Estimativa de Holt-Winters para Detecção de Ataques em Redes WAN

Sidney C. de Lucena[1], Alex Soares de Moura[2]

[1]Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Av. Pasteur, 458 – Urca
22290-240 – Rio de Janeiro, RJ

[2]Rede Nacional de Ensino e Pesquisa (RNP)
Rua Lauro Müller, 116/1103 – Botafogo
22290-906 – Rio de Janeiro, RJ

sidney@uniriotec.br, alex@rnp.br

Abstract. *Attacks against networks and its services are permanent concerns for Internet service providers. Several methods for malicious traffic detection in WANs have been researched in the last years. This article evaluates a method based in the Holt-Winters forecasting algorithm to verify significant changes at the pattern of IP addresses and port numbers, normally affected in the presence of attacks. This work also proposes and evaluates the use of filters to increase the effectiveness of the method for the detection of attacks. Results confirm the usefulness of this proposal to detect malicious traffic related to a TCP SYN flood attack and to the propagation of the Slammer worm, both applied to real traffic samples from RNP's WAN backbone.*

Resumo. *Ataques à segurança de redes e serviços são preocupações constantes para provedores de serviços de Internet. Diversos métodos para detecção de tráfego malicioso em WANs têm sido alvo de pesquisa nos últimos anos. Este artigo avalia um método baseado no algoritmo de Holt-Winters para verificar mudanças significativas nos padrões de IPs e portas, geralmente afetados na presença de atividades maliciosas. O artigo também propõe e avalia o uso de filtros para aumentar a eficácia do método na detecção de ataques. Resultados comprovam a aplicabilidade da proposta na detecção de um TCP SYN Flood e da propagação do worm Slammer, ambos aplicados a amostras reais de tráfego oriundas do backbone da RNP.*

1. Introdução

A cada ano que se passa, provedores de serviço de Internet (ISPs) e outros que oferecem serviços de *datacenter*, como do tipo *hosting* ou *colocation*, tornam-se cada vez mais requisitados por clientes dos mais variados portes. O aumento da popularidade de determinados serviços aliado com o barateamento do custo da banda, em todos os níveis, acarreta numa elevação considerável do volume de tráfego que passa pelas redes destes provedores. É natural, portanto, que a preocupação com a segurança seja cada vez maior, já que aumenta também a probabilidade de clientes e serviços serem vítimas de ataques. Face a esta nova realidade, atuar de maneira reativa mediante reclamações de clientes pode prejudicar consideravelmente a credibilidade na oferta de serviços de

conectividade, sendo necessário, portanto, uma postura proativa visando a detecção de ataques que possam estar em curso.

Há várias formas de se fazer detecção de ataques, entretanto, nem todas são apropriadas para uso em redes WAN, que é o tipo de rede empregado pelos ISPs. Os chamados *Intrusion Detection Systems* (IDSs), por exemplo, amplamente adotados em redes LAN para a sinalização de ataques, utilizam mecanismos para inspeção de pacotes IP em busca de dados pertinentes à camada de aplicação. Tal procedimento, também chamado de *deep packet inspection*, não é apropriado para um ambiente de rede WAN em função do alto volume de tráfego que passa pelos enlaces. Um roteador que venha a realizar uma inspeção profunda nos pacotes que por ele trafegam, por exemplo, sacrificaria por demais o processamento e, conseqüentemente, o encaminhamento dos pacotes. No caso de um processamento "externo" dessas informações, isto também exigiria um alto investimento em infra-estrutura de rede e armazenamento, além do custo do processamento em si.

Para redes WAN, a estratégia mais apropriada para detecção de ataques é o uso de assinatura estatística de tráfego. Trata-se de um procedimento para caracterizar estatisticamente o tráfego normal que passa por uma rede. Desta forma, qualquer desvio significativo destas características pode ser sinal de um ataque. Neste caso, usa-se também o termo anomalia, que compreende não somente os ataques à segurança da rede, mas também qualquer outro tipo de evento anormal, como falhas de enlace ou mudanças bruscas no roteamento.

Alguns métodos foram propostos nos últimos cinco anos objetivando a detecção de anomalias em redes WAN ([Paschalidis 2009], [Androulidakis 2009], [Zonglin 2009], [Lakhina 2005]). No método apresentado em [Lakhina 2005], a assinatura de tráfego utiliza medidas de entropia extraídas a cada cinco minutos para IPs e portas dos pacotes que passam pela rede. Para a identificação e categorização dos fluxos anômalos, em [Lakhina 2005] é utilizado o método de *Principal Component Analisys* (PCA), sendo criticado em [Brauckhoff 2009] por não capturar correlações temporais, o que pode acarretar numa taxa relevante de "falsos negativos". Um ponto importante com relação ao método proposto em [Lakhina 2005] é que ele é do tipo *network-wide*, ou seja, analisa todo o tráfego que entra e sai de uma rede num dado intervalo de tempo.

Em [Lucena 2009] é apresentado um método mais simples para detecção de anomalias baseado em [Lakhina 2005], valendo-se também das mesmas medidas de entropia, porém numa abordagem *single-link*. Ou seja, as medidas são extraídas para IPs e portas dos pacotes que trafegam por uma única interface. Portanto, somente ataques que passam pela interface monitorada são detectados. Para a detecção de padrões anômalos, o trabalho comparou o uso do EWMA (Exponentially Weighted Moving Average) e da estimativa de Holt-Winters como mecanismos de *forecasting* aplicados a séries temporais, sendo que o Holt-Winters se mostrou mais eficaz uma vez que as medidas de entropia apresentam sazonalidade. Dada uma série temporal de entropias, um desvio significativo do padrão estimado pelo mecanismo de *forecasting* sinaliza uma possível anomalia.

