MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
CURSO DE MESTRADO EM SISTEMAS E COMPUTAÇÃO

NÍCOLAS ROCHA E SILVA

AVALIAÇÃO DA SENSIBILIDADE DE MÉTRICAS PARA A DETECÇÃO DE ATAQUES DE INUNDAÇÃO

Rio de Janeiro
2012

INSTITUTO MILITAR DE ENGENHARIA

NÍCOLAS ROCHA E SILVA

AVALIAÇÃO DA SENSIBILIDADE DE MÉTRICAS PARA A DETECÇÃO DE ATAQUES DE INUNDAÇÃO

Dissertação de Mestrado apresentada ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Sistemas e Computação.

Orientador: Ronaldo Moreira Salles - Ph.D

Rio de Janeiro
2012

INSTITUTO MILITAR DE ENGENHARIA
Praça General Tibúrcio, 80 – Praia Vermelha
Rio de Janeiro - RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do autor e do orientador.

003 Silva, Nícolas Rocha e.
S506a Avaliação da sensibilidade de métricas para a detecção de ataques de inundação / Nícolas Rocha e Silva; orientado por Ronaldo Moreira Salles. - Rio de Janeiro: Instituto Militar de Engenharia, 2012.

79 f. : il.

Dissertação (mestrado). - Instituto Militar de Engenharia.- Rio de Janeiro, 2012.

1. Sistemas e Computação. 2. DDoS. 3. Ataques DDoS - detecção. I. Salles, Ronaldo Moreira. II. Título. III. Instituto Militar de Engenharia.

CDD 003

INSTITUTO MILITAR DE ENGENHARIA

NÍCOLAS ROCHA E SILVA

AVALIAÇÃO DA SENSIBILIDADE DE MÉTRICAS PARA A DETECÇÃO DE ATAQUES DE INUNDAÇÃO

Dissertação de Mestrado apresentada ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Sistemas e Computação.

Orientador: Ronaldo Moreira Salles - Ph.D

Aprovada em 03 de julho de 2012 pela seguinte Banca Examinadora:

_____

Ronaldo Moreira Salles – Ph.D, do IME - Presidente

_____

Sidney Cunha de Lucena – D. C. da UniRio

_____

Anderson Fernandes Pereira dos Santos – D. C. do IME

Rio de Janeiro

2012

AGRADECIMENTOS

Agradeço a todas as pessoas que me incentivaram, apoiaram e possibilitaram esta oportunidade de ampliar meus horizontes. Em especial, desejo agradecer à minha esposa Maria Jeanedite, pelo apoio, e ao TC Ronaldo Moreira Salles e ao Maj Sérgio dos Santos Cardoso Silva, por suas disponibilidades, atenções e, principalmente, pelas relevantes observações realizadas ao longo do curso.

Agradeço, também, ao professor Sidney Cunha de Lucena e ao Alex Soares de Moura, cujas contribuições referentes à aquisição de amostras de tráfego do *backbone* da RNP e à extração de dados foram determinantes para o desenvolvimento deste trabalho.

A todos os meus amigos, que contribuíram direta ou indiretamente para realização do meu trabalho.

Por fim, a todos os professores e funcionários da Seção de Engenharia de Computação (SE/8) do Instituto Militar de Engenharia.

Nícolas Rocha e Silva

SUMÁRIO

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS

ABREVIATURAS

ASD	–	*AhnLab Smart Defense*
COSSACK	–	*Coordinated Suppression of Simultaneous Attacks*
DARPA	–	*Agência de Projetos de Pesquisa Avançada em Defesa dos Estados Unidos*
DefCOM	–	*Defensive Cooperative Overlay Mesh*
DDoS	–	*Ataques Distribuídos de Negação de Serviço*
DH	–	*Divergência de Hellinger*
EWMA	–	*Exponential Weighted Moving Average*
FNEG	–	*Falsos Negativos*
FPOS	–	*Falsos Positivos*
HW	–	*Holt-Winters*
IDMEF	–	*Intrusion Detection Message Exchange Format*
IPFIX	–	*IP Flow Information Export*
ISC	–	*Internet Systems Consortium*
QoS	–	*Quality of Service*
RNP	–	*Rede Nacional de Pesquisa*
SA	–	*Sistemas Autônomos*
WAN	–	*Wide Area Network*

RESUMO

Cada vez mais ataques distribuídos de negação de serviço (DDoS) se tornam mais sofisticados e difíceis de detectar. O emprego de métricas mais sensíveis pode permitir a detecção antecipada destes ataques, onde há um baixo percentual de pacotes maliciosos. Sendo assim, a presente dissertação tem como objetivo analisar a sensibilidade de preditores de suavização exponencial usados para detectar ataques DDoS, e de medidas de Divergência. Comparou-se a capacidade de detecção de dois preditores (EWMA e Holt-Winters), com diferentes configurações e cenários, e duas medidas de Divergência (Divergência de Hellinger e Chi-Square). Foi verificado o desempenho dessas métricas a partir das taxas de falsos positivos e falsos negativos observados. Foi inserido ataques em *traces* reais (MAWILab), *traces* simulados (DARPA) e em amostras reais de tráfego oriundas do *backbone* da RNP, com o intuito de realizar simulações de ataques com diferentes volumes de inundação. Os experimentos mostram que a otimização de parâmetros dos preditores trazem melhores resultados, e que, desde que se conheça as características do tráfego monitorado e dos enlaces observados, pode-se selecionar as métricas mais eficientes para cada cenário.

ABSTRACT

Distributed Denial of Service (DDoS) attacks are becoming more sophisticated and harder to detect. The use of sensitive metrics may provide the earlier detection of DDoS attacks, with a low rate of flooding. Therefore, the goal of this dissertation is to analyze the sensitivity of typical exponential smoothing predictors used to detect DDoS flooding attacks and measures of divergence. We compare two predictors (EWMA and Holt-Winters) and two measures of divergence (Hellinger Distance and Chi-Square). We evaluate their detection accuracy within different settings and scenarios. The performance is investigated in terms of false positive and false negative ratios. We insert attacks on real traces (MAWILab), and simulated traces (DARPA), and on real traffic samples from RNP's WAN backbone to perform simulations with different levels of flooding. The experiments show that optimal parameters of predictors provide better results. When the characteristics of monitored traffic are known, It is possible to select the most effective metrics.

1 INTRODUÇÃO

A Internet, desde a sua criação, tem se expandido a cada ano, tanto em número de usuários como em tecnologias envolvidas. De acordo com a *Internet Systems Consortium, Inc.* (ISC), em julho de 2011, foram computados 849.869.781 *host's* permanentemente conectado a essa rede (FIG. 1.1). Trata-se de um poderoso meio de comunicação que permite o compartilhamento de informações através da transferência de dados. Através desse meio, vários serviços, oferecidos por entidades governamentais ou privadas, podem ser acessados por usuários de qualquer lugar através de um dispositivo conectado à rede.

Por outro lado, o tráfego de dados maliciosos, capazes de causar os mais diversos tipos de danos, também está presente nesse meio, e que podem advir de ataques planejados. Por isso, cuidados com a confidencialidade, a autenticidade, a integridade e a disponibilidade tornaram-se ainda mais imprescindíveis, e constituem preocupações constantes tanto para empresas como para governos.

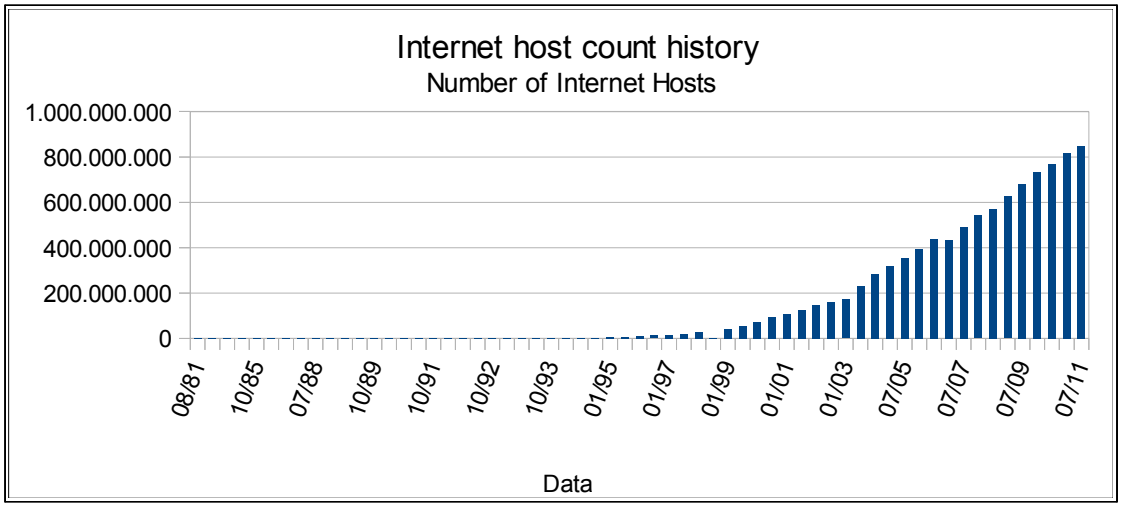

FIG. 1.1: Número de *host's* interconectados através da Internet. Fonte: (ISC, 2011)

A Internet atual é vulnerável a ataques distribuídos de negação de serviço (DDoS). Ataques desse tipo têm como objetivo fazer com que uma rede ou serviço oferecido por ela

fique inacessível aos usuários legítimos, o que geralmente é alcançado quando um atacante envia pacotes com uma taxa maior do que a vítima pode processar (CASTELÚCIO, 2009). Este é um dos diversos tipos de ataques que se aproveitam das vulnerabilidades causadas por erros de programação, presentes em serviços e protocolos empregados, que possibilitam ao invasor explorar a enorme assimetria de recursos que existe entre a Internet e a vítima.

1.1 Motivação

Cada vez mais os ataques DDoS se tornam mais sofisticados e difíceis de detectar. Por isso mesmo trata-se de uma tarefa bastante desafiadora. Embora existam estudos desde o ano de 2000, ainda não há uma resposta definitiva para a solução desse problema, de forma que ainda há espaço para contribuições.

Uma das maneiras de combater esse tipo de ataque se dá através da observação das fases que antecedem um ataque. Controlar um número expressivo de máquinas requer medidas de comando e controle para se garantir a eficiência do ataque DDoS, gerando um tráfego que pode fornecer dados importantes sobre um possível ataque.

Existem várias iniciativas de estudos, com diferentes abordagens, voltados para a detecção de *botnet's* maliciosas (WANG, 2009). Encontrar os componentes desta rede antes mesmo do ataque ocorrer constitui uma medida preventiva muito vantajosa, já que pode evitar muitos danos a partir de ações tomadas contra os agentes previamente identificados. Entretanto, quando essas identificações não ocorrem, as redes ficam vulneráveis aos ataques, caso não haja outro mecanismo de defesa. Além disso, podem ocorrer ataques de inundação sem a formação de uma *botnet*, quando, por exemplo, usuários são convocados a participar voluntariamente de um ataque organizado (JORNALNH, 2012). Sendo Assim, torna-se necessário o estudo de mecanismos eficientes de detecção de ataques DDoS.

1.2 Objetivo

São vários os trabalhos relacionados à detecção de ataques DDoS, cada qual com uma abordagem própria. Muitos deles não buscam a causa dos picos de tráfego, mas baseiam seus

estudos nos sintomas causados por eles, tais como congestionamento dos links de comunicação, sobrecarga de solicitações SYN não concluídas, ou desequilíbrio entre os tráfegos de entrada e saída (CHEN, 2007). Nesses casos, os danos causados pelos ataques, caracterizados pela negação do serviço, já foram consumados e podem ser facilmente percebidos. Por outro lado, se uma detecção for realizada antecipadamente, em pontos mais distantes da vítima, a defesa pode ser muito mais eficiente, por estar mais próxima das fontes de ataque.

As mudanças de tráfego nesses pontos mais afastados podem ser menos expressivas, devido à diversidade de origens e rotas dos pacotes maliciosos, dificultando ainda mais a sua detecção. Nesse caso, são necessários detectores mais sensíveis, capazes de perceber as flutuações mais brandas causadas pelos ataques. Como consequência, espera-se que o número de falsos positivos aumente de modo expressivo, de forma a tornar a técnica menos confiável. Entretanto, para contornar essa deficiência, sem a necessidade de diminuir o rigor da técnica empregada, pode-se adotar uma abordagem colaborativa, cujo compartilhamento de alertas permita o descarte de boa parte dos falsos positivos e a detecção antecipada do ataque em pontos longe da vítima.

A detecção em locais afastados da vítima pode exigir uma quantidade maior de locais de inspeção, dada as possíveis rotas que ligam a vítima à Internet. Uma forma de reduzir o número de detectores necessários pode ser alcançada colocando-os em pontos de concentração de fluxo, tais como as bordas de Sistemas Autônomos (SA), que são interligados por enlaces de grande capacidade.

Como o volume do tráfego nessas bordas pode ser muito alto, a quantidade de memória em disco rígido e recursos do processador demandados durante a inspeção dos pacotes tornam impeditiva tal abordagem. Entretanto, pode-se reduzir esse consumo através da inspeção de extratos que retratem o tráfego num determinado intervalo ao invés de verificar todo o conteúdo que passa pelo detector. Os roteadores de borda dos Sistemas Autônomos podem exportar esse tipo de informação através de protocolos como o *sFlow*, *NetFlow*, ou IPFIX (QUITTEK, 2004).

O objetivo do presente trabalho consiste na avaliação da sensibilidade de métricas baseadas na identificação de anomalias no tráfego, para a detecção de ataques DDoS. Foi desenvolvida uma metodologia onde foram realizadas inserções de ataques de inundação, com volume ajustável, nos *traces* adquiridos. Foi verificado o desempenho das métricas através dos índices de falsos positivos e falsos negativos observados em cada cenário, com a inserção de pacotes maliciosos em diferentes níveis de concentração.

