Análise dos Estimadores EWMA e Holt-Winters para Detecção de Anomalias em Tráfego IP a partir de Medidas de Entropia

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, 116/3902 – Botafogo
22290-906 – Rio de Janeiro, RJ

sidney@uniriotec.br, alex@rnp.br

Abstract. *To detect anomalies in wide-area network traffic is a relatively complex task. Most promising propositions are based in time series of entropy measurements to describe traffic patterns. This present work evaluates the use of traditional predictors of simple implementation, applied to entropy measurements, to signalize events that may compromise the well behavior of the network and that cannot be easily detected by commonly used network management tools. Experimental results, obtained for traffic samples of Rede Ipê, show the use of EWMA and Holt-Winters estimators on signalizing artificially injected anomalies in the traffic samples.*

Resumo. *Detectar anomalias de tráfego em redes WAN é uma tarefa de relativa complexidade. Propostas mais promissoras se baseiam em séries temporais contendo medidas de entropia para descrever padrões de tráfego. Este trabalho avalia o uso de estimadores tradicionais e de simples implementação, aplicado às medidas de entropia, para sinalizar a ocorrência de eventos que possam comprometer o bom funcionamento da rede e que não sejam facilmente detectados pelas ferramentas de gerência comumente usadas. Os resultados experimentais, extraídos para amostras de tráfego da Rede Ipê, comparam o uso dos estimadores EWMA e Holt-Winters na sinalização de anomalias artificialmente injetadas nas amostras de tráfego.*

1. Introdução

A forma como é realizado o gerenciamento das redes que trafegam grandes volumes de tráfego vem mudando substancialmente nos últimos tempos. Um dos motivadores desta mudança são os crescentes problemas de segurança que afligem os clientes destas redes, tais como ataques de negação de serviços e a proliferação de "vermes" (*worms*). Tais problemas pertencem a um conjunto maior de eventos freqüentemente denominados como "anomalias de tráfego". Entende-se por anomalia de tráfego tudo aquilo que foge a um padrão de normalidade no tráfego de pacotes de uma rede. Tais anomalias podem ser intencionais (por exemplo, ataques) ou não (por exemplo, falhas de roteamento). Qualquer que seja o caso, vários prejuízos podem ser causados de acordo com a

natureza da anomalia, seja na infra-estrutura da rede, na utilização de serviços ou no acesso destes por diversas comunidades de usuários.

Diagnosticar a ocorrência de uma anomalia em meio a um grande volume de tráfego é uma tarefa complexa. Além de questões ligadas ao processamento das informações contidas nos pacotes IP, esta complexidade advém também das diferentes causas e formatos das anomalias que podem ocorrer numa rede. As anomalias podem decorrer de erro na configuração de um protocolo, falhas no fornecimento de energia elétrica ou de ataques de negação de serviço, por exemplo. Pode-se dizer também que um dos grandes desafios para a detecção de anomalias reside na necessidade de se analisar uma grande quantidade de dados a procura de padrões. Em geral, quanto maior o volume de tráfego na rede, maior o volume de dados a ser tratado num período de tempo que se deseja ser o mais curto possível. Quanto mais curto o intervalo de tempo entre a ocorrência de uma anomalia no tráfego e a sinalização desta ocorrência aos administradores de rede, mais cedo se poderá mitigar o problema e, conseqüentemente, menor será o dano causado.

Os "ataques de segurança" são as formas de anomalia mais difíceis de se detectar e que mais causam transtornos aos operadores de rede. Podemos classificar estes ataques em alguns tipos básicos: ataques de negação de serviço (*DoS*), ataques de negação de serviço com origem distribuída (*DDoS*), infestações viróticas automatizadas (*worms*), exércitos de máquinas controladas sem autorização (*botnets*), varreduras de portas maliciosas (*port scans*) e correio eletrônico não solicitado em massa (*spam*). Há também os tráfegos repentinos – muitas vezes legítimos – e temporários (*flash crowds*), facilmente confundidos com ataques de DoS ou DDoS. Na maioria dos casos, cada tipo de anomalia possui características específicas que se diferem na maneira como cada uma se manifesta e interfere no tráfego da rede.

Uma das maiores referências na área de detecção de anomalias é o trabalho encontrado em [Lakhina 2005], onde é apresentada uma metodologia baseada em medidas de entropia, extraídas para o conjunto de endereços IPs e números de portas que circulam por uma rede a cada intervalo de 5 minutos. A partir desta análise, os autores são capazes de classificar as anomalias em uma lista de tipos, incluindo as do tipo *day-zero*, ou seja, anomalias que ainda não são conhecidas. Para tal, são usadas técnicas de estruturação e agrupamento de dados para mineração, aprendizado e classificação. A eficiência desta proposta foi aferida com sucesso usando amostras de tráfego da principal rede acadêmica dos EUA, a Abilene, e da rede acadêmica que agrega as diversas redes acadêmicas da Europa, a Géant.

Há também outras formas de detecção, para alguns tipos de anomalia, que não são recomendadas para grandes volumes de tráfego. São técnicas mais indicadas para redes locais e se baseiam na identificação de elementos que estejam explicitamente associados a algum tipo de anomalia, como uma determinada seqüência de caracteres dentro de um pacote ou uma seqüência específica de pacotes de determinado tipo ([Estevez-Tapiador 2004]). Este tipo de detecção é dito como sendo "por assinatura" e costuma ser onerosa do ponto de vista computacional, pois exige a inspeção de cada pacote IP que trafega por uma ou mais conexões de rede. Já detecções baseadas em "assinaturas estatísticas" de tráfego são mais adequadas para redes grandes ([Estevez-Tapiador 2004]). Através do comportamento dos fluxos de pacotes, é possível verificar se há ou não a presença de alguma anomalia correlata, sem que para isso seja necessário

inspecionar cada pacote IP trafegado. Este é o caso dos métodos propostos em [Lakhina 2005] e em [Zhang 2005], sendo que este último apresenta um algoritmo para eficientemente rastrear anomalias no nível da rede, refletidas em mudanças de rotas, e traz também um *framework* englobando diversos métodos para detecção de anomalias.