O presente trabalho avalia a eficiência da estimativa de Holt-Winters na detecção dos dois tipos de ataques que mais causam transtorno aos operadores de redes WAN: o DDoS (*Distributed Denial of Service*) e a proliferação de *worms*. Baseado nos resultados obtidos, o presente trabalho também propõe e atesta uma melhoria ao método

baseada num procedimento simples durante o processo de medição do tráfego. É importante ressaltar que este trabalho se preocupa com métodos que possam ser facilmente acoplados aos sistemas de gerenciamento tradicionalmente adotados pela maioria dos centros de operação de redes WAN. A análise realizada utiliza amostras reais de tráfego do *backbone* da RNP (rede Ipê) e se vale de um mecanismo para injeção de tráfego malicioso, artificialmente gerado com base nas características reais dos ataques. No caso, os ataques usados foram o *worm Slammer* e o TCP SYN *Flood*.

A Seção 2 do artigo descreve o método de detecção de ataques usando a estimativa de Holt-Winters, a Seção 3 destaca aspectos da implementação do método, a Seção 4 apresenta a metodologia experimental, a Seção 5 traz os resultados obtidos e a Seção 6 apresenta a conclusão e propõe trabalhos futuros.

2. Método de *Forecasting* para Detecção de Ataques

2.1. Assinatura Estatística de Tráfego Usando Entropia de Shannon

A entropia de Shannon ([Shannon 1948]) é definida como:

$$E_S = -\sum_{i=1}^{N} p_i \log_2(p_i) \quad , \tag{1}$$

onde N é o número de diferentes ocorrências no espaço amostral e p_i a probabilidade associada a cada ocorrência i. No trabalho de [Lakhina 2005], N corresponde ao número de diferentes valores para IP de origem, IP de destino, porta de origem ou porta de destino, valores estes que ocorreram durante um dado intervalo de tempo. p_i é, portanto, a fração de ocorrência de cada um desses diferentes valores no intervalo de tempo medido. O resultado varia entre *zero* e log_2N, onde *zero* indica concentração máxima na distribuição medida e log_2N indica máxima dispersão na distribuição medida. Em [Lakhina 2005], foi mostrado que variações inesperadas no padrão de uma ou mais dessas séries temporais de entropias servem como indicativo da presença de algum tipo de anomalia.

2.2. Estimativa de Holt-Winters como Método de *Forecasting*

A predição ou estimativa de Holt-Winters divide uma série temporal em três partes superpostas: um termo que denota a periodicidade da série, um segundo termo que indica a tendência de crescimento da série e um terceiro que expressa a parte residual da série, resultante da dissociação desta das duas partes anteriores. Cada um desses três termos é tratado de forma separada através de um EWMA (*Exponentialy Weighted Moving Avarage*). O EWMA é um estimador aplicado a séries temporais que estabelece uma ponderação entre o valor atual da série e a estimativa anterior. Sua expressão é dada por:

$$X_{t+1} = \alpha x_t + (1-\alpha) X_t \quad , \tag{2}$$

onde X_t é a média histórica e x_t o valor corrente. $0 < \alpha < 1$. Assim sendo, será atribuído um EWMA para cada termo e a soma dos três resulta na estimativa de Holt-Winters.

Supondo o uso da sazonalidade aditiva, que é aquela onde o termo que representa a variação periódica da série temporal possui comportamento estatístico que independe da taxa de crescimento (positiva ou negativa) da série, as expressões para a estimativa de Holt-Winters seguem abaixo, onde a_t corresponde à componente residual,

b_t à componente de tendência de crescimento, c_t à componente periódica e m é o tamanho do período:

$$X_{t+1} = a_t + b_t + c_{t+1-m} \quad , \tag{3}$$

$$a_t = \alpha(X_t - c_{t-m}) + (1-\alpha)(a_{t-1} + b_{t-1}) \quad , \tag{4}$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \quad , \tag{5}$$

$$c_t = \gamma(X_t - a_t) + (1-\gamma)c_{t-m} \quad . \tag{6}$$

A vantagem de se usar a estimativa de Holt-Winters, ao invés de um EWMA puro, por exemplo, reside justamente na capacidade do Holt-Winters de ser aderente não somente ao padrão de sazonalidade da série temporal, mas também a sua tendência de crescimento. Nos resultados mostrados na Seção 5, é possível verificar que a entropia extraída para certos parâmetros pode apresentar sazonalidade com períodos de 24 horas.

2.3. Detecção de Ataques Usando Método de *Forecasting*

A simples adoção de medidas de entropia, por si só, não se configura num processo eficiente de detecção de anomalias. Faz-se necessário o uso de alguma técnica que seja capaz de verificar automaticamente, e em tempo real, se a medida de entropia fugiu de seu padrão de normalidade. No caso do presente trabalho, a técnica adotada é a estimativa de Holt-Winters: medidas de entropia que difiram substancialmente daquelas previstas pelo Holt-Winters são tratadas como uma anomalia, possivelmente um ataque. Conforme a metodologia apresentada em [Lakhina 2005], devem ser verificadas as quatro séries temporais de entropia extraídas para IPs e portas, tanto de origem quanto de destino. A forma como se pode automaticamente verificar o desvio do padrão de normalidade apontado pelo Holt-Winters é mostrada na Seção 3.4.

3. Aspectos de Implementação

3.1. Definição de Fluxo IP

Um fluxo IP é definido como uma seqüência unidirecional de pacotes onde cada pacote contém os mesmos valores para IP de origem, IP de destino, porta de origem, porta de destino e campo *protocol*. O intervalo de tempo entre pacotes de um mesmo fluxo não deve ultrapassar um valor máximo, que por *default* é igual a 15 segundos na maioria das implementações dos fabricantes de roteadores. Caso o intervalo de tempo ultrapasse este limite, o fluxo expira e um novo se inicia, conforme descrito em [Cisco 2008]. Outros critérios adotados para decretar o término de um fluxo IP são os pacotes RST e FIN, para conexões TCP, e o tempo de vida máximo do fluxo, geralmente configurado como sendo 30 minutos.

3.2. Formato *NetFlow*

A maneira mais utilizada para a coleta de fluxos IP é através da funcionalidade que os roteadores possuem de exportar estas informações cada vez que um fluxo expira. Dentre os padrões usados para estruturar estas informações, o *NetFlow* versão 5 ([Cisco 2008]) é o mais utilizado. A leitura destas informações é realizada para cada interface lógica do roteador mediante configuração. Os registros *NetFlow* são então exportados, usando pacotes UDP, para uma ou mais estações, chamadas de "coletores *NetFlow*".