1.3 Organização da Dissertação

Esta dissertação está organizada da seguinte forma: após a introdução, são discutidos no capítulo 2 os trabalhos relacionados. No capítulo 3 são abordados conceitos referentes à detecção de ataques DDoS. No capítulo 4, são apresentadas as características das simulações realizadas, bem como a metodologia utilizada e a avaliação dos resultados obtidos. No capítulo 5, por sua vez, são realizadas as considerações finais e apresentadas sugestões de trabalhos futuros.

2 TRABALHOS RELACIONADOS

Existem várias pesquisas cujo objetivo consiste na minimização de danos causados por um ataque DDoS ou na compreensão dos mecanismos envolvidos nesse processo (FIG. 2.1). Boa parte dos estudos concentra seus esforços na detecção e classificação de anomalias, empregando as mais diversas técnicas, como em (LAKHINA, 2005). Alguns desenvolvem métodos para a identificação dos focos de ataque, cujo objetivo é traçar a rota seguida pelos pacotes maliciosos, como em (LAW, 2002) e (DEMIR, 2010). Outros estudam o comportamento nas fases que antecedem ao ataque, a fim de identificar os vetores de ataque antes mesmo dele ocorrer, como em (CABRERA, 2001), (DITTRICH, 2008) e (FERRER, 2010).

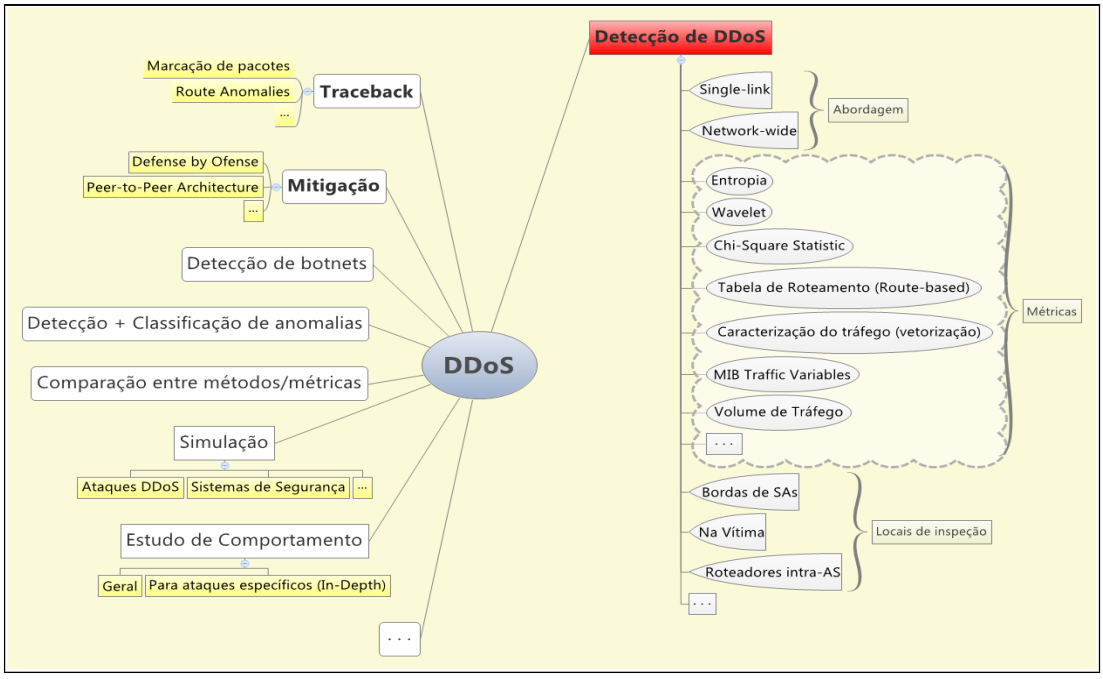

FIG. 2.1: Trabalhos relacionados a ataques DDoS

De maneira geral, o combate contra ataques DDoS requer a execução de 3 etapas: (i) identificar a ocorrência de um ataque; (ii) rastrear a origem dos pacotes maliciosos; e (iii) acionar contramedidas que contribuam para a mitigação ou eliminação dos danos causados pelo ataque, tais como filtragem e bloqueio de pacotes (CASTELUCIO, 2009).

No que diz respeito à detecção, são propostos diferentes mecanismos baseados em wavelet (KAUR, 2010), entropia (LAKHINA, 2005) (LUCENA, 2008), tabela de roteamento (PARK, 2000), *defense by ofense* (WALFISH, 2010), caracterização do tráfego (FENG, 2009), marcação de pacotes (LAW, 2002), que podem adotar uma abordagem *single-link* (LUCENA, 2008) (DEMIR, 2010) ou *network-wide* (CHEN, 2006). Algumas arquiteturas independem de base histórica (LIN, 2005), enquanto que outras têm seus parâmetros adaptados de acordo com uma *baseline* para melhor se ajustar a fatores sazonais (KLINE, 2008).

Diversos pesquisadores sugerem que a detecção deste tipo de anomalia seja realizada junto à vítima, e que os alertas, bem como o rastreamento e as contramedidas, sejam realizados no sentido contrário do fluxo, como ocorre em COSSACK (PAPADOPOULOS, 2003) e DefCOM (MIRKOVIC, 2005). Neste caso, o intuito dos gerentes de rede é proteger a sua própria rede. Entretanto, mesmo com a detecção, o ataque já pode ter comprometido a vítima de alguma maneira e a execução de qualquer medida torna-se mais difícil devido ao esgotamento de recursos. Sendo assim, é altamente desejável que a detecção do ataque DDoS ocorra o mais rápido possível, antes que a inundação torne-se generalizada (CHEN, 2007).

A detecção antecipada pode ser alcançada através da distribuição de detectores em pontos afastados (SARDANA, 2010), que não são objetivos do ataque, mas que constituem vias por onde passam os fluxos destinados à vítima. Obviamente, há um custo adicional para se manter tal arquitetura, mas que pode ser reduzido através de algumas medidas. Uma delas pode ser realizada por meio de uma abordagem colaborativa, onde o esforço conjunto de instituições pode proteger um número muito maior de redes e fortalecer o sistema como um todo, tornando-o mais viável e eficiente. Para reduzir o número de detectores necessários e ainda manter essa distribuição, basta realizar a inspeção de pacotes nos roteadores de borda de Sistemas Autônomos, já que constituem pontos de concentração de fluxo (PARK, 2000) (LIN, 2005).

Nestes roteadores a concentração de volume de tráfego é bem elevada, tornando a inspeção dos pacotes bastante custosa. Por isso é comum utilizar métodos que minimizem a carga de processamento na coleta e tratamento de dados. Normalmente, isto é feito

capturando-se apenas uma determinada percentagem dos pacotes que passam pela interface monitorada. Desta forma é possível realizar uma detecção baseada numa assinatura estatística do tráfego de rede associado a determinado tipo de anomalia (ESTEVEZ-TAPIADOR, 2004). Uma outra abordagem bem adequada para redes de *backbone*, dado o grande volume de pacotes que costuma atravessar seus roteadores, se dá através da inspeção dos fluxos de pacotes, sendo possível verificar se há ou não a presença de alguma anomalia correlata, sem a necessidade de inspecionar cada pacote IP trafegado na rede (LUCENA, 2008).

No que diz respeito ao rastreamento da origem, são vários os sistemas propostos que adotam diferentes abordagens, tais como marcação de pacotes (CASTELUCIO, 2009) (LAW, 2002) e armazenamento de resumos baseados em filtros (LAUFER, 2005) (SNOEREN, 2002). A detecção dos ataques tem papel determinante para a eficiência desses sistemas, devido à dependência existente entre essas etapas. Por conta disso, alguns trabalhos propostos já sugerem uma solução conjunta, como em (XIANG, 2011) e (CHEN, 2007).

No que diz respeito a mitigação e eliminação dos efeitos causados pelo ataque, em (WALFISH, 2010) os autores propõem uma solução baseada numa ação dos usuários legítimos provocada pela vítima, cujo intuito é diferenciá-los das máquinas pertencentes à *botnet* que esteja atacando. Em (WANG, 2009) a solução prevê o compartilhamento de informações que auxiliam na manutenção de filtros cujo objetivo consiste em bloquear tráfegos indesejados. Em ambos os casos, mecanismos de detecção também são necessários.

Apesar dos avanços na detecção do tráfego não desejado, especialmente sobre *backbones* de alta velocidade, muitas das abordagens apresentam um custo computacional elevado, requerem mudanças na infraestrutura ou mesmo apresentam resultados imprecisos. Por isso, a demanda por métodos mais eficientes justifica o desenvolvimento de estudos nessa área. Soluções baseadas na correlação de dados e agregação do tráfego em fluxos parecem ser a tendência para soluções futuras (FEITOSA, 2008).

(MOURA, 2009) propôs em sua dissertação de mestrado uma abordagem *single-link* para a detecção de anomalias em enlaces de uma rede WAN a partir da observação da entropia de fluxos de pacotes IP que passam por uma dada interface, combinado ao uso de um estimador

de comportamento para estas séries temporais, no caso a estimativa de Holt-Winters (BRUTLAG, 2000). Tal abordagem parece ser adequada também para tráfegos entre WAN's, devido às características dessa métrica e do estimador empregado. Entretanto, foram empregados neste trabalho, para os estimadores estudados, parâmetros padrão, baseados nos valores adotados em (BRUTLAG, 2000), e que, supostamente, se adequariam a qualquer tráfego e tipo de anomalia. Adotar a mesma abordagem entre SA, talvez não seja a mais eficiente. Parâmetros mais ajustados ao tráfego podem ser mais adequados, mesmo com um maior índice de falsos positivos.

(CHEN, 2007), por outro lado, propôs um novo esquema distribuído de detecção com agregação de informações através da comunicação entre vários domínios de rede e alcançou bons resultados, no que diz respeito ao número de falsos alertas e prevenção da inundação causada. Neste esquema a detecção é realizada em três camadas. Na camada mais baixa, todos os roteadores executam um algoritmo capaz de detectar flutuações suspeitas de tráfego, enviando um alerta a um servidor. Na segunda camada, fica a cargo deste servidor construir uma subárvore que mapeia dentro do SA o caminho do suposto ataque. Na camada mais alta, os servidores dos diferentes SA, que participam deste Sistema Colaborativo e formam uma rede sobreposta, comunicam-se entre si, mapeando todo o percurso do ataque.

Tais sistemas tem se mostrado promissores no combate aos ataques DDoS e, por conta disso, parece bastante plausível que a ação colaborativa de sistemas *single-link* se adeque a um cenário de detecção inter-SA, cujo o intuito é detectar ataques DDoS o mais longe da vítima possível, prevenindo antecipadamente a inundação. No ataque conhecido como *3.4 DDoS Attack* (AHNLAB, 2011), que ocorreu no dia 04/03/2011, dezenas de *websites* da Coréia do Sul, incluindo *websites* de agências financeiras, de bancos, de shoppings, de fóruns, de portais, e, principalmente, governamentais, tornaram-se alvos de um ataque bem expressivo, que mobilizou milhares de máquinas. Devido a uma ação colaborativa fornecida pelo sistema *AhnLab Smart Defense* (ASD), boa parte dos usuários tiveram garantida a continuidade do serviço apesar do ataque massivo (AHNLAB, 2011).

Entretanto, a seleção da métrica adequada pode ser determinante para a eficiência de um sistema colaborativo. Métricas mais sensíveis, como a entropia de *Shannon* e Divergência,

aplicadas nas distribuições estatísticas dos endereços IP ou dos tamanhos dos pacotes, têm sido consideradas eficazes na detecção de tráfego anormal (XIANG, 2011). Entretanto, os elevados índices de falsos positivos e a dificuldade encontrada na configuração dos parâmetros que determinam as margens de segurança são problemas que precisam ser contornados para viabilizar o seu emprego.

Parece ser bastante plausível que o compartilhamento de alertas e a correlação de dados pode colaborar fortemente para o descarte dos falsos positivos e permitir o uso de técnicas mais sensíveis para a redução do nível de falsos negativos.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Negação Distribuída de Serviço

De acordo com a cartilha de Segurança para Internet, disponibilizada pelo Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil (CERT.br), ataque DDoS (*Distributed Denial of Service*) consiste num ataque distribuído de negação de serviço, onde um conjunto de computadores é utilizado para tornar indisponíveis um ou mais serviços ou computadores conectados à Internet (CERT.BR, 2011).

Na maioria das vezes, o intuito desses ataques não é realizar uma invasão, mas impedir que usuários legítimos utilizem um determinado serviço de um computador ou rede. Para tanto, o atacante promove na vítima um aumento do consumo de recursos, tais como memória, poder de processamento, espaço em disco e, principalmente, largura de banda. Estes efeitos são provocados quando a vítima recebe uma quantidade de pacotes ou solicitações maior do que o serviço pode suportar (FEITOSA, 2008).

Um ataque DDoS baseia-se no emprego de centenas ou mesmo milhares de máquinas, normalmente comprometidas, que juntas são usadas numa ação coordenada. A quantidade de elementos empregados serve para potencializar o ataque e dificultam as ações de mitigação, mas são requeridas algumas etapas para que o ataque ocorra de forma organizada e eficiente. De forma geral, são 3 as fases: (i) "intrusão em massa", na qual ferramentas automáticas são usadas para comprometer máquinas e obter acesso privilegiado; (ii) instalação nas máquinas invadidas de softwares para a realização dos ataques DDoS; (iii) lançamento de pacotes contra uma ou mais vítimas, consolidando efetivamente o ataque (RNP, 2011).

A estrutura geral adotada nesse tipo de ataque é composta pelos seguintes elementos:

- **Atacante**: Máquina que coordena o ataque e controla os mestres;
- **Mestre**: Máquina que recebe ordens do atacante e as repassa para os escravos;

Client: Aplicação residente em cada mestre usada para receber as ordens do atacante

e repassá-las aos escravos, enviando as instruções para os respectivos *daemons*;

- **Escravo**: Máquina que recebe ordens dos mestres e gera o tráfego contra um ou mais alvos, conforme definido pelo atacante. O conjunto de escravos constitui uma *botnet;*

  Daemon: Aplicação residente em cada escravo usada para receber as ordens do mestre e executá-las, efetivando assim o ataque;

- **Vítima**: Máquina que sofre o ataque.