Tanto em [Lakhina 2005] quanto em [Zhang 2005], o arcabouço necessário para todo um processo de detecção, classificação e identificação dos fluxos anômalos é bastante complexo. Nestes trabalhos, a abordagem é do tipo *"network-wide"*, ou seja, faz-se necessário extrair informações de toda a rede para que seja possível detectar uma anomalia qualquer. Isto significa que é necessário ter visibilidade da rede como um todo. Entretanto, em [Silveira 2008], é mostrado que, surpreendentemente, a grande maioria das anomalias encontradas por métodos do tipo *network-wide* são também encontradas por métodos mais simples, do tipo *"single-link"*, ou seja, baseados em informações de apenas uma única interface. No caso dos métodos *single-link*, somente as anomalias que afetam a interface monitorada podem ser detectadas. Se, por exemplo, a interface monitorada é o único enlace de conexão da rede com a Internet, então é por este enlace que passam todos os ataques oriundos de redes externas.

O presente artigo se propõe a abordar o problema de detecção e sinalização de anomalias para um cenário de gerenciamento de redes típico de várias instituições brasileiras, principalmente aquelas ligadas a governo, como as redes acadêmicas das IFES (Instituições Federais de Ensino Superior) e outras similares na América Latina e no mundo. Deste cenário, podemos destacar as seguintes características: possuem soluções de gerenciamento de redes baseadas em *software* livre e arquitetura SNMP; monitoramento restrito às interfaces de rede dos equipamentos sob responsabilidade dos centros de gerência, na maioria destes centros não há visibilidade do sistema autônomo como um todo; histórico de estatísticas de tráfego, coletadas via SNMP, armazenado em bases RRD (*Round-Robin Database*), geralmente usando medidas extraídas a cada intervalo de 5 minutos; alarmes de ataques baseados em limiares pré-estabelecidos para a taxa de bits/s e pacotes/s nas interfaces monitoradas.

Diante deste cenário, este trabalho propõe e analisa o uso de estimadores tradicionais, e de simples implementação, aplicados às medidas de entropia extraídas para endereços IP e portas dos pacotes que trafegam numa dada interface de rede (modelo *single-link*) – tais estimadores sendo usados apenas como forma de sinalizar possíveis anomalias, mais especificamente ataques de negação de serviço. No caso presente, os estimadores analisados foram o *Exponetial Weighted Moving Average* (EWMA) e o Holt-Winters (HW). Para melhorar a eficiência do ferramental adotado pela maioria dos administradores das redes mencionadas, o método aqui proposto se valerá da ferramenta de *software* livre *RRDTool* ([Bogaerdt 2008]), amplamente usada no apoio a diversas ferramentas destinadas à gerência de redes. Até onde foi possível verificar, a combinação de estimativa de Holt-Winters, ou de EWMA, com medidas de entropia é uma abordagem nova para a detecção de anomalias em tráfegos de rede.

Para validação da proposta, foram coletadas amostras de tráfego da Rede Ipê, nome dado ao *backbone* da Rede Nacional de Ensino e Pesquisa (RNP). Esta coleta se dividiu em duas partes: uma parte registrou dez dias de tráfego do enlace entre os Estados de BA e PE, período em que não houve registro de ação maliciosa; a outra parte registrou um ataque intenso no enlace entre SC e RS, que durou aproximadamente 20 horas. De posse destas amostras, os fluxos correspondentes ao ataque foram

artificialmente ajustados e inseridos no tráfego coletado para o enlace BA-PE. Posteriormente, aplicou-se o método proposto ao tráfego contendo a injeção do ataque. Os resultados mostraram a validade do método para detecção de ataques do tipo DoS, que não necessariamente seriam verificados através da simples monitoração do tráfego de pacotes na interface de rede.

As contribuições do presente trabalho são: a proposta de uma metodologia simples para detecção de anomalias usando medidas de entropia e abordagem *single-link*, compatível com cenários mais comuns de gerenciamento de redes; uma análise comparativa entre os estimadores EWMA e HW quanto à capacidade de sinalizar anomalias nas séries temporais de medidas de entropia; a criação de um *framework* para a implementação da metodologia proposta utilizando ferramentas de *software* livre amplamente adotadas no gerenciamento de redes; e a obtenção de resultados usando dados reais da Rede Ipê.

A Seção 2 do artigo descreve a medida de entropia como forma de detecção de anomalias, a Seção 3 mostra as estimativas de Holt-Winters e EWMA aplicadas às medidas de entropia, a Seção 4 mostra os resultados obtidos para o tráfego da Rede Ipê e a Seção 5 apresenta conclusões e trabalhos futuros.

2. Detecção de Anomalias usando Medida de Entropia

Conforme apresentado por [Lakhina 2005] e [MacKey 2003], a técnica de detecção de anomalias a partir de medidas de entropia é algo relativamente recente. Em [Lakhina 2005], a entropia de Shannon é usada para medir o grau de concentração de uma distribuição de probabilidade. No caso, o enfoque está nas distribuições dos números de portas e nas distribuições dos endereços IP, tanto de origem quanto de destino nos dois casos. Estas distribuições são obtidas a partir dos fluxos de pacotes amostrados a cada intervalo de 5 minutos. As séries temporais das entropias medidas, para cada um desses parâmetros de tráfego, são então usadas para identificar a ocorrência de anomalias, assim como para identificar os fluxos anômalos.

A aplicabilidade das medidas de entropia na identificação de ataques pode ser ilustrado para o caso dos *DDoS*, ataques distribuídos de negação de serviço, nos quais diversas máquinas infectadas disparam pacotes para uma vítima específica. Durante estes ataques é possível notar a presença de um número significante de fluxos com diversos valores de IP origem e um mesmo valor de IP destino. Ou seja, tem-se uma dispersão de endereços IP de origem e uma concentração de endereços IP de destino nos fluxos IP direcionados à vítima. No caso de ataques do tipo *port scan*, uma máquina atacante, em busca de algum tipo de vulnerabilidade, faz uma varredura de diversas portas de uma ou mais máquinas numa rede. Portanto, os fluxos durante um *port scan* possuem concentração de endereços IP, tanto de origem como de destino, e dispersão de portas de destino ([Lakhina 2005]).

