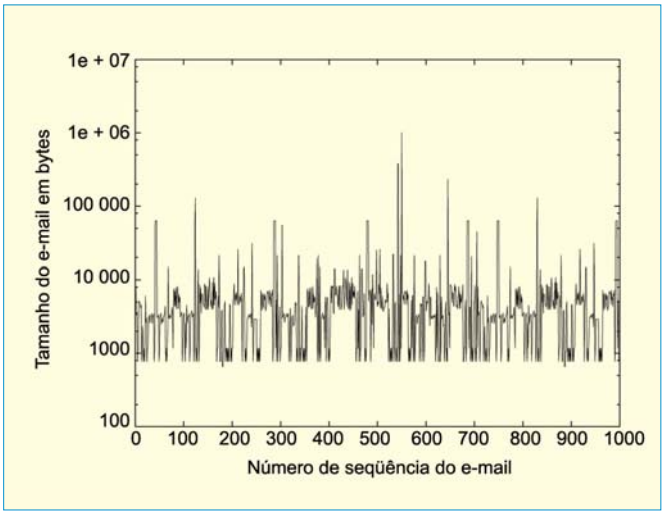

Fig. 16 – Distribuição de chegada de e-mails temporais em massa (tamanho + tempo)

Conclusão e trabalho futuro

Implementamos uma abordagem em duas fases para detectar spam no nível do roteador. A primeira identifica e-mail em grande volume por correspondência com padrão e a segunda fase aplica o classificador bayesiano sobre a mensagem em massa identificada para classifica-la como spam. Esse trabalho demonstra que uma quantidade significativa de controle de spam pode alcançar sucesso no nível do roteador. A abordagem não apenas protege os usuários finais de

Algoritmo de correspondência de conteúdo

1: correspondência ← 0;
2: contagem ← 0;
3: padrões [] ← dividir a mensagem que entra em padrões de tamanho fixo;
4: **para** cada elemento de cache de tamanho k **fazer**
5: **para** cada P em padrões []; fazer
6: correspondência ← procurar P no elemento do cache
7: **se** equivalente **então**
8: contar ← contar + 1
9: **encerrar se**
10: **encerrar para**
11: **se** contar > limite de conta **então**
12: fazer processamento de spam
13: voltar;
14: **se terminar**
15: contar ← 0;
16: **encerrar para**
17: fazer outro processamento

excessivos volumes de mensagens não solicitadas, mas também limita o congestionamento da rede causado por spams. Os testes conduzidos em vários spams coletados de diferentes fontes deram um nível médio de precisão de 97%. O método foi mais efetivo com os "falsos positivos". O número de "falsos positivos" foi quase desprezível (tabela I). Pode ser implementado com o uso de um processador de rede projetado para aplicações de rede de grande velocidade, que oferece possibilidade de programação, flexibilidade e boa eficiência. Nossa abordagem traz um custo extra e encargos para o envio de spams e controla o abuso do tráfego da Internet.

Apesar disso, ainda há muito espaço para aperfeiçoamentos. Nosso esquema pode ser complementado pelo esquema de *Vipul* para alcançar uma plataforma de filtragem mais colaborativa. O esquema proposto também pode ser usado para detectar vírus, armazenando assinaturas de vírus separadamente. A detecção de vírus, juntamente com detecção de spam no nível do roteador, pode ser um grande passo na direção de se construir um backbone seguro na Internet.