Essa forma de atuação contribui não apenas para a constituição de tráfegos bastante significativos, mas também para o anonimato do atacante no momento do ataque. A maneira como os elementos são organizados pode ser visualizado na FIG. 3.1.

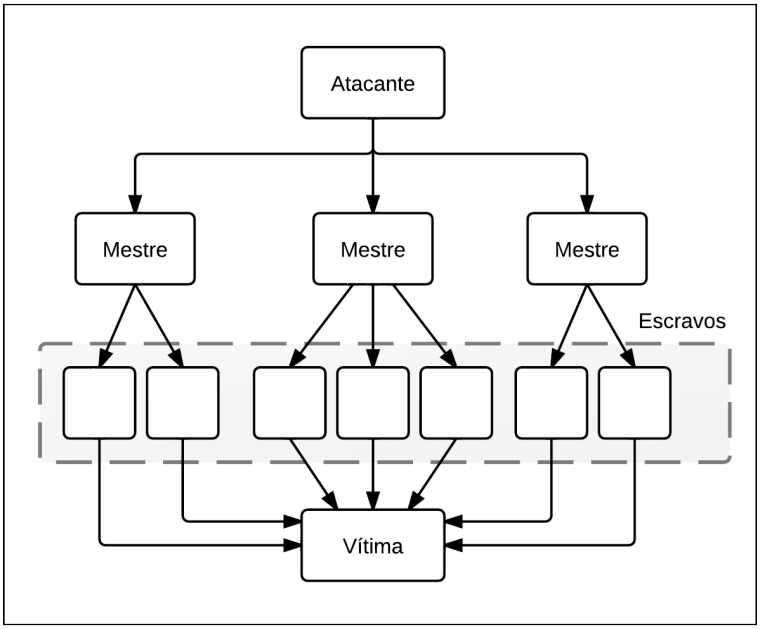

FIG. 3.1: Estrutura de um ataque DDoS

Recentemente, grupos de hackers têm convocado internautas interessados em participar de ataques de negação de serviço contra diversas instituições, muitas vezes como forma de protesto ou retaliação (LABOVITZ, 2010). Normalmente, os ataques são previamente combinados através de calendários organizados, e os programas utilizados para esse fim são disponibilizados na Internet para *download* por esses grupos (JORNALNH, 2012). Neste caso, o ataque de inundação não segue a estrutura descrita na FIG. 3.1, uma vez que os ataques são iniciados voluntariamente, não havendo a necessidade de máquinas que assumem

o papel de atacante ou de mestres. Vale salientar que milhares de usuários que realizaram o *download* desse programa tiveram suas máquinas infectadas (SYMANTEC, 2012).

Existe uma grande variedade de ataques DDoS, e que podem ser classificados de acordo com o grau de automação, vulnerabilidade explorada, mecanismos de propagação, etc. Foi proposto em (SPECHT, 2004) uma taxonomia que agrupa estes ataques em duas classes principais: ataques de esgotamento da largura de banda e ataques de esgotamento de recursos (FIG. 3.2). No primeiro caso, a inundação da rede da vítima atrapalha o acesso ao serviço, enquanto que no segundo, o consumo de recursos impede a vítima de processar requisições de usuários legítimos.

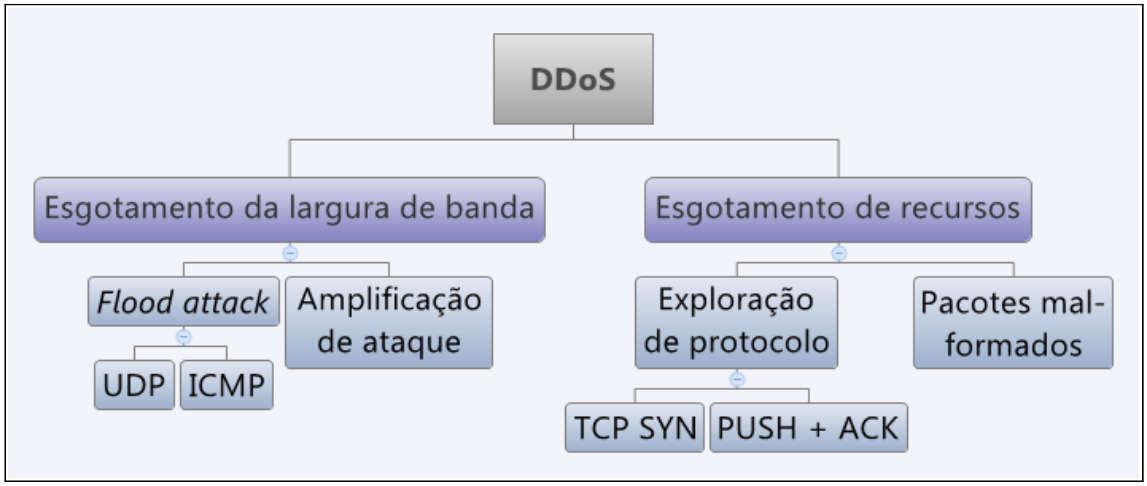

FIG. 3.2: Taxonomia de um ataque DDoS

Os ataques de esgotamento da largura de banda caracterizam-se pelo envio de grande quantia de tráfego (*Flood Attacks*), tais como *UDP Flooding* e *ICMP Flooding*, ou pela amplificação de ataque, quando os zumbis enviam pacotes para um endereço de *broadcast IP* e toda a sub-rede associada a esse endereço manda mensagens de resposta à vítima (FIG. 3.3).

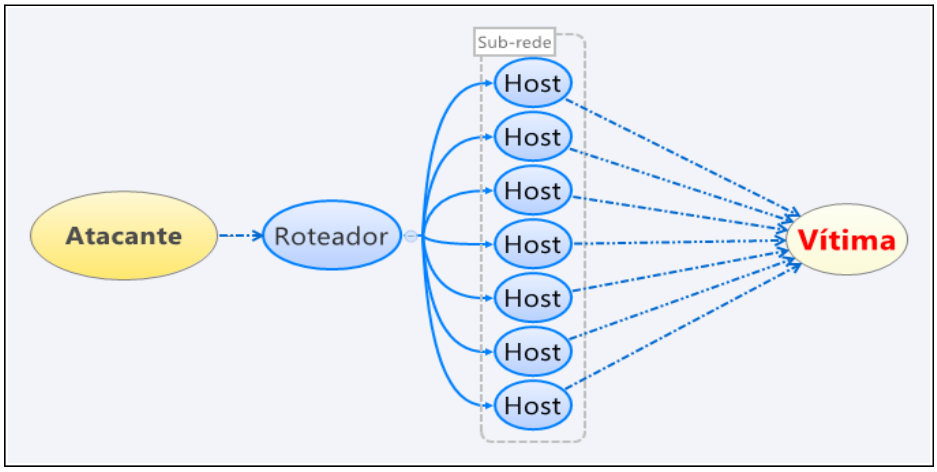

FIG. 3.3: Amplificação de ataque

No caso nos ataques de esgotamento de recurso, podem ser exploradas algumas fragilidades da pilha de protocolos TCP/IP, e por conta disso, utilizar esta modalidade contra qualquer computador conectado à Internet. O uso adulterado do TCP SYN e PUSH+ACK constituem exemplos deste modalidade de ataque. O envio de pacotes mal formados constitui outra maneira de consumir recursos da vítima. Neste caso, os pacotes IP apresentam algum tipo de incoerência em seu cabeçalho, como, por exemplo, todos os bits do campo *QoS* com o valor 1, ou mesmo o endereço de origem igual ao endereço de destino. Em todos esses casos, e em outros não citados, o aumento do número de processos ativos causado pela atividade dos zumbis podem consumir todo o poder de processamento da vítima ou mesmo derrubar o sistema. Um dos fatores que contribuem para a dificuldade na identificação dos fluxos maliciosos advém da natureza das requisições, que geralmente possuem conteúdo aparentemente legítimo e são originados a partir de conexões TCP válidas.

3.2 Métodos de Detecção

De maneira geral, combater ataques DDoS requer a execução de 3 etapas: (i) identificar a ocorrência de um ataque; (ii) rastrear a origem dos pacotes maliciosos; e (iii) acionar contramedidas que contribuam para a mitigação ou eliminação dos danos causados pelo ataque, tais como filtragem e bloqueio de pacotes (CASTELUCIO, 2009). O presente trabalho tem seu enfoque na etapa de identificação do ataque, considerando-se diferentes volumes de ataque.

Atualmente, os métodos de detecção de ataques DDoS podem ser agrupados em duas classes principais, cujas métricas se baseiam ou na assinatura de ataques ou na presença de anomalias no tráfego (XIANG, 2011). Na primeira classe, deve-se conhecer um conjunto de assinaturas, tais como padrões ou conjuntos de caracteres, presentes em pacotes maliciosos. Neste caso, além de ser necessário conhecer a assinatura do ataque, o custo computacional para identificar os ataques é bem elevado, e o seu emprego em tráfegos *backbone* torna-se inviável, a não ser que sejam empregados métodos estatísticos para reduzir o número de pacotes inspecionados. Na outra classe, alterações no comportamento da rede servem como indicação da presença de tráfego malicioso, sendo desnecessário conhecer assinaturas para a identificação de ataques.

Nas métricas baseadas em anomalias, o comportamento do tráfego de rede em condições normais serve de base para a determinação de limites que são ultrapassados na presença de ataques. Nesse caso, a principal vantagem reside na capacidade de detectar ataques desconhecidos. Podem ser baseados em limiares fixos (*threshold*) ou limites que são atualizados em tempo de execução a partir de estatísticas geradas durante o período em que ataques não são detectados (SANTOS, 2007).

Os alertas com limiares fixos são adequadas para medidas como, por exemplo, consumo de memória ou processamento, mas são limitados quando há algum tipo de variação, mesmo que esperada (sazonalidade) (MOURA, 2009). Neste caso, o número de falsos positivos pode crescer de forma indesejável. Por outro lado, em ataques mais elaborados, limites atualizados em tempo de execução podem ser treinados pelos atacantes, de forma que o sistema vai, gradualmente, interpretando como normal o comportamento anômalo da rede. Além disso, equilibrar as taxas de falsos positivos com as de falsos negativos pode não ser uma tarefa simples (XIANG, 2011).

Entretanto, várias métricas baseadas na teoria da informação têm sido propostas para tentar compensar essas limitações. A entropia de Shannon, por exemplo, fornece o grau de concentração de uma dada distribuição de valores, enquanto que a divergência mede a diferença entre duas distribuições de probabilidade. Essas técnicas têm sido consideradas efetivas na detecção de ataques DDoS, particularmente quando se avaliam as distribuições de

endereçamento IP e de tamanhos de pacotes. No corrente trabalho, será avaliado o desempenho de algumas técnicas de detecção agrupadas em duas categorias: a primeira baseada no emprego de métricas de suavização exponencial, e a segunda baseada em medidas de divergência. Será verificado o comportamento da entropia de Shannon associado a estimadores, que se encaixa na primeira categoria, e das divergências de Chi-Square e de Hellinger, que se enquadra na segunda.

3.2.1 Entropia de Shannon

Em (SHANNON, 1948) foi desenvolvida uma teoria da comunicação com o intuito de tornar melhores os projetos de sistemas de telecomunicações. Trata-se de uma medida da informação contida numa mensagem que Shannon chamou de entropia, e pode ser definida como:

$$E_S = -\sum_{i=0}^{N} p_i \log_2(p_i) \tag{3.1}$$

onde N é o número de diferentes ocorrências no espaço amostral e p_i é a probabilidade associada a cada ocorrência *i*. O resultado varia entre *zero* e $\log_2 N$, onde *zero* indica concentração máxima na distribuição medida, quando ocorre um único valor de *i*, e $\log_2 N$ indica máxima dispersão na distribuição medida, quando todas as ocorrências têm a mesma probabilidade de ocorrência.

Em (MOURA, 2009) percebeu-se que os valores de Entropia podem representar a dispersão de valores presentes numa dada distribuição, e apresentam boa sensibilidade na identificação de ataques DDoS, quando a entropia dos endereços de origem e destino observados num determinado intervalo sofrem alterações expressivas em seus valores. A concentração de pacotes com o mesmo destino tende a aumentar repentinamente, ou seja, o valor de entropia associado ao endereço IP de destino diminui de forma inesperada. Quanto à entropia dos endereços de origem, o valor tende a aumentar, uma vez que aumenta a dispersão. Este último valor está diretamente ligado ao número de atacantes ativos e a taxa de pacotes que cada um envia à vítima. Vale salientar que técnicas de *spoofing* podem ou não ser

empregadas nos endereços de origem, da maneira que o atacante achar mais conveniente para tentar mascarar a quantidade real de atacantes, bem como a sua localização, e a identificação da anomalia.

3.2.2 Divergência

Medidas de Divergência são amplamente empregadas em problemas estatísticos cujo foco reside na identificação de mudanças em séries temporais (BASSEVILLE, 1993). Existem vários tipos de funções que podem ser empregadas para medir a diferença entre dois conjuntos (LI, 1995), tais como divergência de Hellinger (SENGAR, 2008), divergência de Kullback-Leibler (BASSEVILLE, 1993), divergência *Chi-Square* (FEINSTEIN, 2003), etc.

Diferente do que ocorre no cálculo da entropia, que retrata a dispersão dos valores em uma dada distribuição probabilística, a divergência permite a comparação entre duas distribuições consecutivas, permitindo a identificação de mudanças abruptas. Quando ocorre um ataque DDoS, espera-se que o número de pacotes destinados a um determinado endereço cresça repentinamente, alterando significativamente sua distribuição de probabilidade. No corrente trabalho, será avaliada a eficiência de duas modalidades: divergência de *Chi-Square*, empregada em diversos estudos; e divergência de Hellinger, medida simétrica que permite a identificação de mudanças no início e no final dos ataques. Tratam-se de métricas amplamente empregadas na detecção de ataques DDoS.