2.1. Entropia de Shanon

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

$$E_S = -\sum_{i=0}^{N} p_i \log_2(p_i) \quad (1)$$

onde N é o número de diferentes ocorrências no espaço amostral e p_i a probabilidade associada a cada ocorrência i. Em [Lakhina 2005], N corresponde ao número de

diferentes valores do parâmetro em análise (IP de origem, IP de destino, porta de origem ou porta de destino), que ocorreram durante um intervalo de 5 minutos, e p_i é a probabilidade de cada um desses diferentes valores *i* no intervalo medido. O resultado varia entre *0* e $log_2 N$, onde *0* indica concentração máxima na distribuição medida – ou seja, um único valor *i* ocorreu durante todo o intervalo de observação – e *1* indica máxima dispersão na distribuição medida – ou seja, uma probabilidade igual a *1/N* para todas as ocorrências dentro do intervalo de observação. Isto significa que, quanto menor o valor da entropia, mais concentrada é a distribuição e, quanto maior seu valor, mais dispersa é a distribuição.

2.2. Obtenção dos Fluxos de Pacotes

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, por default 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.

A maneira mais otimizada de se capturar os fluxos IP que passam por um roteador é usar a capacidade destes roteadores de exportar esta informação. Trata-se de um recurso muito comum em roteadores de gerações mais recentes, porém nem sempre encontrada em roteadores de pequeno porte. Os fluxos são estruturados segundo um formato padrão e exportados, usando o protocolo UDP, para uma estação coletora (ver [Cisco 2008]). Dentre os padrões usados para estruturar as informações dos fluxos, podemos citar o *Sflow* [Phaal 2001] e o *NetFlow* [Cisco 2008]. Destes, as versões 5 e 9 do padrão *NetFlow* [Claise 2004], desenvolvido e patenteado pela Cisco Systems em 1996, são as mais usadas. A versão 9 serviu como ponto de partida para um grupo de trabalho do IETF desenvolver um novo padrão, chamado IPFIX [Leinen 2004]. Embora proprietário, o *NetFlow* é um formato aberto e amplamente usado em equipamentos de diversos fabricantes. O registro *NetFlow* de cada fluxo é composto pelos seguintes campos: versão do *NetFlow*, número seqüencial, interfaces de entrada e saída no roteador, carimbos de tempo de início e fim do fluxo, número de *bytes* e pacotes observados no fluxo, endereços IP de origem e destino, portas de origem e destino, campo *protocol*, valor do campo *Type of Service* (ToS) e, nos fluxos TCP, a união de todas as *flags* TCP observadas durante o tempo de vida do fluxo. A captura destas informações é realizada para cada interface lógica do roteador mediante configuração. À medida que os fluxos expiram, os registros *NetFlow* são exportados pelo roteador para um sistema de coleta responsável por organizá-los em arquivos. Em geral, cada arquivo armazena os registros dos fluxos que estavam ativos num certo período de tempo (por exemplo, um arquivo para cada intervalo de 5 minutos).

Em geral, os roteadores são configurados para usar amostragem de pacotes como forma minimizar a carga de processamento na geração dos registros dos fluxos. Isto é muito comum em redes que concentram um volume considerável de tráfego, como nos *backbones*. Esta amostragem se dá capturando-se apenas um pacote a cada tanto que passa pela interface monitorada. Valores típicos são 1 para 100 ou mesmo 1

para 1000, dependendo do volume de tráfego. Outro motivo para se fazer amostragem é diminuir a quantidade de registros gerados e, conseqüentemente, usar menos espaço de armazenamento da estação coletora.

2.3. Abordagem *Network-Wide* versus *Single-Link*

A abordagem *network-wide* para detecção de anomalias se baseia numa visão completa da rede. Em [Lakhina 2005], utilizou-se o termo "par origem-destino" (*OD pair*) 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 realizadas para todos os possíveis pares OD. Isto significa que, para cada par OD, as entropias de IP de origem, IP de destino, porta de origem e porta de destino são calculadas a cada cinco minutos de acordo com os fluxos que passam neste par. O resultado desta operação são quatro matrizes, uma para cada parâmetro, com os pares OD num eixo, a seqüência de intervalos de 5 minutos no outro eixo e as respectivas entropias em cada posição da matriz. Toda a metodologia para identificação e classificação das anomalias, utilizada pelos autores, baseia-se nestas matrizes.

No caso do uso de pares OD, a simples informação contida nos 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) e, caso esta interface não esteja na borda do AS, será necessário um ferramental mais complexo que considere topologia e tabelas de roteamento para ser possível saber que pontos são esses. Além disso, estas tabelas mudam dinamicamente por conta de variações, comuns a qualquer rede WAN, que são reportadas pelos protocolos de roteamento (quedas de enlace, manutenções, ampliações da rede, etc). Esta investigação torna-se mais difícil ainda quando se deseja saber, por exemplo, o ponto de entrada de um fluxo em um AS *multihomed*, ou seja, que possua múltiplas conexões com um mesmo AS vizinho. Há também outros casos muito comuns onde, por exemplo, um AS possui múltiplos pontos de saída para outro AS, vizinho ou não, e várias políticas de balanceamento de rotas. Isto dificulta a estratégia para descobrir o ponto de saída de um dado fluxo IP.

Portanto, a construção das matrizes contendo a série temporal de entropias para cada par OD e, mais especificamente, a identificação do par OD para o qual um determinado fluxo pertence, adiciona um grau de complexidade razoável na implementação da solução. Torna-se necessário conhecer de antemão o ponto de entrada e de saída da rede a partir dos IPs de origem e de destino do fluxo, respectivamente, o que nem sempre é trivial. Conforme dito anteriormente, a proposta deste trabalho se restringe a analisar informações de fluxos IP colhidos de uma única interface. Esta abordagem *"single-link"* simplifica significativamente a implementação de uma solução capaz de sinalizar anomalias de um modo geral. Além disso, em [Silveira 2008], é mostrado que a grande maioria das anomalias encontradas na abordagem *network-wide* é também encontrada na abordagem *single-link.* Uma restrição desta abordagem é que somente as anomalias que afetam a interface monitorada podem ser detectadas. Todavia, em [Lakhina 2005], os autores alegam que a metodologia usando pares OD amplia bastante as possibilidades de análise, como no caso da detecção de novas classes de anomalias (por exemplo, ataques do tipo *day-zero*).