3.3. Extração das Entropias

O cenário de aplicação deste trabalho considera sistemas de gerenciamento de redes onde o histórico das estatísticas de tráfego é armazenado em bases RRD (*Round-Robin Database*, [Bogaerdt 2008]), geralmente usando medidas extraídas a cada intervalo de cinco minutos. É então proposto que registros *NetFlow* sejam capturados e que tenham seus valores de entropia calculados para cada intervalo de cinco minutos, e que os resultados sejam armazenados em bases do tipo RRD, uma para cada parâmetro (IP de origem, IP de destino, porta de origem e porta de destino). Vale observar que, em [Lakhina 2005], as medidas de entropia são também calculadas para intervalos de cinco minutos a partir dos registros de fluxo coletados. Assim sendo, as entropias devem ser calculadas contabilizando-se todos os pacotes dos fluxos registrados a cada intervalo de cinco minutos, montando-se os histogramas de cada parâmetro e aplicando-se a expressão descrita em (1). De maneira a uniformizar o grau de concentração/dispersão, os valores calculados podem ser normalizados por $log_2 N$, onde N corresponde ao número de ocorrências do respectivo parâmetro para cada intervalo de cinco minutos.

3.4. Aplicação de Estimadores e Verificação de Ataques

A estimativa de Holt-Winters (HW) pode ser calculada utilizando uma função especial da ferramenta *RRDtool*, conforme descrito em [Brutlag 2000]. O *RRDtool* é um aplicativo para manipulação de bases RRD amplamente adotado como componente básico de várias ferramentas de *software* livre para gerenciamento de redes.

A partir da série de predições do HW, o *RRDtool* calcula um limiar superior e inferior para o que pode ser considerado como comportamento normal. Ou seja, se o valor de uma nova amostra de entropia está fora deste intervalo, é sinal de que pode ser um ataque. O cálculo deste intervalo nada mais é do que um EWMA para o valor do desvio, que é a diferença absoluta entre valor estimado e valor real. No caso, este EWMA considera o ciclo sazonal da série temporal e usa o mesmo γ como coeficiente de amortização:

$$desvio_t = \gamma |x_t - X_t| + (1-\gamma) desvio_{t-m} \quad , \tag{7}$$

onde x_t é o valor real e X_t o valor estimado. Assim, tem-se que o intervalo adotado como limite para flutuações dos valores reais equivale a

$$(X_t - \delta.desvio_{t-m}, X_t + \delta.desvio_{t-m}) \quad , \tag{8}$$

onde δ é um fator multiplicador. Valores razoáveis para δ estão entre *2* e *3*, segundo [Ward 1998]. Os resultados gerados para os desvios são armazenados numa fila circular cujo tamanho também é um parâmetro do *RRDtool* e seu valor deve ser maior que m.

3.5. Abordagem *Single-link* versus *Network-wide* para Detecção de Anomalias

A abordagem *network-wide* se baseia numa visão completa da rede. Ela se vale dos pares "origem-destino" (OD) para indicar todos os fluxos que possuem um mesmo ponto de entrada (origem) e um mesmo ponto de saída (destino) na rede. Cada fluxo faz parte de um determinado par OD e as medidas de entropia são extraídas a intervalos de tempo regulares para todos os possíveis pares OD.

No caso do uso de pares OD, a simples coleta dos fluxos que passam por uma

determinada interface nem sempre é suficiente. Faz-se necessário saber o ponto de entrada e de saída deste fluxo no âmbito do sistema autônomo (AS), obrigando o uso de um ferramental mais complexo que considere topologia e tabelas de rota. Esta complexidade aumenta ainda mais se considerarmos a dinâmica dos protocolos de roteamento e situações de *multihoming*.

A abordagem *single-link* se restringe a analisar informações de fluxos IP colhidos para uma única interface da rede, geralmente usada por um ISP para conectar um determinado cliente ou mesmo outro ISP, possivelmente um *upstream provider*. Esta abordagem simplifica significativamente a implementação de uma solução capaz de sinalizar anomalias de um modo geral, principalmente em termos de sistemas de armazenamento e processamento. Em [Silveira 2008], é mostrado que a grande maioria das anomalias encontradas na abordagem *network-wide* é também encontrada na abordagem *single-link*. A metodologia usada neste artigo segue a abordagem *single-link*.

3.6. Arquitetura para Detecção de Ataques

A arquitetura proposta para a implementação do método emprega recursos de ferramentas de gerenciamento de redes amplamente usadas pelos centros de operações de redes WAN. A partir de uma coleta de dados *NetFlow*, estes são armazenados em arquivos e pós-processados por ferramentas comerciais ou de *software* livre, como o *Nfcapd* e o *Nfdump*. A partir destes arquivos, geralmente contendo cinco minutos de amostras de fluxos, é possível efetuar o processamento das entropias e aplicar o algoritmo estimador, tendo como resultado final séries temporais armazenadas que podem ser representadas graficamente por ferramentas usuais para visualização de séries históricas, como o já citado *RRDtool*.

4. Metodologia Experimental

Face às dificuldades para se obter amostras reais de tráfego malicioso em *backbones*, com respectivos *timestamps* devidamente identificados e num volume que impeça uma fácil verificação em gráficos de pacotes por segundo, a metodologia para a validação experimental do método proposto consiste em obter dados de ataques artificialmente gerados para que estes sejam inseridos numa seqüência de dados oriunda de um enlace de rede WAN, cujo tráfego esteja supostamente livre de ataques (aqui chamado de "tráfego de fundo"). Desta forma, é possível verificar se os ataques gerados conseguem ser corretamente detectados pelo método. No caso, para representar o tráfego de fundo foram usadas amostras de tráfego da rede Ipê, nome dado ao *backbone* acadêmico nacional operado pela RNP. Foi necessário solicitar ao Centro de Engenharia e Operações (CEO) da RNP, que mantém seus atendimentos registrados em um sistema de *trouble-ticket*, acesso às informações de fluxo relativas a uma interface que, num período de tempo qualquer, preferencialmente superior a uma semana, não apresentasse indícios de ataques.