3.2.2.1 Chi-Square

A Distância ou Divergência de *Chi-Square*, ou de X^2, constitui uma medida que retrata a diferença entre duas distribuições de probabilidade, e pode ser definida da seguinte forma.

Sejam duas distribuições discretas de probabilidade, $P = (p_0, p_1, \ldots, p_{k-1})$ e $Q = (q_0, q_1, \ldots, q_{k-1})$, onde $p_i \geq 0$, $q_i \geq 0$ e $\sum p_i = \sum q_i = 1$. Então, a divergência/distância de X^2 entre a distribuição atual P e anterior Q é dada pela expressão:

$$X^2(P\|Q)=\sum_{i=0}^{k-1}(p_i-q_i)^2/q_i \quad (3.2)$$

Onde X^2 pode assumir valores entre 0 e infinito, devido à presença do q_i no denominador. Quando $P = Q$, X^2 $(P||Q) = 0$. Quanto maior a distinção entre as distribuições, maior o valor de X^2 $(P||Q)$, de forma que, quando $P \neq Q$, X^2 $(P||Q) \approx \infty$.

Vale salientar que esta medida de divergência é assimétrica, de forma que o pico formado devido à mudança no tráfego só ocorre no início do ataque. Para contornar o problema de divisão por zero na expressão 3.2, que ocorre quando $q_i = 0$, este número pode ser substituído por um valor ε, tão pequeno quanto se queira.

3.2.2.2 Divergência de Hellinger

A Divergência de Hellinger (DH) também constitui uma forma de medir a divergência entre duas distribuições de probabilidade, independentemente da configuração de parâmetros (SENGAR, 2008). Pode ser definida da seguinte maneira:

Sejam duas distribuições discretas de probabilidade, $P = (p_0, p_1, \ldots, p_{k-1})$ e $Q = (q_0, q_1, \ldots, q_{k-1})$, onde $p_i \geq 0$, $q_i \geq 0$ e $\sum p_i = \sum q_i = 1$. Então, a DH entre a distribuição atual P e anterior Q é dada pela fórmula:

$$DH(P,Q)=\frac{1}{2}\sum_{i=0}^{k-1}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2 \quad (3.3)$$

Onde DH pode assumir valores entre 0 e 1, de forma que DH(P,Q) = 0, quando P = Q, e DH(P,Q) = 1, quando P e Q apresentam uma divergência máxima.

3.2.2.3 Limites de Segurança

O emprego de medidas de divergência não garante a identificação de um ataque DDoS. Para tanto, faz-se necessário o emprego de limites de segurança, que são ultrapassados quando

uma anomalia é gerada. No corrente trabalho, calculou-se esses limites para as medidas de divergência a partir das seguintes expressões:

$$lim_i = \mu_i + 2\sigma_i \tag{3.4}$$

$$\mu_i = \alpha\,\mu_{i-1} + (1-\alpha)\,DIV_i \tag{3.5}$$

$$\sigma_i^2 = \alpha\,\sigma_{i-1}^2 + (1-\alpha)(DIV_i - \mu_i)^2 \tag{3.6}$$

onde μ_i e σ_i representam, respectivamente, a média e o desvio padrão dos valores de divergência, atualizados através das expressões 3.5 e 3.6, cuja suavização das séries é ajustada por α. DIV_i constitui o valor de divergência calculado no instante i.

3.2.3 Estimadores

O emprego de métricas, como as descritas nas subseções acima, fornece informações relevantes sobre o tráfego de dados inspecionados e, quando utilizadas adequadamente, permitem a detecção de ataques DDoS. Para isso, a distribuição probabilística das variáveis selecionadas deve sofrer uma alteração significativa ao ocorrer uma anomalia desta natureza. Além disso, existe a necessidade de determinar os limites que separam o tráfego com ataque daquele considerado normal.

Durante a verificação de um tráfego livre de ataques DDoS, o valor da entropia referente aos endereços de destino, por exemplo, apresenta um valor diferente a cada intervalo de tempo. A partir da gravação desses valores calculados é possível montar uma série temporal de valores de entropia referente ao tráfego observado. Uma série temporal de predição, gerada com o auxílio de estimadores, pode servir como referência para o cálculo das margens de segurança que seriam ultrapassadas após o início do ataque. No presente trabalho será abordada a aplicação de dois estimadores bastante empregados, *Exponential Weighted Moving Average* (EWMA) e Holt-Winters (HW).

3.2.3.1 EWMA

Como o próprio nome diz, trata-se de um método que faz o cálculo da média móvel exponencialmente ponderada, também conhecido como suavização exponencial simples (*simple exponential smoothing*). Pode ser expressa da seguinte maneira:

$$x_{t+1} = \alpha X_t + (1 - \alpha)x_t \tag{3.7}$$

onde x_t representa a média estimada no instante t e X_t é o valor atual real. O valor de α reflete o peso conferido ao valor mais recente, e assume valores entre 0 e 1.

Note-se que, a cada iteração, as estimativas mais antigas perdem exponencialmente a influência no cálculo, sendo creditado maior o peso aos valores mais recentes. Dessa forma, o valor estimado representa uma média ponderada cujos valores mais recentes tem maior peso. Por conta disso, o traçado da série de estimativa gerada se assemelha ao traçado obtido com os valores reais. Quanto menor o valor de α, maior a suavidade no traçado da série.

3.2.3.2 Holt-Winters

Trata-se de um método de suavização exponencial tripla (*triple exponential smoothing*), que costuma ser empregado quando os dados da série apresentam tendência e sazonalidade. Para lidar com essas duas características, são utilizados mais dois parâmetros além daquele empregado na suavização exponencial simples em três equações que formam um conjunto resultante denominado *Holt-Winters* (HW). Existem dois modelos principais de HW, aditivo e multiplicativo, que tratam a sazonalidade de maneiras ligeiramente distintas (KALEKAR, 2004). Na TAB. 3.1 são apresentadas as equações empregadas nesses dois modelos.

TAB. 3.1: Equações do Modelos de Holt-Winters

Componente	HW Aditivo	HW Multiplicativo
Residual	$a_t = \alpha(X_t - c_{t-m}) + (1 - \alpha)(a_{t-1} + b_{t-1})$	$a_t = \alpha(X_t \div c_{t-m}) + (1 - \alpha)(a_{t-1} + b_{t-1})$
Tendência	$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$	$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$
Sazonalidade	$c_t = \gamma(X_t - a_t) + (1 - \gamma)c_{t-m}$	$c_t = \gamma(X_t \div a_t) + (1 - \gamma)c_{t-m}$
Estimativa	$x_{t+1} = a_t + b_t + c_{t+1-m}$	$x_{t+1} = (a_t + b_t)c_{t+1-m}$

Onde x_t representa a média estimada no instante t e Xt é o valor atual real. a_t denota a componente residual, b_t, a componente de tendência de crescimento, c_t, a componente de periodicidade da série e m representa o tamanho do período. Os parâmetros α, β e γ refletem a importância conferida a cada componente.

No corrente trabalho, foi utilizado apenas o modelo aditivo, por apresentar melhores resultados nos experimentos iniciais. Quanto aos parâmetros α, β e γ, foram empregadas duas abordagens: numa delas, utilizou-se os valores empregados em (MOURA, 2009), e noutra, foi selecionada a combinação que tornava a previsão mais ajustada, ou seja, com menor somatório de erro na estimativa.

Para as margens de confiança, assim como em (MOURA, 2009), foi utilizada a seguinte expressão de suavização exponencial simples para o erro de estimativa:

$$e_t = \gamma|X_t - x_t| + (1 - \gamma)e_{t-m} \tag{3.8}$$

Onde e_t representa o erro de estimativa no instante t. A componente atualiza-se a cada erro calculado, levando em conta os erros calculados no período anterior.

De acordo com (BRUTLAG, 2000), baseando-se na teoria de distribuição estatística e em algumas suposições, este erro deve ser multiplicado por um valor de escala δ, normalmente entre 2 e 3, para poder compor as margens de segurança. Dessa forma, os limites superior e inferior podem ser calculados a partir das seguintes expressões:

$$lim_{Sup} = x_t + \delta.e_{t-m} \tag{3.9}$$

$$lim_{Inf} = x_t - \delta.e_{t-m} \quad (3.10)$$

No corrente trabalho, para o parâmetro δ, foram utilizados os valores 2 e 3.

3.3 Características dos Tráfegos Inter-SA

Um Sistema Autônomo (SA) é um grupo conectado de redes, executados por um ou mais operadores de rede, que tem uma mesma política de roteamento claramente definida (HAWKINSON, 1996).

Sabe-se que todos os computadores que se conectam à Internet fazem parte de algum SA. Dessa forma, o tráfego de pacotes destinados a máquinas de outro SA deve passar por seus roteadores de borda, que concentram esses fluxos em enlaces inter-SA. Além disso, os SA podem permitir a passagem de tráfego de outros SA através de suas conexões. Por conta disso, o tráfego nestes enlaces tende a apresentar volumes altos, já que abrangem os fluxos de várias conexões, e a variedade de endereços de origem e destino pode ser bastante ampla, particularmente quando o SA permite a passagem de tráfegos de outros SA.

3.4 Padrão Netflow

Tráfegos inter-SA comportam volume de tráfego agregado bastante elevado. Por conta disso, torna-se computacionalmente inviável inspecionar todos os pacotes trafegados na rede. Entretanto, com o auxílio de protocolos, muitos roteadores permitem a exportação de informações relevantes do tráfego, viabilizando o seu monitoramento.

A *Cisco Systems* (CISCO, 2011) desenvolveu um protocolo proprietário aberto para a utilização em roteadores Cisco, embora também seja usado em equipamentos de outros fabricantes, conhecido como *NetFlow*. Os serviços oferecidos por este padrão permitem que administradores de rede tenham acesso a informações referentes aos fluxos IP que passam por suas redes. Os dados exportados podem ser utilizados para vários propósitos, tais como gerenciamento e planejamento de redes, mineração de dados, combate a ataques DDoS, etc.

Os roteadores compatíveis com esse padrão gravam o tráfego que passa pelas interfaces e enviam para um ou mais servidores, também conhecidos como "coletores *NetFlow*", informações detalhadas desses fluxos de dados, utilizando pacotes UDP ou SCTP (*Stream Control Transmission Protocol*).

De acordo com (CISCO, 2011), um fluxo IP é uma sequência unidirecional de pacotes que compartilham os mesmos valores para: IP de origem, IP de destino, porta de origem, porta de destino e protocolo. Além disso, caso o intervalo entre os pacotes exceda um dado valor, geralmente 15 segundos, considera-se que um novo fluxo foi iniciado. Um fluxo IP também pode ser encerrado através de pacotes RST e FIN, para conexões TCP, ou quando o tempo de vida máximo do fluxo é alcançado, geralmente 30 minutos na maioria dos casos.

4 ANÁLISE DA SENSIBILIDADE DAS MÉTRICAS

4.1 Metodologia

A metodologia utilizada para a avaliação das métricas de detecção, envolveu a observação de anomalias causadas por ataques gerados artificialmente. A ideia inicial da pesquisa consistia na realização de experimentos que retratassem ao máximo o mundo real. Por conta disso, procurou-se por *traces* de tráfegos com ataques reais, cuja identificação dos períodos com ataques DDoS fosse bem clara.

O ideal seria a obtenção de *traces* representados através de pacotes do tipo *Netflow*, já que fornecem as informações do cabeçalho IP necessárias para os cálculos envolvidos no presente trabalho, além de serem relativamente menores quando comparados a arquivos do tipo .PCAP ou .DUMP, por exemplo. Dessa forma, o esforço computacional necessário no monitoramento se reduz consideravelmente, tornando-o mais viável.

Entretanto, arquivos com essas características são difíceis de se obter. Por isso, para garantir um maior número de resultados, foi necessário trabalhar com *traces* de diferentes origens. Inicialmente, foram utilizados *traces* do tipo .DUMP, disponibilizados pelo Laboratório Lincoln, do Instituto de Tecnologia de Massachusetts (MIT) (LINCOLN, 1999) e pelo grupo de trabalho MAWI do *WIDE Project* em (MAWILAB, 2011), cada qual com suas peculiaridades. Mais tarde, foram obtidos *traces* do tipo *NetFlow*, produzidos a partir da observação do tráfego real de um dos enlaces da Rede Ipê, que é o *backbone* da Rede Nacional de Ensino e Pesquisa (RNP).

Foram estudados os desempenhos das métricas: entropia de Shannon com limites baseados em estimadores de Holt-Winters e EWMA, que apresentou bons resultados em (MOURA, 2009); e divergências de Hellinger e *Chi-Square*, com limites baseados na média móvel e no desvio-padrão móvel, que possui fácil implementação e permite a comparação entre duas distribuições vizinhas.

4.2 Dificuldades Encontradas

Na primeira etapa desta pesquisa, foram realizados experimentos com diferentes bases de dados a fim de verificar a eficácia de diferentes métodos na identificação de ataques DDoS em enlaces inter-SA.

Houve grande dificuldade para encontrar *traces* de enlaces mais robustos no formato *NetFlow*, desejáveis para a realização dos experimentos. Foram obtidos, inicialmente, arquivos .DUMP, que são relativamente grandes, uma vez que contêm toda informação dos pacotes transitados pela interface observada. Por conta disso, numa primeira tentativa de extrair as informações contidas nestes arquivos, empregou-se algumas soluções prontas para converter os arquivos .DUMP no padrão *NetFlow*. Para tanto, foi seguido o seguinte algoritmo:

ENTRADA: *FILE.dump*

1. Reprodução do tráfego presente no arquivo .DUMP através do programa *tcpreplay*;
2. Emulação de roteadores através do programa *softflowd* ou *fprobe* – estes programas registram os fluxos recebidos por uma interface, geram pacotes do tipo *Netflow* e os enviam para o endereço escolhido;
3. Armazenamento dos arquivos *Netflow* através do programa *nfcapd* – este assume o papel do servidor que recebe os pacotes *Netflow* provenientes dos roteadores de borda, monta as séries temporais e armazena os dados coletados.