3. Estimadores Tradicionais Aplicados a Medidas de Entropia

Para se detectar um comportamento que fuja a um dado padrão histórico, faz-se necessária a construção de um *baseline* para a série analisada. Um exemplo simples seria usar a taxa média do tráfego ao longo de uma janela de tempo e seu desvio padrão para arbitrar um limiar mínimo e máximo de variação desta taxa. Qualquer novo valor fora destes limites seria considerado uma anormalidade. A seguir, são mostrados dois estimadores bastante conhecidos, o *Exponential Weighted Movin Average* (EWMA) e o Holt-Winters (HW), assim como um método tradicionalmente usado para definir os limares de normalidade.

3.1. *Exponential Weighted Moving Average* (EWMA)

Esta técnica é também chamada de aproximação exponencial, ou *exponential smoothing* [Brutlag 2000]. Trata-se de um estimador que faz uma soma ponderada entre o valor atual e um valor representando o acúmulo destas ponderações ao longo do tempo. A expressão do EWMA é:

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

onde x_t é a média histórica, X_t o valor corrente e $0 < \alpha < 1$. A constante α indica o peso que uma amostra recente tem sobre a previsão da próxima amostra. Usando de recursividade, é fácil verificar que o peso das amostras passadas, em relação a uma nova previsão, decai exponencialmente à medida que estas amostras tornam-se mais antigas. Valores típicos de α costumam ser inferiores a 0.1.

Embora muito aplicado em diversos cenários da computação (por exemplo, na estimativa de *timeouts* a partir de medidas de RTT no protocolo TCP), o EWMA não foi concebido para casos onde exista algum tipo de periodicidade ou tendência de crescimento. Para estes casos, um estimador mais apropriado é o Holt-Winters (HW) [Brutlag 2000].

3.2. Estimativa de Holt-Winters (HW)

O Holt-Winters divide a 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, por fim, um termo que expressa uma parte residual. Cada um desses três termos é tratado de forma separada através de um EWMA, ou *exponential smoothing*. Isto significa que três coeficientes, α, β e γ, devem ser atribuídos, um para cada EWMA. Entretanto, a maneira como estes termos são combinados pode refletir dois tipos de sazonalidade: a aditiva e a multiplicativa [Koehler 1999].

A sazonalidade aditiva é aquela onde o termo que representa a variação periódica da série temporal possui comportamento estatístico que independente da taxa de crescimento (positiva ou negativa) da série. A sazonalidade multiplicativa é aquela onde o termo que representa a variação periódica da série temporal possui comportamento estatístico proporcional à taxa de crescimento da série. No trabalho aqui descrito, utilizou-se apenas a forma aditiva do HW uma vez que ela apresentou melhores resultados durante os experimentos. Abaixo estão as expressões para o HW usando modelo aditivo, onde a_t corresponde à componente residual, b_t à componente de tendência, c_t à componente periódica e m é o tamanho do período:

$$x_{t+1} = a_t + b_t + c_{t+1-m} \tag{3}$$

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

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

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

Em [Brutlag 2000], o modelo aditivo da estimativa de Holt-Winters é usado para detecção de comportamentos anômalos numa série temporal contendo a taxa de bits por segundo do tráfego de saída na interface de um roteador.

3.3. Medição de Entropia para os Registros *NetFlow*

Conforme mencionado na seção 1, o cenário em que este trabalho se baseia considera sistemas de gerenciamento de redes onde o histórico das estatísticas de tráfego é armazenado em bases RRD (*Round-Robin Database*), geralmente usando medidas extraídas a cada intervalo de 5 minutos. Assim sendo, propõe-se que os registros *NetFlow* capturados tenham os valores de entropia calculados para cada intervalo de cinco minutos, e que os resultados sejam armazenados em bases do tipo RRD [Bogaerdt 2008], uma para cada parâmetro (IP de origem, IP de destino, porta de origem e porta de destino). O armazenamento em bases RRD permite que os administradores de rede tratem estas medidas a partir de ferramentas tradicionais de visualização, como o *Cacti* [Cacti 2007], ou mesmo de manipulação de bases RRD, como o *RRDtool*, possibilitando a geração de alarmes conforme os recursos destas ferramentas. Vale notar que, em [Lakhina 2005], as medidas de entropia são também calculadas para intervalos de 5 minutos.

A entropia de cada parâmetro, a cada intervalo de cinco minutos, é calculada contabilizando-se todos os pacotes de todos os fluxos registrados neste intervalo, montando-se os histogramas de cada parâmetro e aplicando-se a expressão descrita em (1). De maneira a uniformizar o grau de concentração/dispersão informado pela medida de entropia, os valores calculados são normalizados por $log_2 N$, onde N corresponde ao número de ocorrências do respectivo parâmetro para cada intervalo de cinco minutos. No caso de não se fazer a normalização, a cada intervalo de 5 minutos poderá haver um novo valor de $log_2 N$ e, portanto, um limite superior diferente dos demais para a respectiva entropia. Todavia, na prática, é esperado que estes diversos limites superiores guardem pouca diferença entre si devido ao ajuste pelo logaritmo.

Diferente do trabalho em [Lakhina 2005], deseja-se analisar um sistema cujo objetivo se resume a sinalizar que determinado tipo de anomalia, dentre algumas mais conhecidas, pode estar em curso. Posteriormente, o administrador de rede poderá usar outras ferramentas para uma investigação mais direcionada ao tipo de anomalia sinalizada. A identificação do tipo de anomalia em curso pode ser conseguida verificando-se o conjunto dos valores de entropia, calculados para os quatro parâmetros dentro do mesmo intervalo de tempo. Segue abaixo o exemplo de uma possível classificação a partir dos quatro valores de entropia:

a) *DoS*: entropia baixa para IP de origem (origem específica), entropia baixa para IP de destino (alvo específico);

b) *DDoS*: entropia alta para IP de origem (origem dispersa), entropia baixa para IP de destino (alvo específico), entropia alta para porta de origem (portas aleatórias);

c) *Port Scan*: entropia baixa para IP de destino (mesmo IP com portas sendo

varridas), entropia alta para porta de destino (muitas portas sendo varridas);

d) Proliferação de *Worms*: entropia alta para IP de origem (possível *Botnet*), entropia alta para IP de destino (procura por possíveis vítimas), entropia alta para porta de origem (várias conexões para múltiplos destinos), entropia baixa para porta de destino (explora a vulnerabilidade de alguns serviços).