4.1. Coleta de Fluxos na Rede Ipê

No caso da RNP, o sistema usado para coleta de fluxos é o *Nfcapd*. O *Nfcapd* recebe os registros referentes a todos os fluxos IP que estavam ativos num certo intervalo de cinco minutos e os armazena num arquivo específico. Ou seja, um novo arquivo contendo os registros dos fluxos IPs ativos no intervalo é gerado a cada cinco minutos. Não há perda de detalhamento das informações registradas pelo *Nfcapd*, o

único elemento que ocasiona perda na informação coletada é a taxa de amostragem dos pacotes dos roteadores, que captura apenas o primeiro pacote de uma seqüência contendo um número definido de pacotes que passam pela interface num dado sentido (entrada ou saída, dependendo da configuração). A taxa de amostragem empregada na rede Ipê é de 1/100 pacotes.

4.2. Obtenção das Entropias para os Fluxos Coletados

A obtenção das entropias utiliza a ferramenta *Nfdump*. Através do *Nfdump* é possível filtrar os fluxos contidos nos arquivos gerados pelo *Nfcapd* para obter informações relacionadas a IP de origem, IP de destino, porta de origem ou porta de destino, gerando arquivos separados contendo todos os valores ocorridos em cada fluxo, para cada um dos parâmetros, durante um intervalo específico. Um programa especialmente elaborado para calcular as entropias de cada parâmetro, para cada intervalo de cinco minutos, é então aplicado.

O armazenamento do *Nfcapd* permite que as entropias sejam computadas para outros intervalos de tempo, entretanto, optou-se por manter o intervalo de cinco minutos para que a granularidade dos gráficos gerados seja igual a dos gráficos utilizados pela grande maioria dos sistemas de gerenciamento de redes. Esta é também a granularidade adotada em [Lakhina 2005].

4.3. Geração Artificial de Ataques

Para a geração de amostras artificiais de ataque, foi usada uma ferramenta *web* que se vale de códigos maliciosos conhecidos para gerar ataques segundo parâmetros especificados pelo usuário. Esta ferramenta pode ser encontrada em *www.pcapr.net* e dela é gerado um arquivo descrevendo o ataque desejado. Este arquivo deve ser baixado e dele se gera um ataque real com auxílio da ferramenta *MuDos*, também encontrada em *www.pcapr.net*. O *MuDos* lê o arquivo com a descrição do ataque, gera os respectivos pacotes e os envia para a rede de destino. Através de um aplicativo de *sniffing*, como *tcpdump* ou *tshark*, faz-se a captura do ataque. É necessário tomar cuidado para não haver nenhum outro aplicativo gerando tráfego para a interface de rede monitorada, de maneira que somente o ataque esteja presente no arquivo *pcap* gerado pelo *tcpdump* ou pelo *tshark*. O passo seguinte para se obter os arquivos de fluxo é dividir o arquivo *pcap* gerado em arquivos menores, cada qual contendo cinco minutos de tráfego. Esta operação pode ser realizada através de um *script* Perl de domínio público chamado *pcap-util*. Após a geração dos vários arquivos *pcap* contendo cinco minutos de ataque, é possível aplicar amostragem de pacotes, segundo uma taxa específica, utilizando o comando *pcapdump*, que pertence ao pacote de ferramentas *pcaputils*. Após esta etapa, são executados mais dois programas para se gerar arquivos de fluxos no formato *NetFlow* versão 5: o *softflowd* e o *flowd*. O *softflowd* é capaz de ler um arquivo *pcap* e dele exportar conteúdo *NetFlow* versão 5 para um determinado IP e porta. O *flowd* é um *daemon* que escuta os pacotes *NetFlow* em uma determinada porta e os armazena em arquivo. Após a gravação de cada arquivo, executa-se o comando *flowd-reader* para ler e filtrar os parâmetros necessários para o cálculo das entropias – IPs e portas de origem e destino – gravando cada métrica em arquivos separados e mantendo os *timestamps* dos arquivos originais. Estes arquivos são então usados para se fazer a injeção, que nada mais é do que um *merge* destes com os respectivos arquivos provenientes do tráfego de fundo.

5. Resultados Obtidos para o Tráfego da Rede Ipê

5.1. Tráfego de Fundo

As amostras de tráfego da rede Ipê, representando um período de normalidade (não houve registros nem indícios de ataques massivos no período amostrado), foram obtidas do enlace de 2,5 Gbps entre BA e PE, sentido PE, das 00h do dia 20/11/2008 às 00h do dia 01/12/2008, totalizando onze dias de amostragem. A taxa média de pacotes por segundo medida para estas amostras é de 300 pps, o que significa que, devido à amostragem de 1:100, o tráfego de fundo possui taxa média real de 30 Kpps.

5.2. Ataques Artificialmente Gerados

As taxas médias de pacotes dos ataques inseridos foram ajustadas de maneira a situá-las numa faixa de 10 a 20% da taxa média de pacotes do tráfego de fundo. Estes são valores típicos onde ataques de certa magnitude não conseguem ser facilmente identificados por inspeção visual de gráficos exibindo estatísticas da taxa de pacotes.

O *TCP SYN flood* é um ataque de negação de serviço (DoS), podendo ter origem distribuída (portanto um DDoS), que procura exaurir os recursos computacionais da vítima através da criação de múltiplas seções TCP. Os parâmetros usados para gerar este ataque foram os seguintes: IPs de origem randômicos a partir de uma rede /22, portas de origem randômicas, IP de destino único e porta de destino 80/TCP, com duração de 1,5 h (5.400 segundos) e taxa de pacotes por segundo variando entre 4.000 e 6.000 pps. Estas taxas correspondem a uma variação entre 13,3% e 20% da taxa média real do tráfego de fundo. Durante a geração dos fluxos deste ataque foi usada uma taxa de amostragem de 1:100.