SAÍDA: arquivo no formato *nfcapd.YYYYMMDDhhmm*

Entretanto, houveram problemas na execução desse algoritmo. Foram verificadas várias ocorrências de descarte de pacotes, devido ao volume de tráfego gerado. Além disso, tanto o *softflowd* como o *fprobe* encerram sua execução repentinamente.

Por conta disso, modificou-se a forma de extrair os dados contidos nos *traces* e eliminou-se a etapa de conversão de *traces* para pacotes do tipo *Netflow*, como visto na FIG. 4.1.

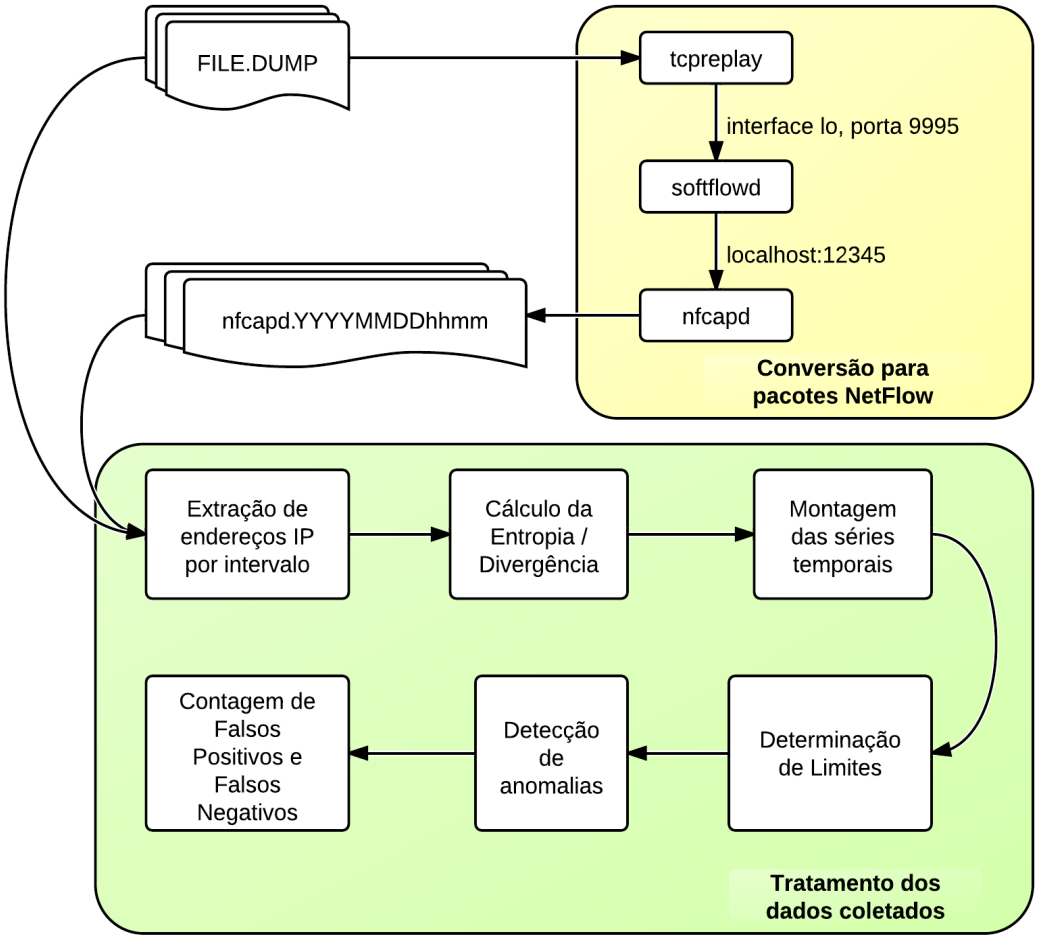

FIG. 4.1: Algoritmo para extração e tratamento de dados

Dessa forma, a extração dos endereços IP foi realizada diretamente dos *traces* através do programa *Wireshark*, versão 1.4.4, que permite a exportação de diversos campos contidos no trace. Para o tratamento dos dados coletados, foi necessário implementar em C++ um programa capaz de realizar todos os cálculos referentes à montagem das séries temporais, desde a agregação dos endereços contidos no mesmo intervalo considerado até a contabilização de falsos positivos (FPOS) e falsos negativos (FNEG). Essa linguagem foi escolhida pelo fato de fornecer boa performance na execução do programa. Também foi implementada uma função para a seleção de parâmetros ótimos para EWMA e Holt-Winters, explicada na seção 4.4. Verificou-se que o custo computacional necessário em cada atualização das séries temporais foi relativamente baixo para todos os experimentos, por volta de 1 segundo em um computador pessoal comum.

No corrente estudo, as únicas informações extraídas dos *traces*, relevantes para a identificação de anomalias, foram os endereços IP e o instante em que cada pacote foi coletado. Com esses dados as séries temporais podem ser montadas, baseando-se na distribuição de probabilidade em cada intervalo considerado. Assim como em (MOURA, 2009), apenas para os cálculos de entropia, os valores obtidos foram normalizados por $log_2 N$, com o intuito de uniformizar o grau de dispersão observado, onde N corresponde ao número de endereços presentes em cada intervalo.

Para as métricas e limites estudados, existem parâmetros configuráveis que influenciam diretamente na utilidade dos mesmos. No caso do Holt-Winters, que possui o maior número de variáveis, são três os parâmetros: α, β e γ, explicados na Seção 3.2.3.2. Entretanto, percebeu-se durante a codificação que, além disso, a escolha dos valores iniciais de a_0, b_0 e c_0, três componentes que representam o resíduo, a tendência e a sazonalidade, respectivamente, também influencia de maneira expressiva nos valores estimados e, consequentemente, nos limites gerados. Portanto, no caso do estimador HW, tornou-se necessária a configuração de seis variáveis para uma estimativa mais precisa.

4.3 Traces

Foram utilizadas as informações de *traces* reais (MAWILab), *traces* simulados (DARPA) e de amostras reais de tráfego oriundas do *backbone* da RNP. A partir desses *traces*, foram montadas séries temporais das divergências e das entropias dos endereços de destino em intervalos fixos de 5 minutos, para DARPA e RNP, e de 1 segundo, para MAWILab. Na TAB. 4.1 são apresentadas algumas características desses *traces*.

Devido ao pequeno período de observação disponibilizado nos *trace* do MAWILab, 15 minutos por dia, foi necessário realizar o estudo com séries temporais de 1 segundo. Além disso, verificou-se uma sazonalidade nos *traces* da RNP e do DARPA, não presente nos *traces* do MAWILab. Também foi observado que apenas o RNP possui tráfego sem interrupção, o que pode permitir transições mais suaves entre os dias.

TAB. 4.1: Características dos *traces* utilizados

		BASES DE DADOS		
		DARPA	**RNP**	**MAWILab**
Período de observação	**Datas**	De 01Mar99, 8:00h, a 06Mar99, 6:00h	De 20Nov08, 0:00h, a 27Nov08, 0:00h	07Jun10, de 14:00h a 14:15h e 06Dez10, de 14:00h a 14:15h
	Tempo por dia	22 horas	24 horas	15 minutos
	Total	5 dias (de 22 h)	7 dias (ininterruptos)	30 min (07Jun10 e 06Dez10)
Intervalo		5 minutos	5 minutos	1 segundo
Capacidade do enlace		10 Mbps	2.5 Gbps	150 Mbps
Amostragem		Sem amostragem	1:100	Sem amostragem
Formato		.tcpdump	*NetFlow* versão 5	.dump
Taxa média real		20 Kpps	30 Kpps	48 Kpps
Observação		Tráfego entre uma rede local e a internet, gerado artificialmente.	Tráfego intra-SA supostamente livre de ataques.	Tráfego em um enlace que liga EUA e Japão (trans-pacífico).

4.3.1 DARPA

O Laboratório Lincoln do Instituto de Tecnologia de Massachusetts, juntamente com a Agência de Projetos de Pesquisa Avançada em Defesa dos Estados Unidos (DARPA), desenvolveu experimentos envolvendo uma rede de computadores de uma base aérea americana, cujo objetivo consistia no fornecimento de dados para o estudo de detecção de invasão de redes. Foram realizadas simulações nos anos de 1998, 1999 e 2000, com tráfegos e ataques gerados artificialmente. Os ataques inseridos foram classificados em quatro grupos: dos - *denial of service*; r2l - *remote to local*; u2r - *user to root*; *probe* - levantamento de vulnerabilidades; e *data* - acesso não autorizado a dados (HAINES, 2001).

Para efeito de estudo, foram utilizados *traces* referentes às simulações realizadas no ano de 1999, constituída por três semanas sem ataques e duas semanas com ataques. Além desses *traces*, produzidos através da ferramenta tcpdump, foram disponibilizados arquivos XML com informações sobre os ataques, dentre os quais, os endereços IP, o início e a duração.

Apesar da variedade de ataques inseridos nessas simulações, poucos são apropriados para a avaliação das métricas estudadas no que diz respeito à detecção de ataques DDoS. Em (SANTOS, 2009), por exemplo, apenas os ataques caracterizados pela sobrecarga de rede foram considerados: *apache2, smurf, warezmaster e udpstorm*. Além disso, boa parte das ocorrências de ataques com grande volume de tráfego e de atacantes são de curta duração. Das quatorze ocorrências da semana 5, por exemplo, apenas quatro possuem duração maior que cinco minutos, quatro são de apenas um segundo, e, das seis restantes, quatro delas não chegam a um minuto de duração. Por conta disso, no corrente trabalho, forma empregados apenas os *traces* da semana 1 e 3 do ano de 1999, que não possuem ataques.

Embora esses *traces* não sejam típicos de um enlace inter-SA, além de serem relativamente antigos, tratam-se de *traces* amplamente citados em trabalhos relacionados à detecção de ataques. São dados referentes a cinco semanas de tráfego com tempo de captura diária de aproximadamente 22 horas.

4.3.2 MAWILab

(MAWILab, 2011) é uma base de dados criada com o intuito de ajudar pesquisadores na avaliação de métodos de detecção de anomalias de tráfego. Através de uma metodologia própria, baseada em grafos, realiza-se uma comparação e combinação de diferentes detectores de anomalia para a criação de arquivos com a identificação de anomalias de tráfego contidas em alguns *traces* mantidos pelo grupo de trabalho MAWI, integrante do *WIDE Project* (WIDE, 2011). A base de dados é atualizada diariamente, permitindo a inclusão de novos tráfegos e métodos de identificação de anomalias.

Atualmente, no MAWILab, existem amostras de tráfegos de 2001 até 2011. Entretanto, cada uma possui apenas 15 minutos, iniciadas sempre às 14:00h, o que pode tornar sua utilização mais limitada. A coleta dos *traces* é realizada num enlace de 150 Mbps localizado entre Japão e EUA. Devido ao baixo volume dos ataques presentes nos *traces*, os mesmos foram considerados como livres de ataques, para fins de estudo.

4.3.3 RNP

Foram obtidos registros no formato *NetFlow* versão 5, referentes a 07 (sete) dias de observação, no ano de 2008, de um enlace de 2,5 Gbps da Rede Ipê (FIG. 4.2), que é uma infraestrutura de rede Internet operada pela RNP e desenvolvida para atender as principais universidades e institutos de pesquisa do país, que se beneficiam de um canal de comunicação rápido e com suporte a serviços e aplicações avançadas.

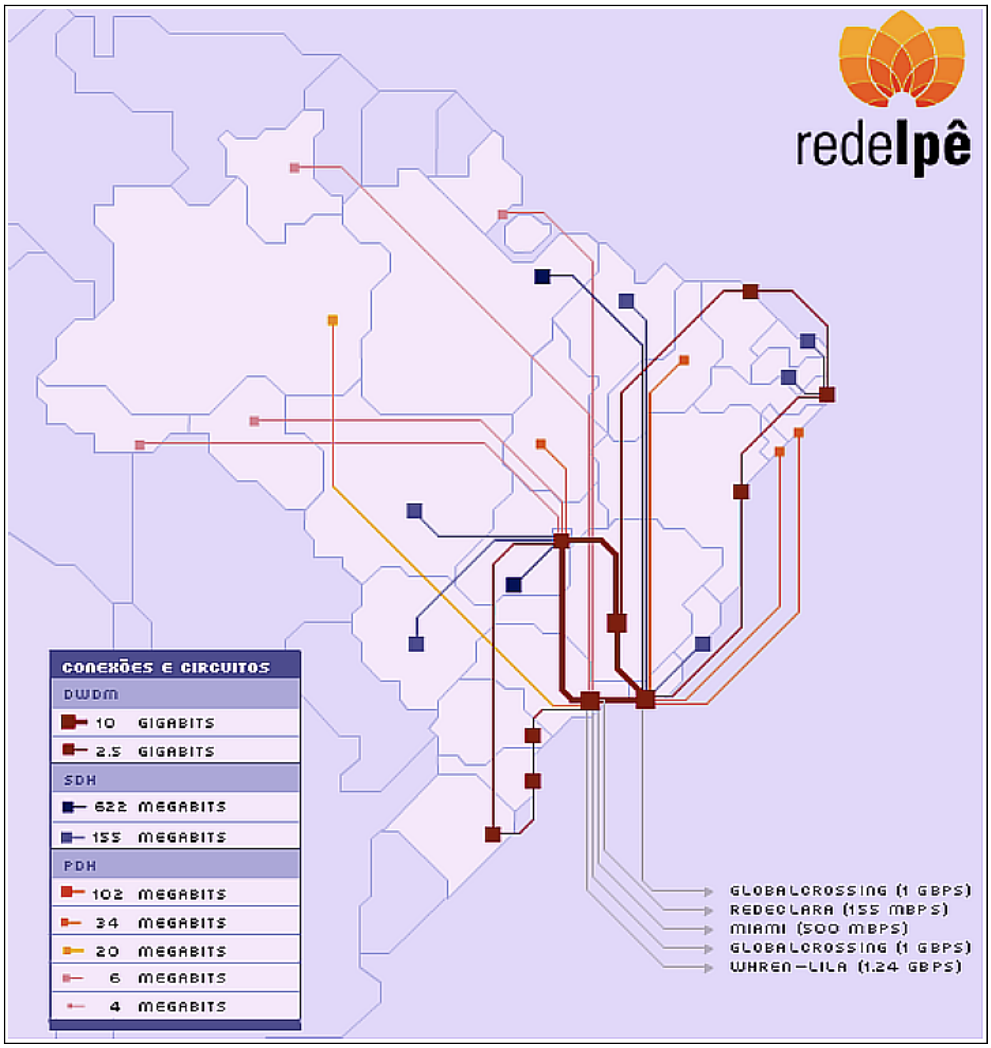

FIG. 4.2: Topologia da Rede Ipê em 2008. Fonte: (RNP, 2011)

A coleta dos dados utilizados foi realizada desde o dia 20/11/2008 até o dia 26/11/2008, a partir das 00:00h, de forma ininterrupta, no roteador localizado no ponto de presença do Estado de Pernambuco, e se refere ao tráfego proveniente da Bahia. Neste período, não houve

indícios de ataques com sobrecarga da rede, representando um período de suposta normalidade. A taxa de amostragem de pacotes *NetFlow* empregada para os registros recebidos é de 1:100, de forma que a taxa média real de 30 Kpps se reduz a 300 pps (LUCENA, 2010). As informações referentes ao tráfego observado foram cedidas pelo Centro de Engenharia e Operações da RNP, mediante solicitação.