4. Resultados Obtidos para o Tráfego da Rede Ipê

4.1. Processamento dos Dados dos Fluxos

Das ferramentas de *software* livre que processam informações de fluxo, uma das mais populares é o *Nfsen* [Haag 2005]. O *Nfsen* é uma interface gráfica para o *Nfdump*, que é uma ferramenta de comando de linha similar ao *TCPdump*, capaz de filtrar e extrair diversas informações de fluxos *NetFlow* armazenados pelo programa *Nfcapd* [Nfdump 2007]. É comum usar o *Nfsen* para verificar se há uma porta desconhecida dentre as *n* mais usadas, o que pode ser um indicativo de algum *worm* se alastrando. Ou ainda para olhar os fluxos IP de maior volume em busca de algum possível ataque do tipo *DoS*.

No caso da RNP, o sistema usado para coleta dos registros *NetFlow* é o *Nfcapd*. O *Nfcapd* armazena em arquivo os registros referentes a todos os fluxos IP que estavam ativos num certo intervalo de 5 minutos [Nfdump 2007]. Ou seja, a cada período de 5 minutos, um novo arquivo contendo os registros dos fluxos IPs ativos é gerado. Este processo de armazenamento não realiza nenhum tipo de média que implique na perda de detalhamento das informações registradas. O único elemento que ocasiona perda na informação coletada é a taxa de amostragem dos pacotes, que captura apenas o primeiro pacote a cada *X* pacotes que passam pela interface num dado sentido (entrada ou saída, dependendo da configuração).

O processo para obtenção das entropias começa pela utilização do *Nfdump*. Cada arquivo gerado pelo *Nfcapd* possui todos os registros referentes a todos os fluxos capturados para todas as interfaces do roteador que tenham sido configuradas para tal, isso para cada intervalo de cinco minutos. Através do *Nfdump* é possível filtrar os fluxos capturados na entrada de uma interface específica e, ao mesmo tempo, obter um dos parâmetros desejados (IP de origem, IP de destino, porta de origem e porta de destino). O comando é repetido para cada um dos quatro parâmetros, gerando arquivos separados contendo todos os valores ocorridos em cada fluxo durante um intervalo específico. A partir daí, usa-se um programa escrito para calcular as entropias de cada parâmetro para cada intervalo de cinco minutos.

A forma de armazenamento do *Nfcapd* não obriga que as entropias sejam computadas para cada período de 5 minutos. Optou-se por manter este intervalo para que a granularidade dos gráficos gerados seja igual a dos gráficos utilizados pela grande maioria dos sistemas de gerenciamento de redes. Além disso, esta é a mesma granularidade usada em [Lakhina 2005].

4.2. Amostras de Tráfego e Injeção do Ataque

Para que fosse possível um ambiente de teste controlado para verificação do método aqui proposto, adotou-se a estratégia de usar amostras de um ataque massivo e previamente confirmado, colhidas num determinado ponto da rede Ipê, para serem injetadas numa outra seqüência de amostras (neste caso, de outro ponto da rede) para a

qual não houve registro de ataque.

O ataque amostrado foi um DoS que ocorreu das 16:50h do dia 26/05/2008 às 12:40h do dia 27/05/2008 no enlace de 2,5 Gbps entre SC e RS, sentido RS. Este ataque foi informado ao Centro de Engenharia e Operações (CEO) da RNP, a pedido dos administradores da rede vitimada, para que este fosse bloqueado. Já as amostras de tráfego representando um período de normalidade (não houve indícios de ataques massivos) 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 dez dias de amostragem. A captura destas seqüências usou registros *NetFlow* versão 5 das respectivas interfaces dos roteadores modelo Juniper M40 localizados nos pontos de presença dos Estado de Pernambuco (PoP-PE) e Rio Grande do Sul (PoP-RS). Por padrão, toda coleta de fluxo na Rede Ipê usa taxa de amostragem de 1 a cada 100 pacotes e é habilitada somente na entrada das interfaces. As estatísticas de tráfego para cada um desses pontos podem ser encontradas em [CEO-RNP 2008]. Todas as informações utilizadas, incluindo os arquivos contendo os fluxos IP, foram gentilmente cedidas pelo CEO da RNP mediante solicitação formal dos autores.

Para ser possível a injeção do ataque, foi necessário um tratamento preliminar das amostras deste tráfego. Primeiramente, todos os arquivos contendo os fluxos IP do ataque continham também outros fluxos IPs referentes ao tráfego normal da interface. Portanto, foi necessário identificar os fluxos do ataque para extraí-los dos respectivos arquivos. O perfil do ataque DoS amostrado era simples: um único IP de origem, um único IP vitimado, sempre na porta 80, e uma única porta de origem. Em seguida, o número de pacotes computado para cada fluxo atacante foi multiplicado por um fator corretivo de maneira que o volume do ataque não fosse, em média, superior a 25% do volume médio do tráfego, aparentemente normal, que receberia a injeção (aqui chamado "tráfego de fundo"). Desta forma, evita-se que o ataque injetado tenha um volume de pacotes grande demais, o que o tornaria facilmente detectável por simples inspeção da taxa de pacotes do tráfego resultante. Porém, é preciso também que o volume injetado tenha destaque frente a ataques desconhecidos que possam estar escondidos no tráfego de fundo. Como se deseja um cenário onde apenas o ataque inserido seja detectado, qualquer outro ataque desconhecido e escondido no tráfego normal não pode ser destacado. Por fim, as amostras de ataque, já processadas, são adicionadas ao tráfego de fundo de forma bastante simples. Tendo como base as datas e horas referentes ao tráfego de fundo, o ataque foi injetado das 00h às 19h50 do dia 27/11/2008.

Analisar o método proposto sem o artifício da inserção artificial de uma anomalia previamente conhecida, num determinado ponto do tráfego amostrado, traria uma incerteza sobre qualquer anomalia indicada pelo método, uma vez que não seria possível afirmar ser um falso positivo ou não. Não há garantias de que todo e qualquer ataque presente na Rede Ipê tenha sido percebido e registrado pelos operadores de rede. Portanto, a estratégia usada foi inserir uma anomalia previamente confirmada num tráfego supostamente livre de anomalias para, então, verificar se o método consegue perceber a anomalia introduzida.