O *worm Slammer* é um ataque que explora uma vulnerabilidade do Microsoft SQL Server. Os parâmetros para a geração deste ataque foram os seguintes: IP de origem único, IPs de destino randomizados, portas de origem aleatória e porta de destino 1434/UDP, com duração de 2 h (7.200 segundos) e taxa de pacotes por segundo fixa em 500 pps. Foi usada uma taxa de amostragem de 1:10 durante a geração dos fluxos deste ataque, o que faz com que a taxa média considerada para este ataque corresponda a 16,6% da taxa média real do tráfego de fundo.

5.3. Uso de Filtros para Aumentar a Eficácia na Detecção de Ataques

Como a medida de entropia depende das distribuições de IPs e portas dos fluxos amostrados, um ataque de grandes proporções, cujo volume de pacotes seja muito maior que o volume médio do tráfego normal, causará uma mudança por demais abrupta nestas distribuições que será evidenciada da mesma forma na medida de entropia. Por outro lado, usando da mesma analogia, ataques de baixa intensidade podem causar variações muito sutis na medida de entropia, dificultando que o ataque seja percebido pelos métodos automáticos de detecção.

Um recurso que pode ser usado para evitar este problema é o de filtrar o tráfego monitorado de maneira a somente considerar fluxos que possuam características comuns aos ataques que se deseja identificar. Por exemplo, no caso do *worm Slammer*, pode-se aplicar o método apenas a fluxos que usem o protocolo UDP. Ou, no caso de ataques direcionados a servidores *web*, pode-se extrair a entropia apenas de fluxos direcionados à porta 80. Este recurso foi testado para as duas amostras de ataque

injetadas e os resultados encontram-se na seção 5.4.

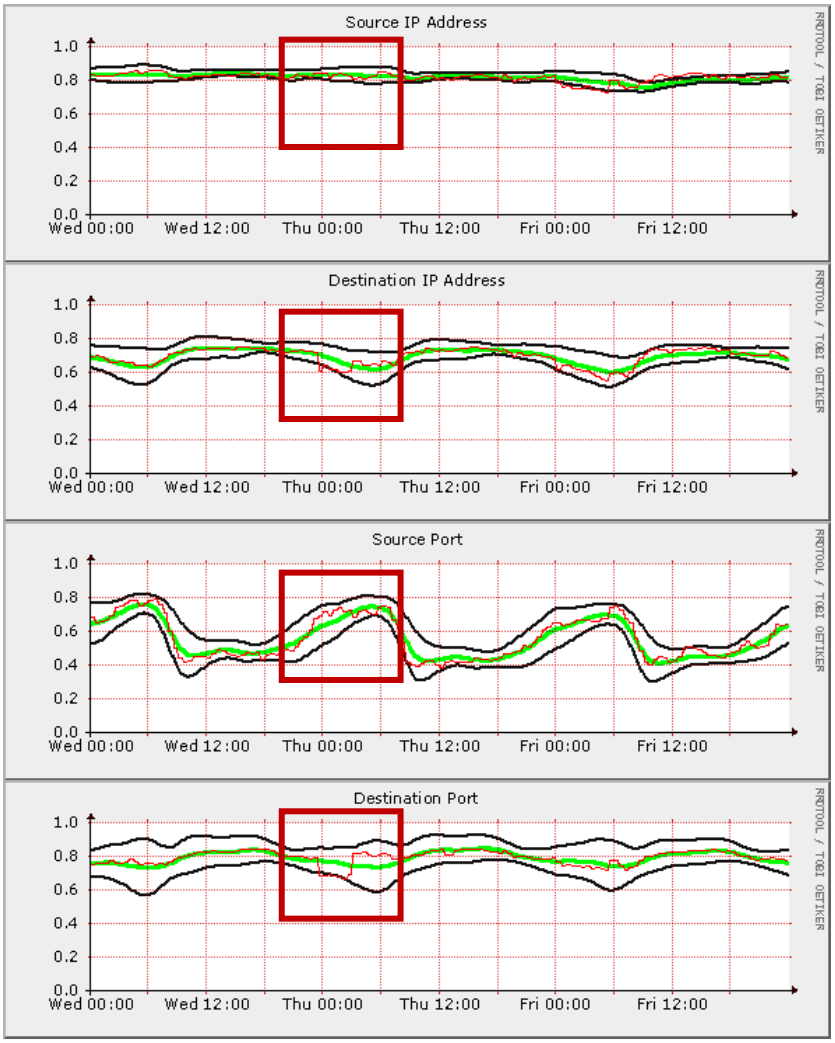

Figura 5.1 - TCP SYN *Flood* injetado no tráfego de fundo

5.4. Resultados

Em todos os testes realizados, foram usados os mesmos valores para os parâmetros da estimativa de Holt-Winters no *RRDtool*. Estes valores foram atribuídos de forma empírica, baseados em sugestões encontradas em [Brutlag 2000]. São eles: $\alpha = 0{,}01$; $\beta = 0{,}0035$; $\gamma = 0{,}01$; $\delta = 2$ e $m = 288$ (equivalente ao número de conjuntos de 5 minutos contidos em um dia). De mesma forma, todos os resultados gerados utilizaram um tamanho de fila circular igual a 1440, equivalente a cinco dias, para armazenamento do EWMA aplicado ao desvio entre a estimativa de Holt-Winters e a medida real.

Nos tempos medidos durante o processamento do método, implementado por código escrito em linguagem Python, o tempo de processamento das entropias para arquivos contendo cinco minutos de fluxo variou de acordo com o tamanho dos arquivos. Foi observada uma média de 0,21 segundo de processamento para cada arquivo da seqüência amostral do tráfego de fundo, usando para tal um *notebook* Apple

Macbook, com CPU Intel Core 2 Duo de 2.4 GHz, 4GB RAM e HD SATA externo Samsung S2, conectado por USB 2.0, com 500GB de capacidade, 5.400 RPM, 8MB cache e barramento Serial ATA/300. Conclui-se que o método proposto é aplicável para situações que requeiram gerenciamento em "tempo real".