A rede da RNP constitui um sistema autônomo com abrangência nacional e, por conta disso, apresenta um grande volume de tráfego.

4.4 Obtenção de Parâmetros Ótimos

Na primeira parte deste trabalho, procurou-se pelos parâmetros dos estimadores EWMA e Holt-Winters que fornecessem os menores somatórios de erros de estimativa num período de tráfego livre de ataques. Para os *traces* do DARPA, serviram de base aqueles referentes à semana 1 e 3 do ano de 1999, que foram criados para fins de treinamento. No caso dos *traces* da RNP e MAWILab, para fins de estudo, considerou-se que os mesmos também estariam livres de ataques.

A partir desses *traces*, foram montadas séries temporais das entropias dos endereços de origem e destino para intervalos fixos de 5 minutos, para DARPA e RNP e de 1 segundo, para MAWILab. Essas séries serviram de base para a montagem das estimativas, que variam de acordo com os parâmetros escolhidos (α, β e γ). Foram testados valores para α, β e γ variando entre 0 e 1. O mesmo foi feito para o parâmetro da EWMA. A combinação que forneceu o menor somatório de erros foi eleita como a mais adequada para os *traces* estudados, como pode ser visto na TAB. 4.2.

Pode-se perceber na TAB. 4.2 que os valores de β e γ são bem menores que os de α. Esses valores evidenciam o peso de cada componente de suavização exponencial na estimativa. Sendo assim, as componentes ligadas à sazonalidade e à tendência de crescimento, explicadas na Seção 3.2.3.2, apresentam menor influência na estimação dos próximos valores.

TAB. 4.2: Parâmetros otimizados para estimadores HW e EWMA

TRACE	**Parâmetros ótimos**			
	Holt-Winters			**EWMA**
	α	β	γ	α
DARPA	0.14	0.01	0.0025	0.368288
RNP	0.775	0.005	0.035	0.6900
MAWILab	0.63	0.01	0.0575	0.6525

4.5 Inserção de Ataques

Para validar os métodos de detecção foi necessária a inserção de ataques gerados artificialmente. Para tanto, alguns pressupostos foram assumidos:

- *Traces* originais livres de ataques ou com percentual irrelevante de fluxos maliciosos;
- Volume máximo de tráfego suportado pelos roteadores desprezado;
- Descartes e atrasos de pacotes, devido ao tráfego inserido, não implementados.

As anomalias causadas pelos ataques são identificadas já no início dos mesmos, quando há uma mudança de comportamento do tráfego, por isso as considerações acima não invalidam os experimentos realizados, já que parte dos efeitos colaterais causados pelo tráfego adicional, e que foram desconsiderados durante a inserção, ocorre após a detecção. Essa abordagem permite uma implementação mais simples, uma vez que basta acrescentar ao tráfego original as informações dos pacotes maliciosos.

Outros fatores foram considerados para determinar a metodologia de inserção:

- Foco na detecção de ataques DDoS caracterizados por grande volume de tráfego;
- Anomalias presentes na distribuição de probabilidade de endereços de destino.

Optou-se pela inserção de ataques de inundação com taxa fixa de pacotes, destinados a uma mesma vítima, e número fixo de zumbis/atacantes. Foram inseridos 15 ataques com duração fixa de 10 unidades de tempo, e espaçados entre si em 50 unidades. Dessa forma, foi possível avaliar a capacidade de detecção em diferentes instantes do período observado.

Para cada *trace*, foram seguidas as seguintes etapas para a inserção:

1- Cálculo do número médio de pacotes por intervalo de tempo para todo o período;

2- Cálculo do percentual a ser acrescentado;

3- Seleção dos intervalos para inserção;

4- Seleção de número de zumbis/atacantes;

Espera-se que a ocorrência de anomalias causadas por ataques DDoS varie de acordo com o volume inserido. Com o intuito de verificar a sensibilidade de cada métrica diante dessa variação, foram inseridas diferentes volumes de ataque em cada experimento, de acordo com a TAB. 4.3, expressos em número de pacotes adicionados.

TAB. 4.3: Pacotes acrescentados na inserção de ataque

Volume médio inserido	**Número de pacotes por janela**		
	DARPA	**RNP**	**MAWI**
50%	2985	42336	23743
25%	1492	21168	11872
20%	1194	16935	9497
15%	895	12701	7123
10%	597	8467	4748
5%	298	4234	2374

A inserção foi realizada durante a extração dos dados, de acordo com a FIG. 4.3, sem a necessidade de alterar o *trace* original.

Inicialmente, os endereços IP de destino dos pacotes presentes em cada intervalo são extraídos do *trace*, de forma a se obter as distribuições de probabilidade para este parâmetro. Em seguida, uma rotina verifica se deve ou não ocorrer a inserção de pacotes, de acordo com os períodos de ataques escolhidos antes da simulação. Caso seja um período de ataque, altera-se o número de pacotes destinados à vítima, adicionando-se o valor listado na TAB. 4.3. A escolha da vítima que recebe os 15 ataques é realizada na primeira inserção através de sorteio. Após isso, pode-se realizar os cálculos referentes a este parâmetro no intervalo considerado, com ou sem inserção de dados.

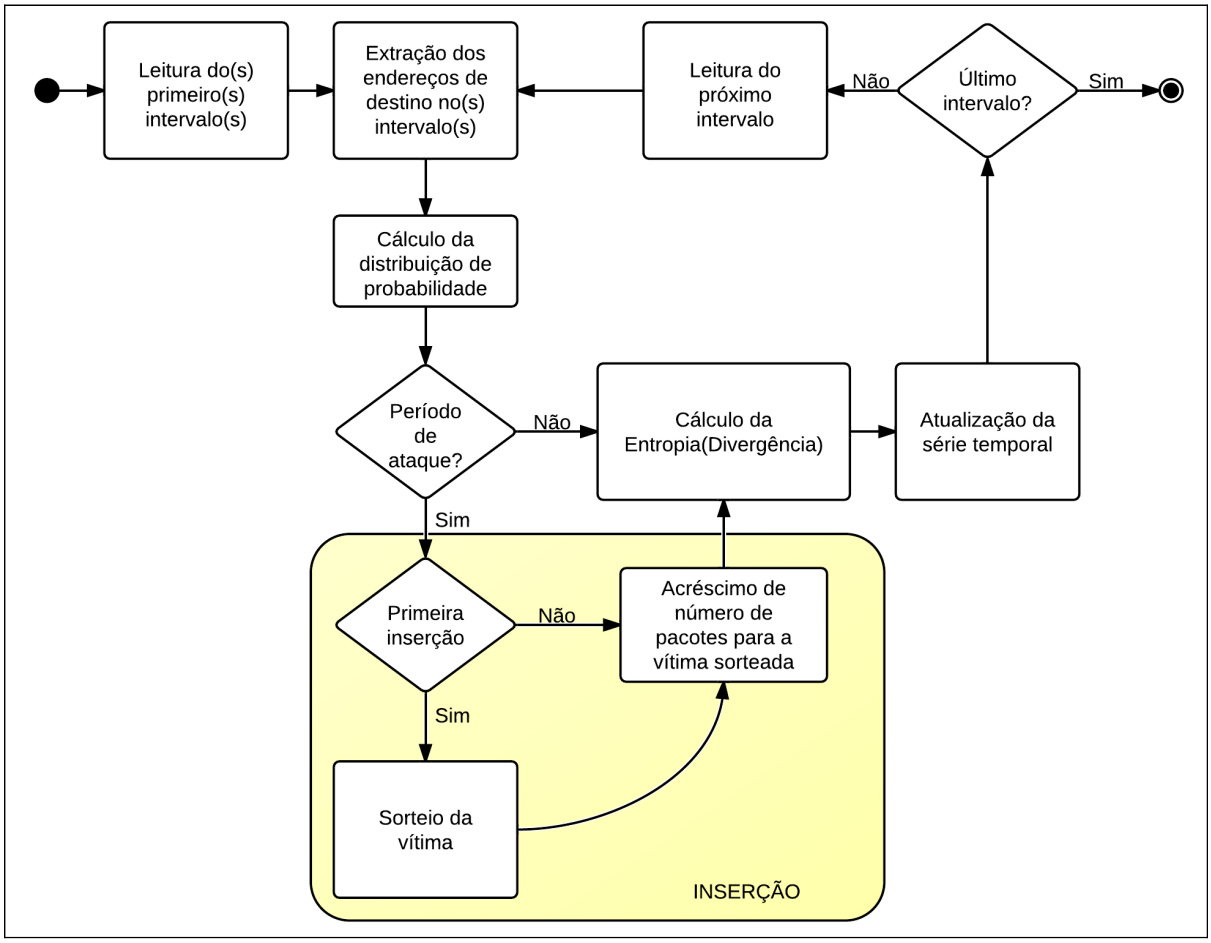

FIG. 4.3: Algoritmo para inserção de ataques

4.6 Experimentos

Para avaliar o desempenho das métricas estudadas, foram realizadas diversas simulações com o intuito de comparar o índice de FPOS e FNEG obtidos com cada configuração de parâmetros. No caso dos estimadores EWMA e HW, duas configurações dos parâmetros α, β e γ foram verificadas: na primeira, aqui chamada de DEFAULT, foram utilizados os valores empregados em (MOURA, 2009), que se baseou em (BRUTLAG, 2000); na segunda, aqui chamada de ÓTIMA, atribuiu-se os valores que garantiam o menor erro de estimativa sem a presença de ataques, de acordo com a TAB. 4.2.

A ocorrência de um ataque DDoS causa uma mudança abrupta na distribuição de endereços IP de destino, de forma que a entropia desta informação sofre uma queda, como pode ser observado na FIG. 4.4. Quando a entropia ultrapassa o limite inferior, emite-se um

alerta, que pode configurar a ocorrência de um FPOS, caso não haja o ataque, ou a detecção do ataque DDoS, caso haja. Na FIG. 4.4, pode-se observar alguns exemplos de resultados obtidos com a utilização dessa arquitetura. Nestes gráficos, foi aplicada nos *traces* da RNP a estimativa de HW com parâmetros PADRÃO (acima) e ÓTIMOS (abaixo). O ataque inserido representa 50% do volume médio calculado em todo o período. A linha azul representa os valores de entropia, calculados a partir dos dados extraídos dos *traces* e ataques inseridos. A linha laranja, por sua vez, representa as estimativas de HW, calculadas a partir dos valores reais, enquanto que as linhas vermelha e ocre indicam os limites inferior e superior, obtidas com a EQ 3.10 e a EQ 3.9, respectivamente.

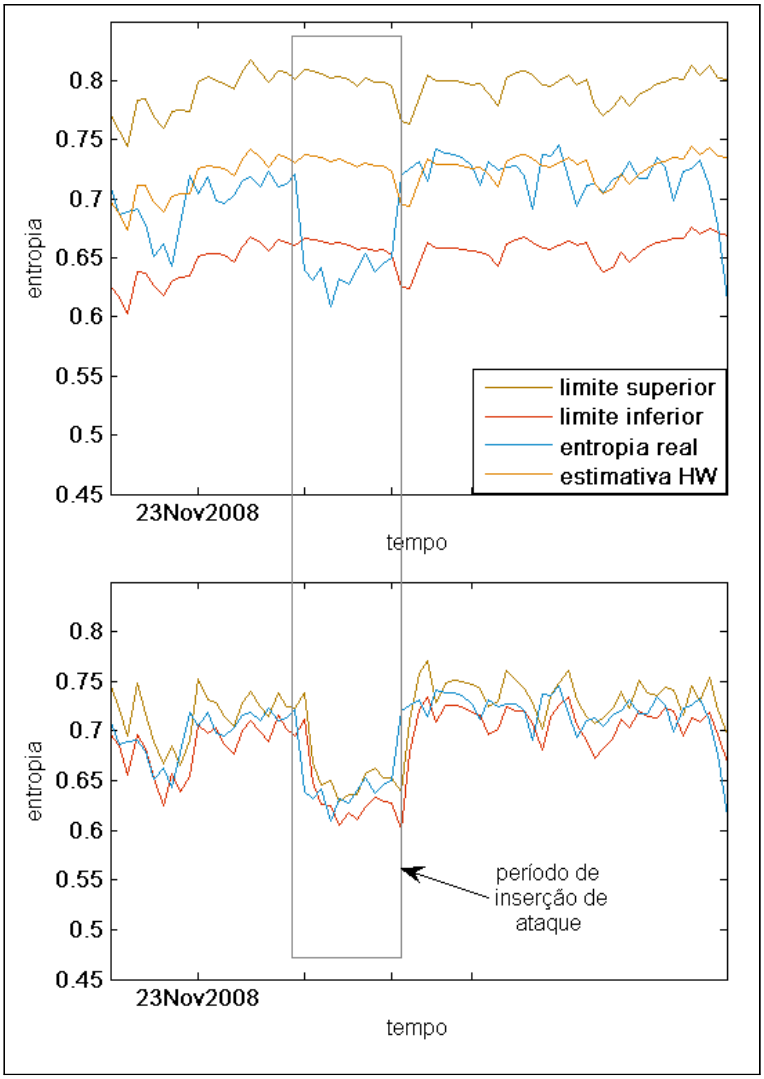

FIG. 4.4: Exemplos de resultados com estimativas de Holt-Winters

Destaca-se nos gráficos da FIG. 4.4 um período de inserção de ataque que foi identificado nos dois gráficos desta figura. Quando o valor real ultrapassa o limite inferior, emite-se um alerta, que pode caracterizar a ocorrência de um FPOS, caso não haja um ataque, ou a detecção do ataque DDoS, caso haja.