Vale ressaltar que, nesta metodologia de teste, não há necessidade de que o tráfego de fundo e o tráfego de ataque tenham sido extraídos para uma mesma interface de rede. O mais relevante é garantir que o tráfego de fundo seja livre de anomalias, ou aproximadamente isto, e que a anomalia inserida tenha as características desejadas,

tanto no tipo quanto no volume. Dentre os dados disponíveis, o enlace PE-BA era aquele que, no período amostrado, aparentava não ter anomalias no tráfego e também não havia registros dizendo o contrário.

Certamente que o tráfego decorrente de um usuário remoto conectado a um servidor baixando arquivos a uma taxa elevada tem características similares a um DoS. Diferenciar tais situações de forma automática é extremamente difícil. Para os objetivos deste trabalho, o importante é que o método usado sinalize rapidamente esta anomalia, seja ela um ataque ou não, cabendo aos operadores da rede investigar e tomar as devidas providências.

4.3. Cálculo das Estimativas e Critérios de Normalidade

4.3.1. Holt-Winters

A estimativa de Holt-Winters foi calculada utilizando uma função especial do *RRDtool* que realiza esta operação, conforme descrito em [Brutlag 2000]. 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 uma anomalia. O cálculo deste intervalo nada mais é do que um *exponential smoothing* para o valor de desvio, que é a diferença absoluta entre valor estimado e valor real. No caso, este *exponential smoothing* considera o ciclo sazonal da série temporal e usa o mesmo γ como coeficiente de amortização:

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

onde X_t é o valor real e x_t o valor estimado. Assim, limiares superior e inferior limitam o intervalo

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

onde δ é um fator multiplicador, geralmente com valor entre 2 e 3 (ver [Brutlag 2000] e [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.

4.3.2. EWMA

Para a estimativa EWMA, foi escrito um programa para ler os valores das bases RRD, contendo as medidas de entropia, e gerar a seqüência correspondente às predições do EWMA. Estes valores são também armazenados numa base RRD e, a partir daí, o *RRDtool* é usado para calcular os limiares superior e inferior que definem o critério de normalidade, tal qual realizado para o HW. No caso do EWMA, o valor de γ, usado em (7), passa a ser igual a α.

4.4. Resultados Obtidos

Foram usados os seguintes valores para os parâmetros do HW no *RRDtool*: $\alpha = 0{,}01$; $\beta = 0{,}0035$; $\gamma = 0{,}01$; $\delta = 2$ e $m = 288$ (equivalente ao número de conjuntos de 5 minutos contidos em um dia). Esses valores não são *default* no *RRDtool* e foram atribuídos de forma empírica, baseados em sugestões encontradas em [Brutlag 2000]. De maneira a favorecer a comparação entre o HW e o EWMA, utilizou-se o mesmo

valor de α para o EWMA. Para os resultados gerados, utilizou-se um tamanho de fila circular igual a 1440 para armazenamento do desvio, o que equivale a cinco dias. A Figura 1 traz os resultados das estimativas usando EWMA e a Figura 2 traz os resultados usando HW.

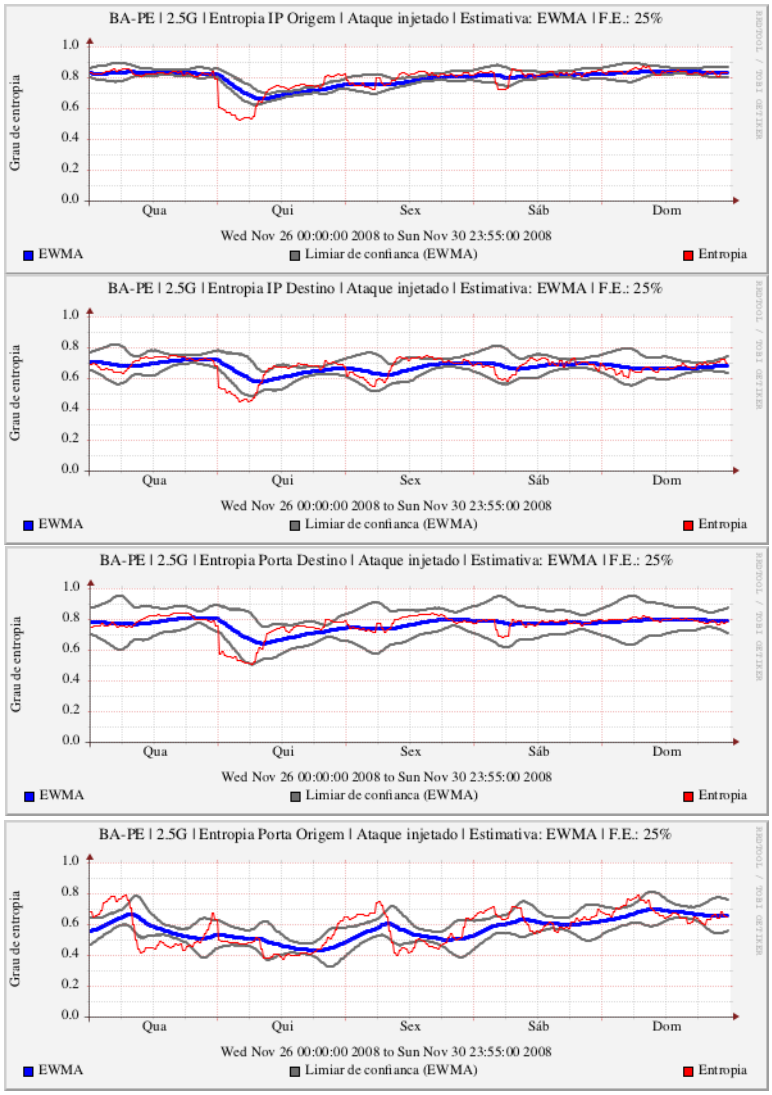

Figura 1. Estimativas das medidas de entropia usando EWMA

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 três 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 indica a respectiva estimativa: azul nos gráficos exibindo o EWMA e verde nos gráficos mostrando o HW.