Tendo como base as datas e horas referentes ao tráfego de fundo, todos os ataques foram injetados às 00h do dia 27/11/2008 (quinta-feira). Apesar dos dez dias de coleta, para uma melhor visualização as figuras mostram apenas uma janela que vai do dia anterior à injeção do ataque até dois dias após o final do mesmo. Entretanto, o cálculo das estimativas considerou toda a série amostral. Os limiares do critério de normalidade estão identificados pelas linhas escuras dos gráficos. A linha mais fina, de cor vermelha, indica a entropia calculada em cada gráfico. A linha mais cheia dos gráficos, de cor verde, indica a respectiva estimativa. O quadrado vermelho indica a região onde o ataque foi inserido.

5.4.1. TCP SYN *Flood* Injetado no Tráfego de Fundo

A Figura 5.1 ilustra o uso do método para a detecção do ataque *TCP SYN flood*. Conforme pode ser observado nas medidas de entropia para IP de destino e porta de destino, o ataque fica evidenciado pela forma como a entropia reduz abruptamente seu valor, indicando uma concentração destes parâmetros. A estimativa de HW acompanha a tendência normal da entropia, podendo ser usada para indicar o momento do ataque (em "Thu 00:00" nos gráficos). No entanto, a indicação baseada no EWMA do desvio (limiares do critério de normalidade), com os parâmetros adotados, não foi eficiente, sendo necessário um melhor ajuste dos mesmos.

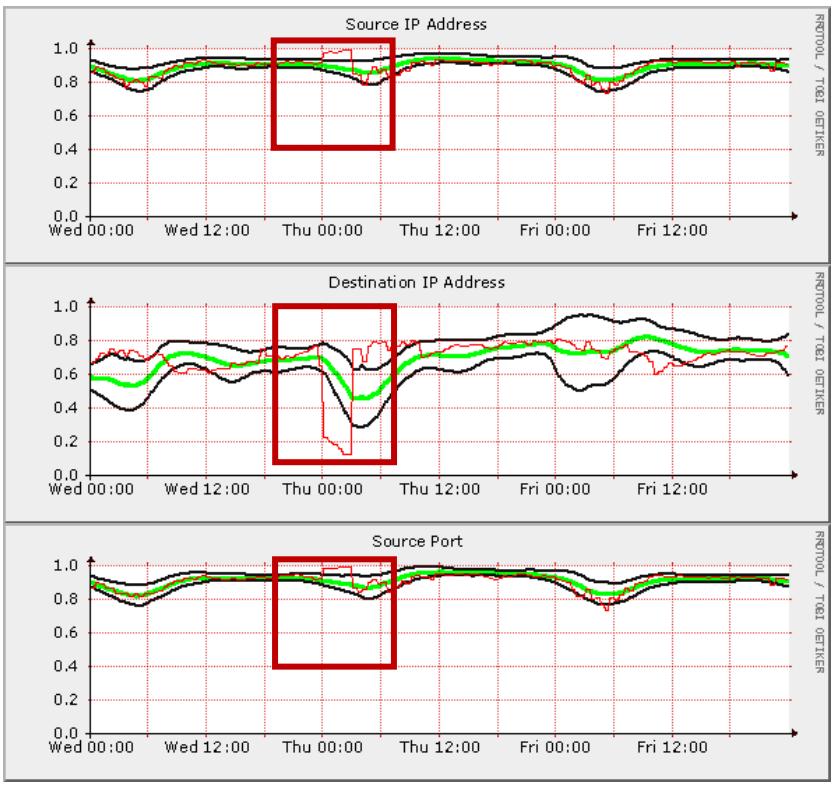

Figura 5.2 - TCP SYN *Flood* injetado, filtro pela porta destino 80

5.4.2. TCP SYN *Flood* Injetado no Tráfego de Fundo, Filtro pela Porta Destino 80

A Figura 5.2 mostra como ficam os resultados do método quando aplicado somente aos fluxos de tráfego com porta de destino 80. Neste caso, entropia da porta de destino (não exibida) será sempre zero, uma vez que há concentração máxima da entropia durante todo o período da amostra. Como houve uma filtragem específica e não houve reajustes dos parâmetros do HW, a estimativa passa a ser um pouco mais falha no caso do IP de destino. Entretanto, na hora do ataque, pode-se observar um grande pico de concentração no IP de destino e uma dispersão maior no IP e na porta de origem, tornando a ocorrência do ataque muito mais evidente e permitindo sua identificação pelo método proposto. Fica claro o aumento na eficiência do método a partir do uso do filtro.