Pode-se perceber, também, alguns pontos com violações do limite inferior que ocorreram fora do período de ataque, particularmente no segundo gráfico desta figura, quando foram empregados parâmetros otimizados. Estes eventos caracterizam a ocorrência de FPOS.

Sabe-se que a largura das margens de segurança influencia no índice de FPOS e FNEG, e que esta largura é diretamente proporcional ao valor atribuído a δ. Sendo assim, verificou-se, também, os resultados obtidos com limites mais largos, que aqui chamamos de AMPLOS, e mais ajustados, que aqui chamamos de RIGOROSOS. No primeiro caso, atribuiu-se o valor 3 a δ, enquanto que no segundo, atribuiu-se o valor 2. De acordo com (BRUTLAG, 2000), os valores mais razoáveis para δ estão entre 2 e 3.

Realizou-se, ainda, uma verificação para duas métricas baseadas em divergência: *Chi-Square* e divergência de Hellinger. Nessa verificação, não há a necessidade de ajuste de configuração, já que não há estimação de valores. Para o cálculo do limite de segurança, foi empregada a EQ 3.4. Neste caso, o limite é violado quando a distribuição de endereços muda bruscamente e a divergência alcança valores mais elevados. Em condições normais de tráfego, espera-se que os valores fiquem próximos de 0.

Na FIG. 4.5, pode-se observar um exemplo obtido com a divergência de Hellinger aplicada nos *traces* do MAWILab. O ataque inserido representa 50% do volume médio calculado em todo o período. A linha azul representa os valores de divergência, calculados a partir dos dados extraídos dos *traces* e ataques inseridos. A linha vermelha, por sua vez, representa o limite superior, calculado a partir da desvio-padrão e média móvel da divergência.

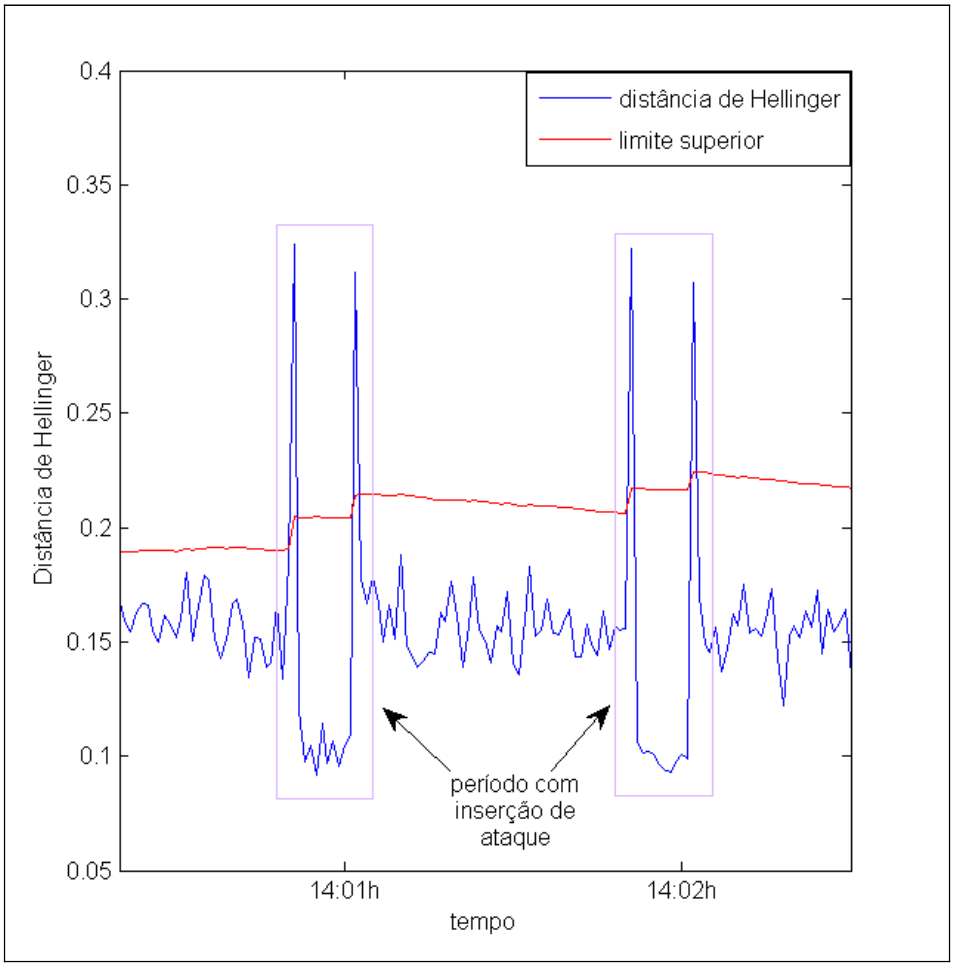

FIG. 4.5: Exemplo de resultado com divergência de Hellinger

Destaca-se no gráfico da FIG. 4.5 dois períodos com inserção de ataque, detectados devido à violação do limite. Quando há essa ultrapassagem, emite-se um alerta, que pode caracterizar a ocorrência de um FPOS, caso não haja um ataque, ou a detecção do ataque DDoS, caso haja. Pode-se perceber que, para cada ataque inserido, existem dois picos que se destacam neste gráfico, indicando o início e o término do ataque. São os dois momentos onde ocorrem mudanças na distribuição dos endereços IP. Pode-se perceber que os valores giram em torno de 0,15, com pequenas variações, e alcança um valor próximo de 0,35 apenas nos inícios e términos de inserção.

Na Figura 4.6, pode-se observar um exemplo obtido com a divergência de *Chi-Square*, aplicada nos *traces* da RNP. O ataque inserido representa 50% do volume médio calculado em todo o período. A linha azul representa os valores de divergência, calculados a partir dos

dados extraídos dos *traces* e ataques inseridos. A linha vermelha, por sua vez, representa o limite superior, calculado a partir da desvio-padrão e média móvel da divergência.

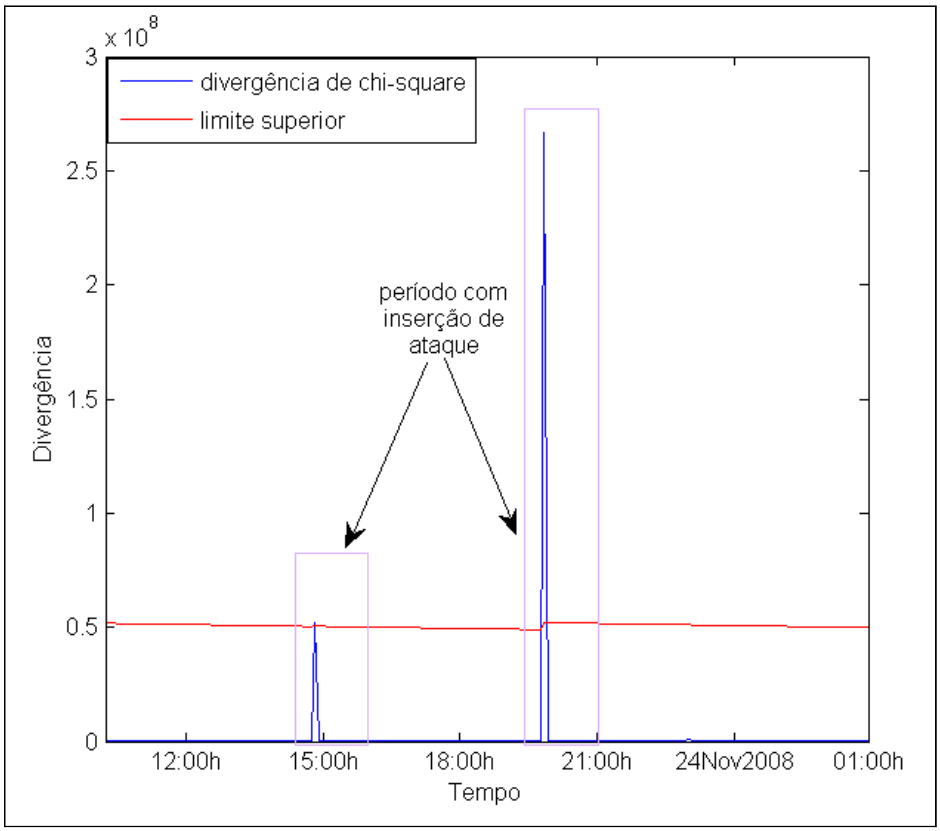

FIG. 4.6: Exemplo de resultado com *Chi-Square*

Destaca-se no gráfico da FIG. 4.6, dois períodos com inserção de ataque, detectados devido à violação do limite. Quando há essa ultrapassagem, emite-se um alerta, que pode caracterizar a ocorrência de um FPOS, caso não haja um ataque, ou a detecção do ataque DDoS, caso haja. Pode-se perceber que, para cada ataque inserido, existe apenas um pico que se destaca no gráfico, devido à assimetria desta métrica. Nesse caso, apenas o início do ataque provoca a anomalia, devido às mudanças na distribuição dos endereços IP que acaba por provocar uma divisão por 0 no algoritmo.

Vale salientar que os valores esperados para a divergência são próximos de 0, mas podem assumir valores bem elevados no início de um ataque, como se pode ver no segundo pico destacado, quando a divergência calculada chega a mais de $2{,}5 \times 10^8$. Pode-se perceber,

também, que a amplitude do primeiro pico em destaque é bem menor que o segundo, apesar do mesmo volume inserido nos dois momentos. No primeiro instante, havia cerca de 110.000 pacotes antes da inserção, enquanto que no segundo havia pouco mais de 41.000 pacotes. Dessa forma, fica bastante evidente que a proporção entre o número de pacotes existentes e o inserido influencia nos resultados observados.

4.6.1 Resultados

Com os valores obtidos, apresentados nos Apêndices 1, 2, 3 e 4, foram montados gráficos para cada *trace* estudado, com o intuito de melhor visualizar os resultados. Dessa forma, foi possível determinar quais os melhores limites, as melhores métricas e configurações em cada caso.

4.6.1.1 Resultados obtidos com estimadores

Para cada *trace* estudado, foram montados quatro gráficos contendo os índices de FNEG e FPOS obtidos para limites RIGOROSOS e AMPLOS.

4.6.1.1.1 Resultados – MAWILab

Observando-se a FIG. 4.7, pode-se verificar que, com a inserção de 15%, 10% e 5% do volume médio transitado, fica mais evidente a capacidade de detecção de cada métrica. Como os ataques são menos expressivos, as mudanças de comportamento são mais brandas e, por isso, mais difíceis de detectar.

Na inserção de 5%, tanto com limites rigorosos quanto com limites amplos, observou-se o menor índice de FNEG com o emprego do estimador EWMA e parâmetros otimizados, cujos índices de erro alcançaram 20% e 45%, respectivamente. O segundo melhor resultado foi obtido com o estimador HW e parâmetros ótimos. Entretanto, no cenário no qual se emprega limites amplos, observou-se um crescimento exponencial do número de FNEG na inserção de 5%, quando mais de 90% de ataques não são detectados.