Em todos os gráficos é possível verificar o início do ataque através de uma queda abrupta nos valores de entropia. Isto é coerente com os dados do ataque, já que o resultado é uma grande concentração de pacotes com mesmo IP de origem, IP de destino, porta de origem e porta de destino. O retorno à "normalidade" do tráfego de

fundo se dá cerca de 20h depois, porém de forma gradual. Interessante verificar que a variação da entropia durante este período indica uma respectiva variação na intensidade do ataque, o que é bastante comum quando este tem duração longa. Outro detalhe interessante é que, por volta das 06h de sábado, pode-se notar em todos os gráficos uma anomalia não registrada pela operação da Rede Ipê.

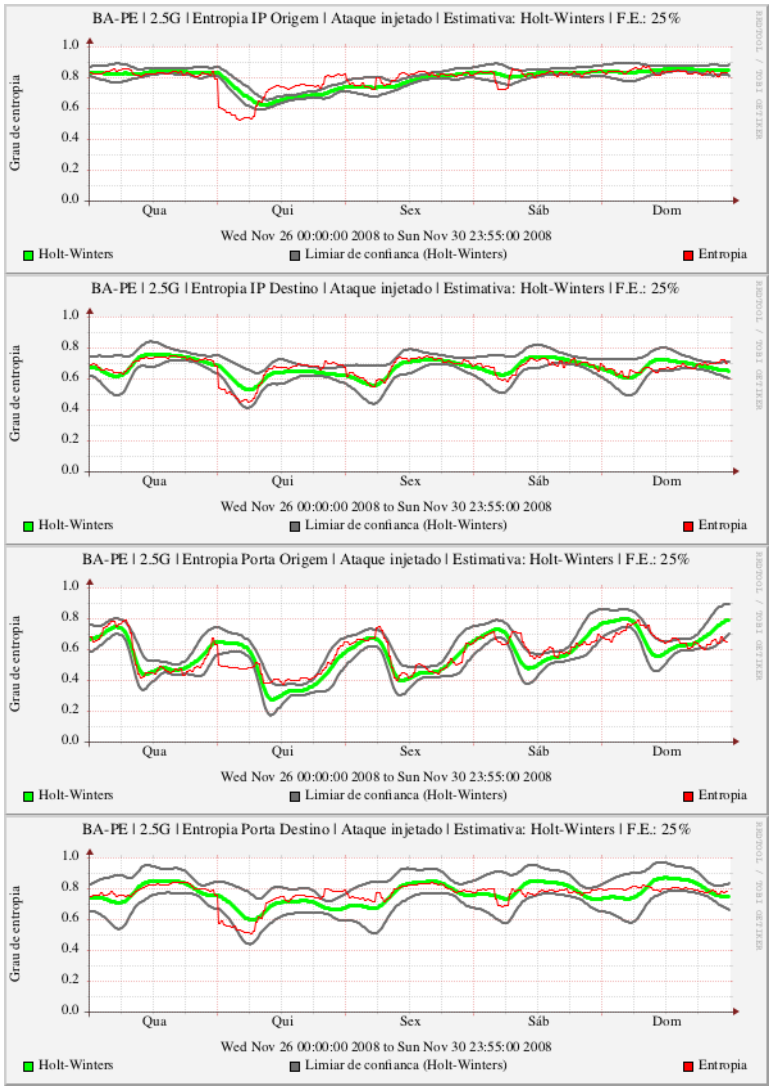

Figura 2. Estimativas das medidas de entropia usando HW

Como era de se esperar, a estimativa usando HW conseguiu aproximar melhor a sazonalidade diária das entropias medidas. Isto fica mais evidente na medida referente à porta de origem, cuja componente sazonal é mais intensa. A explicação está no comportamento dos usuários que, durante o horário comercial, acessam com mais freqüência diversos serviços em portas conhecidas. Estes serviços, por sua vez, respondem com vários pacotes contendo estes mesmos valores conhecidos na porta de origem, o que favorece a concentração. Para os demais parâmetros (porta de destino, IP de origem e IP de destino), esta sazonalidade não é tão presente, o que possibilitaria o uso do EWMA como estimador. Todavia, o HW "copia" de forma muito mais eficaz o padrão de entropia e é a única opção que se adéqua à entropia para a porta de origem.

Com relação à parametrização do HW, a estratégia de fazer γ igual a α simplifica

bastante o trabalho. α passa a ser o elemento principal de ajuste do HW, uma vez que a taxa de crescimento da série temporal, associada a β, tem muito pouca influência em cenários práticos, como o que é aqui mostrado. Os resultados apresentados mostram que o valor escolhido para α foi bastante razoável. Caso α fosse muito menor, o tempo para que a estimador volte a capturar o padrão normal de tráfego, após o término de um ataque, seria bem maior. Se α fosse muito maior, o estimador rapidamente se adequaria à presença da anomalia, dificultando a sinalização da mesma. Testes com o EWMA mostraram que, para o tráfego estudado, α igual a *0,1* é suficiente para que o estimador não consiga sinalizar a anomalia. A Figura 3 mostra como fica a predição de HW com γ diferente de α. No caso, o valor de γ aumentou para *0,1*, mas a diferença entre esta predição e a da Figura 2 é quase imperceptível.