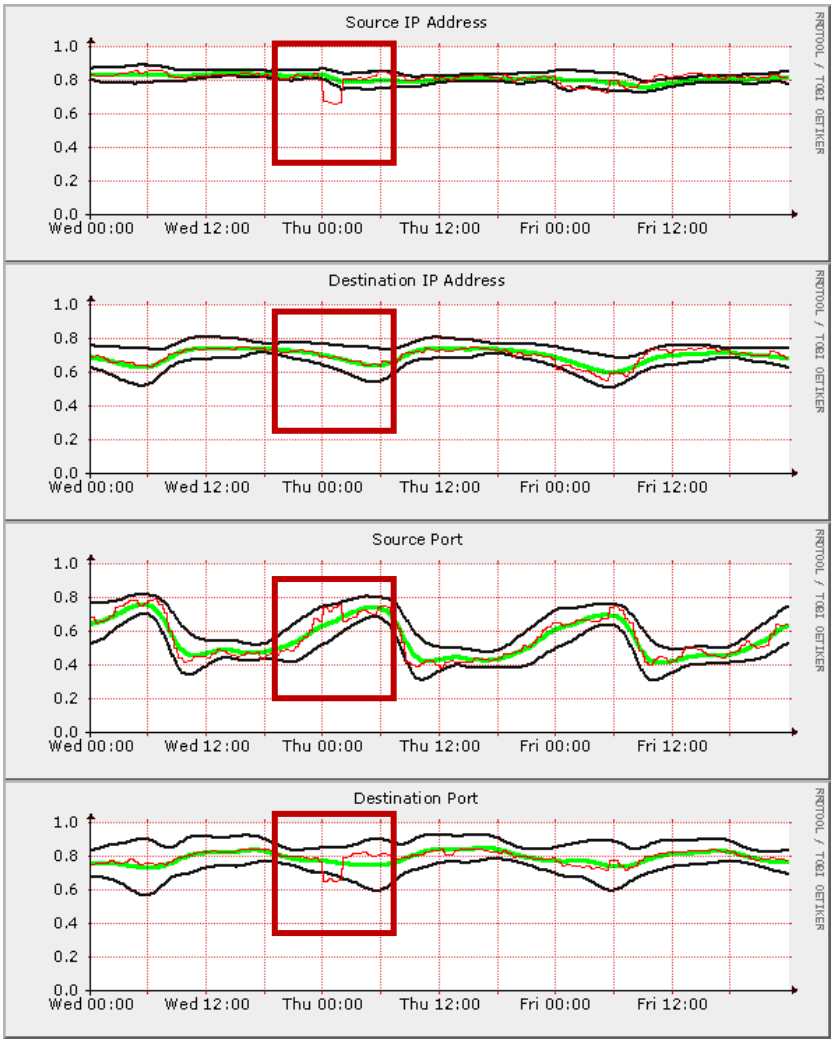

Figura 5.3 - *Worm Slammer* injetado em tráfego de fundo

5.4.3. *Worm Slammer* Injetado no Tráfego de Fundo

A Figura 5.3 ilustra o uso do método para a identificação do ataque do tipo *worm Slammer*. Conforme pode ser observado, a concentração nas medidas de entropia para IP de origem e porta de destino evidenciam o ataque, o mesmo com relação à

dispersão para porta de origem. A estimativa de HW acompanha a tendência normal da entropia, podendo ser usada para indicar o momento do ataque (em “Thu 00:00” nos gráficos). No entanto, a indicação baseada no EWMA do desvio, com os parâmetros adotados, não é tão eficiente quanto desejável, sendo necessário um melhor ajuste dos parâmetros do HW.

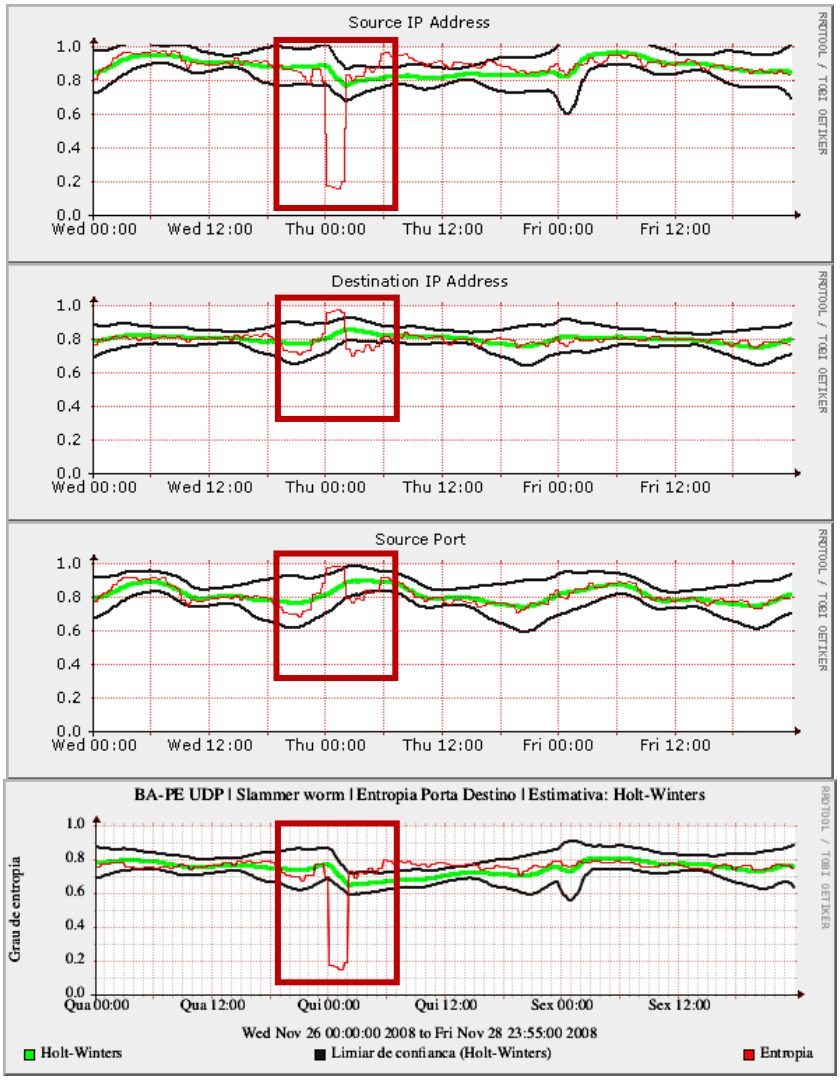

Figura 5.4 - *Worm Slammer* injetado, filtro por protocolo UDP

5.4.4. *Worm Slammer* injetado no tráfego de fundo, filtro pelo protocolo UDP

A Figura 5.4 mostra como ficam os resultados do método quando aplicado somente aos fluxos de tráfego que usam protocolo UDP. Conforme se verifica, as concentrações das entropias para IP de origem e porta de destino ficaram muito mais evidenciadas, assim como as dispersões para IP de destino e porta de origem, graças à aplicação do filtro por protocolo UDP. Da mesma forma, fica claro o aumento da eficiência do método com o uso do filtro.

5.5. Análise geral dos resultados

Os resultados apresentados confirmam a eficácia da proposta, principalmente quando se vale de filtros que buscam características conhecidas de ataques mais tradicionais, como usar o protocolo UDP ou ter a porta 80 como alvo. Fica também claro que os critérios de verificação das anomalias do método podem ser melhorados através de ajustes nos parâmetros de adaptação α, β, e γ do algoritmo Holt-Winters. Vale ressaltar que trabalhos anteriores usaram a estimativa de Holt-Winters para detectar anomalias somente em seqüências de bits por segundo ou pacotes por segundo ([Brutlag 2000]), nunca aplicado a seqüências contendo medidas de entropia.

O impacto da taxa de amostragem na precisão de métodos para detecção de anomalias, baseados em coletas de fluxos, é um assunto bastante extenso. Quanto menor a taxa de amostragem, maior a chance de não se capturar fluxos pequenos, o que impede a captura de alguns ataques e distorce o *baseline* de normalidade. Este estudo não faz parte do escopo deste trabalho. Quanto ao agrupamento das medidas em intervalos de cinco minutos, trata-se de um valor comum em sistemas de gerenciamento de redes, igualmente adotado em [Lakhina 2005]. Granularidades menores ou maiores terão o efeito de dar uma suavização menor ou maior à curva de entropia. Pode-se dizer que o maior problema com o intervalo usado é perder ataques que tenham durado bem menos que cinco minutos.

O uso de métodos de *forecasting* aplicado a séries de entropias traz vantagens em relação à aplicação em séries contendo apenas a taxa de pacotes por segundo, por exemplo. Basta ver o caso de um evento de mudança de rota devido à falha de um enlace. Todo o tráfego que passava pelo enlace que caiu será desviado para outro enlace, o que certamente causará um súbito desvio na taxa de pacotes por segundo, mas não necessariamente haverá uma forte modificação nas distribuições de IPs e portas. Certamente que isto se trata de um evento anômalo, embora não malicioso, e que costuma ser registrado pelo monitoramento SNMP. Entende-se, portanto, que uma análise baseada em medidas de entropia é mais adequada para a detecção de atividade maliciosa.

O tráfego resultante do *download* de arquivos a uma taxa elevada pode ser facilmente confundido com um DoS. Diferenciar tais situações de forma automática não é trivial. O escopo deste trabalho está em avaliar o método descrito para sinalizar uma potencial atividade maliciosa, seja ela um ataque ou não. Caberia, portanto, aos operadores da rede investigar e tomar providências.

6. Conclusão e Trabalhos Futuros

O presente trabalho analisa um método para detecção de ataques em redes WAN baseado na extração de medidas de entropia para IPs e portas de pacotes e na associação de um estimador de simples implementação capaz de se adaptar a sazonalidades presentes em séries temporais, o Holt-Winters. O método analisado usa abordagem *single-link*, mais adequada ao cenário típico de gerenciamento de redes WAN, fortemente baseado em arquitetura SNMP. O presente trabalho também propõe e analisa uma melhoria do método, que consiste da aplicação de filtros, no processo de medição, que capturem características conhecidas de ataques mais tradicionais, como o uso de protocolo UDP e o direcionamento a serviços *web* (porta de destino 80). Os resultados mostram que a aplicação destes filtros aumenta significativamente a eficácia do método. As análises aqui presentes usaram amostras reais de tráfego da rede Ipê e a geração

artificial de ataques do tipo TCP SYN *Flood* e do *worm Slammer*.

Como trabalho futuro, propõe-se o estabelecimento de critérios para um ajuste ótimo dos parâmetros da estimativa de Holt-Winters, em especial do EWMA aplicado aos desvios entre estimativa e medida real de entropia, usado para estabelecer os limites superiores e inferiores que indicam um padrão de normalidade na série temporal.

Referências

Androulidakis, G., Chatzigiannakis, V., Papavassiliou, S. (2009) "Network Anomaly Detection and Classification via Opportunistic Sampling", *IEEE Network*, Volume 23, Issue 1.

Bogaerdt, A. V. D. (2008) "RRD Tutorial", http://oss.oetiker.ch/rrdtool/tut/rrdtutorial.en.html, acessado em 08/04/2009.

Brauckhoff, D., Salamatian, K., May, M. (2009) "Applying PCA for Traffic Anomaly Detection: Problems and Solutions", *Proceedings of IEEE INFOCOM 2009*, Rio de Janeiro, BR.

Brutlag, J. D. (2000) "Aberrant Behavior Detection in Time Series for Network Monitoring", *Proceedings of the 14th Systems Administration Conference (LISA 2000)*.

Cisco Systems, Inc. (2008) "Netflow Services Solution Guide", http://www.cisco.com/en/US/docs/ios/solutions_docs/netflow/nfwhite.pdf, acessado em 08/04/2009.

Lakhina, A., Crovella, M., and Diot, C. (2005) "Mining anomalies using traffic feature distributions", *Proceedings of the ACM SIGCOMM'2005*, Philadelphia, PA, USA.

Lucena, S. C., Moura, A. S. (2009) "Análise de Estimadores EWMA e Holt-Winters para Detecção de Anomalias em Tráfego IP a partir de Medidas de Entropia", *VIII Workshop em Desempenho de Sistemas Computacionais e de Comunicação (WPerformance) 2009*, Bento Gonçalves, RS, BR.

Paschalidis, I. C., Smaragdakis, G. (2009) "Spatio-Temporal Network Anomaly Detection by Assessing Deviations of Empirical Measures", *IEEE/ACM Transactions On Networking*, Volume 17, Número 3.

Shannon, C. E. (1948) "A mathematical theory of communication", *Bell System Technical Journal*, 27:379-423 and 623-656.

Silveira, F., Diot, C., Taft, N., Govindan, R. (2008) "Empirical Evaluation of Network-Wide Anomaly Detection", Thomsom Technical Report, http://www.thlab.net/~fernando/papers/CR-PRL-2008-09-0004.pdf, acessado em 08/04/2009.

Ward, A., Glynn, P., Richardson, K. (1998) "Internet Service Performance Failure Detection", *ACM SIGMETRICS Performance Evaluation Review*, Volume 26, Número 3.

Zonglin, L., Guangmin H., Xingmiao, Y., Dan Y. (2009) "Detecting Distributed Network Traffic Anomaly with Network-Wide Correlation Analysis", *EURASIP Journal on Advances in Signal Processing*, Volume 2009, Artigo Número 2.