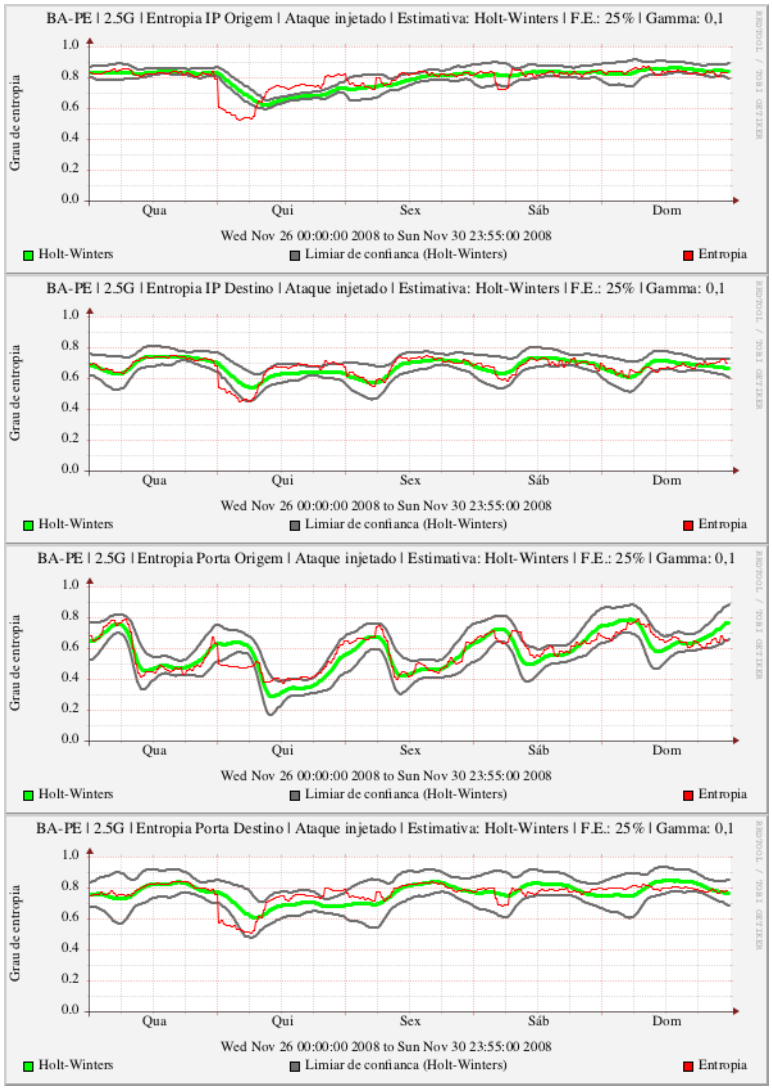

Figura 3. Estimativas das medidas de entropia usando HW com γ igual a *0,1*

5. Conclusão e Trabalhos Futuros

O presente artigo apresentou uma metodologia para detecção e anomalias baseada em medidas de entropia associada a estimadores tradicionais e de simples implementação. Foram estudados dois estimadores, o EWMA e o Holt-Winters, dos quais o HW se

mostrou mais adequado. O método proposto usa a abordagem *single-link* por esta ser mais adequada ao cenário de gerenciamento de redes tipicamente encontrado nas instituições brasileiras, principalmente aquelas ligadas a governo, e nos centros de operação das redes acadêmicas. Esta abordagem, associada ao uso do HW para identificar comportamentos anômalos, simplifica bastante sua implementação, uma vez que ela pode se valer de ferramentas de *software* livre amplamente utilizadas no gerenciamento de redes, como o *RRDtool*. No caso, o objetivo é simplesmente a sinalização de anomalias que dificilmente seriam notadas pelos métodos de gerência tradicionalmente praticados. Uma vez dado o alerta, cabe aos administradores da rede usar os instrumentos adequados para mitigar o problema.

Os resultados mostrados na seção 4.4 confirmam a eficácia do método. O experimento realizado considerou amostras de tráfego real da Rede Ipê. Amostras comprovadas de um ataque do tipo DoS, com duração aproximada de 20 horas, foram injetadas num tráfego para o qual não houve registro de ataque durante o período observado, que foi de dez dias. O método usando HW corretamente identificou a ocorrência deste ataque no tráfego amostrado. Vale salientar que o presente trabalho não se preocupou em estabelecer critérios visando um ajuste ótimo dos parâmetros usados no EWMA e no HW, mas sim uma avaliação qualitativa da eficácia do método e a comparação entre o desempenho destes estimadores.

Como trabalho futuro, deseja-se repetir os experimentos para outros tipos de ataques, mais especificamente os do tipo DDoS e de proliferação de *worms*. Além disso, deseja-se também implementar a metodologia proposta em sistemas de gerenciamento que usam código aberto.

Referências

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

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

Cacti (2007) "The Complete RRDtool-based Graphing Solution", http://cacti.net/features.php, acessado em 08/04/2009.

CEO-RNP (2008) "Operação do Backone RNP", http://www.rnp.br/ceo/, acessado em 08/04/2009.

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.

Claise, B. Ed. (2004) "RFC 3954 - Cisco Systems NetFlow Services Export Version 9", http://www.faqs.org/rfcs/rfc3954.html, acessado em 08/04/2009.

Estevez-Tapiador, J. M., Garcia-Teodoro, P., Diaz-Verdejo, J. E. (2004) "Anomaly Detection Methods in Wired Networks: a Survey and Taxonomy", *Computer Comunications*, 27, 1569-1584.

Haag, P. (2005) "Watch your Flows with Nfsen and Nfdump", *50th RIPE Meeting*, Stockholm, http://www.ripe.net/ripe/meetings/ripe-50/presentations/ripe50-plenary-

tue-nfsen-nfdump.pdf, acessado em 08/04/2009.

Koehler, A. B., Snyder, R. D., and Ord, J. K. (1999) “Forecasting Models and Prediction Intervals for the Multiplicative Holt-Winters Method”, http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/1999/wp1-99.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.

Leinen, S. (2004) “RFC 3955 - Evaluation of Candidate Protocols for IP Flow Information Export (IPFIX)”, http://www.faqs.org/rfcs/rfc3955.html, acessado em 08/04/2009.

MacKey, D. J. C. (2003) “Information Theory, Inference, and Learning Algorithms”, Cambridge University Press, Cambridge, UK.

Monsores, M. L., Ziviani, A., Rodrigues, P. S. S. (2006) “Detecção de Anomalias de Tráfego usando Entropia Não-Extensiva”, *Anais do XXIV Simpósio Brasileiro de Redes de Computadores – SBRC'2006*.

Nfdump (2007) “NFDUMP”, http://nfdump.sourceforge.net/, acessado em 08/04/2009.

Phaal, P., Panchen, S., McKee, N. (2001) “RFC 3176 - InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks”, http://www.ietf.org/rfc/rfc3176.txt, acessado em 08/04/2009.

RRDtool (2008) “Aberrant Behavior Detection with Holt-Winters Forecasting”, http://oss.oetiker.ch/rrdtool/doc/rrdcreate.en.html, acessado em 08/04/2009.

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, number 3.

Zhang, Y., Ge, Z., Greenberg, A., Roughan, M. (2005) “Network Anomography”, *Proceedings of the IMC’05*, Berkeley, CA, USA.