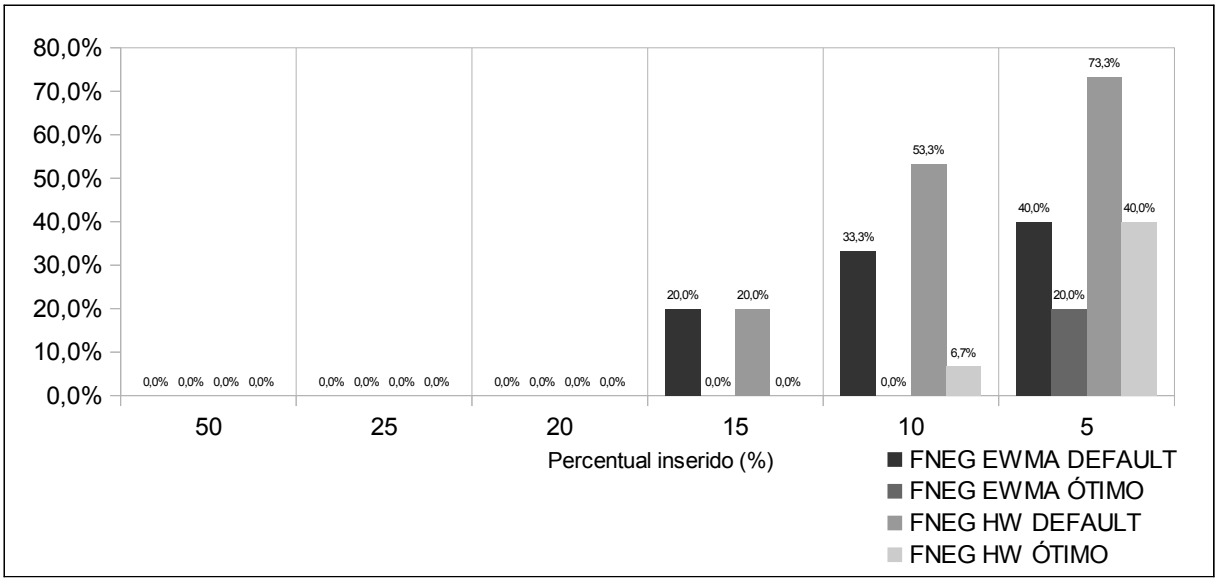

FIG. 4.7: Índice de FNEG para MAWILab com limites rigorosos

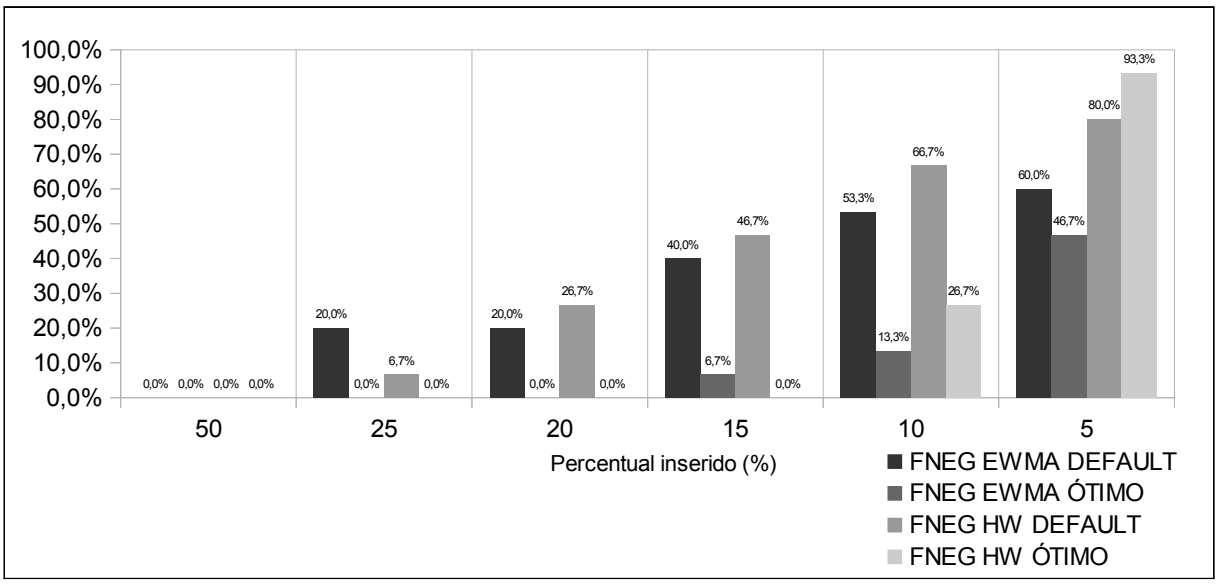

FIG. 4.8: Índice de FNEG para MAWILab com limites amplos

No que diz respeito ao índice de FPOS, observou-se valores relativamente baixos nos experimentos. A variação desse índice, devido à mudança do volume de ataque inserido e ao tipo de limite adotado, também foi relativamente pequena, como pode ser visto na FIG. 4.9 e FIG. 4.10. Observa-se que o valor máximo obtido foi de 5,11%, quando se utilizou o EWMA com parâmetros otimizados e limites rigorosos. As menores taxas de FPOS foram obtidas com o emprego do HW.

Dessa forma, particularmente para ataques menos volumosos, o estimador que apresentou melhor desempenho na detecção de ataques neste enlace foi o EWMA com parâmetros otimizados. Quanto aos limites, fica evidente a vantagem de se empregar o rigoroso, dado o pequeno aumento dos índices de FPOS, e a grande diminuição nos índices de FNEG.

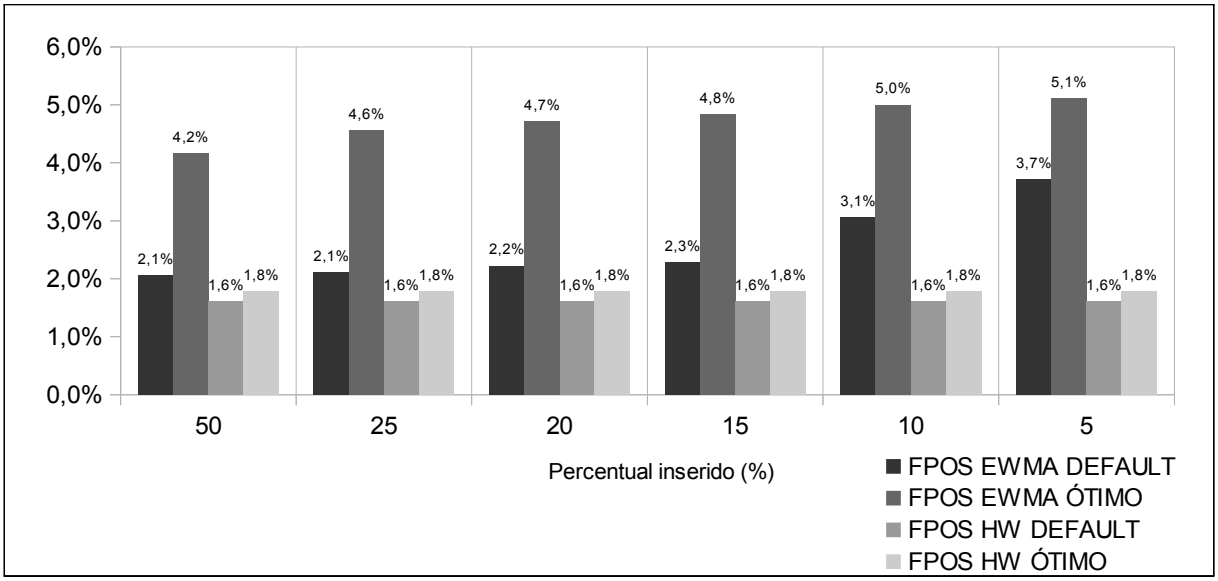

FIG. 4.9: Índice de FPOS para MAWILab com limites rigorosos

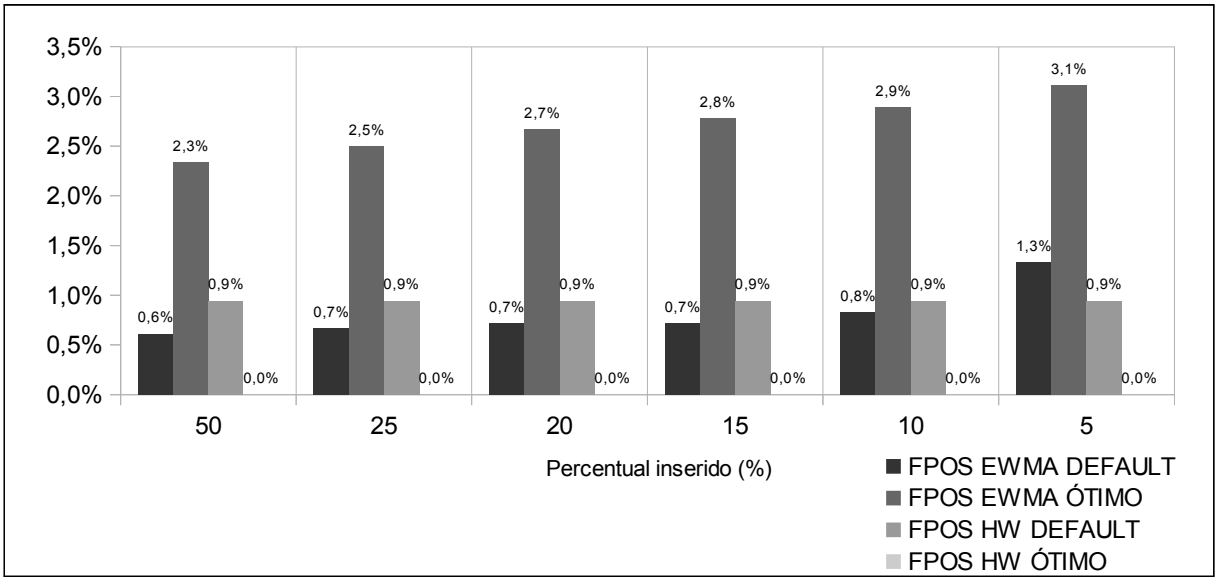

FIG. 4.10: Índice de FPOS para MAWILab com limites amplos

4.6.1.1.2 Resultados – RNP

Pode-se perceber pela FIG. 4.11, que, nos *traces* da RNP, os menores índices de FNEG foram obtidos com o emprego de parâmetros otimizados. Pode-se observar, também, que nos experimentos com limites amplos, o número de FNEG cresce mais rapidamente com o EWMA, de forma que, a 10% e 5% de inserção, o HW com parâmetros otimizados apresentou menores índices de erro.

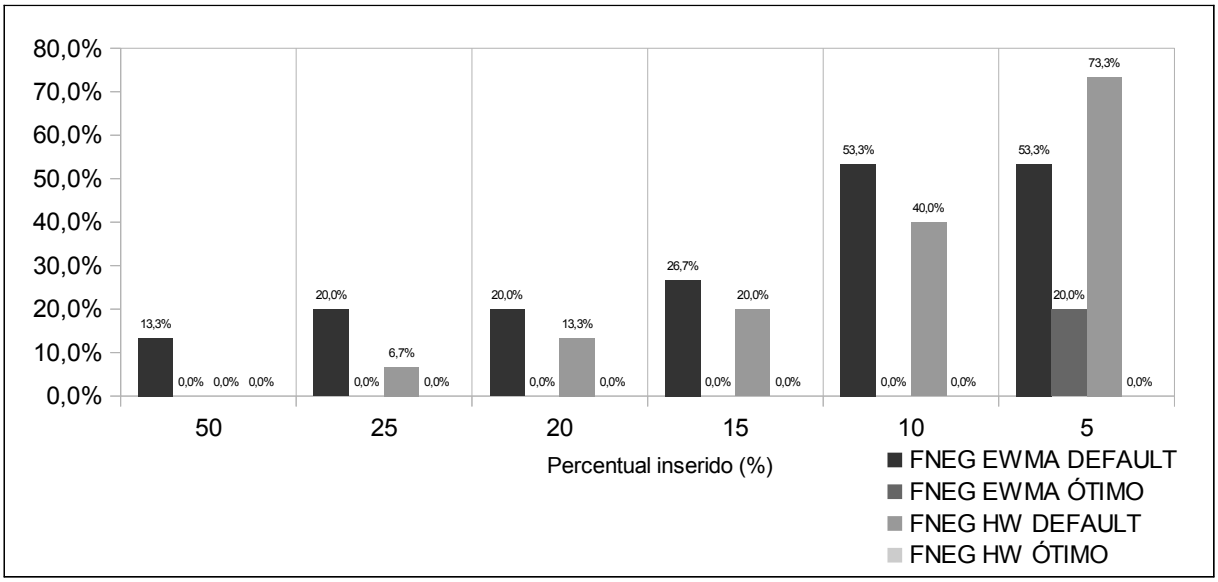

FIG 4.11: Índice de FNEG para RNP com limites rigorosos

Além disso, verificou-se que, mesmo a 5% de inserção, o índice de FNEG do HW com parâmetros otimizados foi de 0%. Dessa forma, pode-se dizer que, para esse enlace, o estimador que apresentou o melhor desempenho foi o HW com parâmetros otimizados, tanto para limites amplos como para rigorosos.

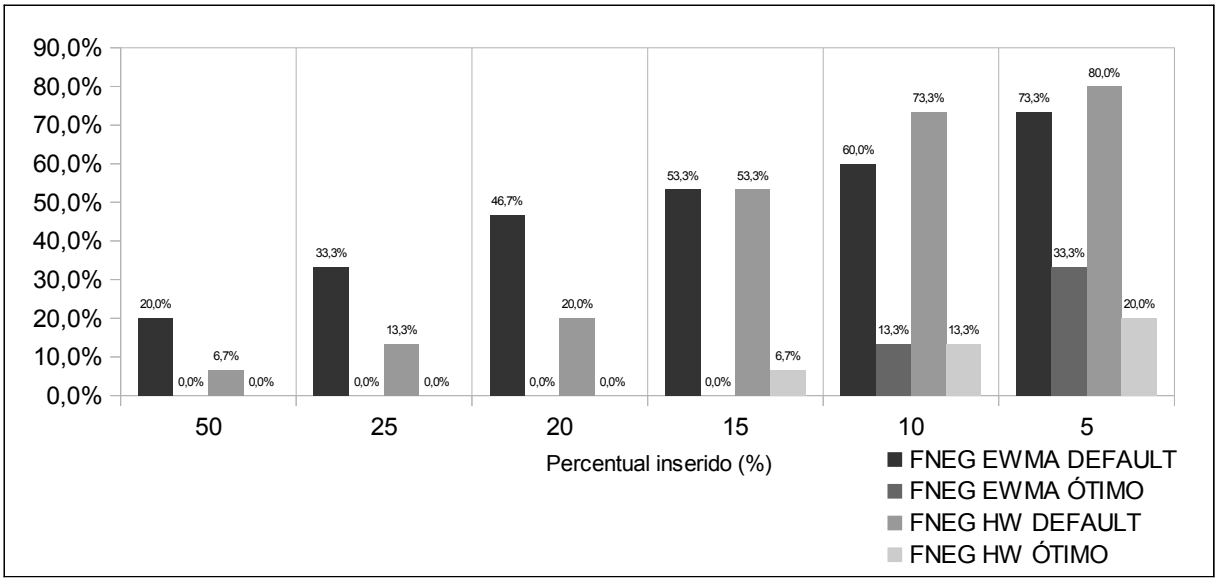

FIG 4.12: Índice de FNEG para RNP com limites amplos

Assim como ocorre com o MAWILab, o índice de FPOS muda muito pouco com a variação do volume de ataque inserido. Entretanto, a utilização de limites mais rigorosos quase que duplica o número de FPOS, que, ainda assim, não alcança percentuais muito expressivos. No pior caso, observado com o HW otimizado e limites rigorosos, o índice de FPOS alcançando foi de 13,64%. Com limites mais amplos esse valor cai para 7,32%. Pode-se observar na FIG. 4.13 e na FIG. 4.14, que os índices de FPOS obtidos com o HW equivalem a pouco mais que o dobro daqueles alcançados com o EWMA, e que os mesmos não ultrapassam 5,43%, com limites rigorosos, e 3,48%, com limites amplos.

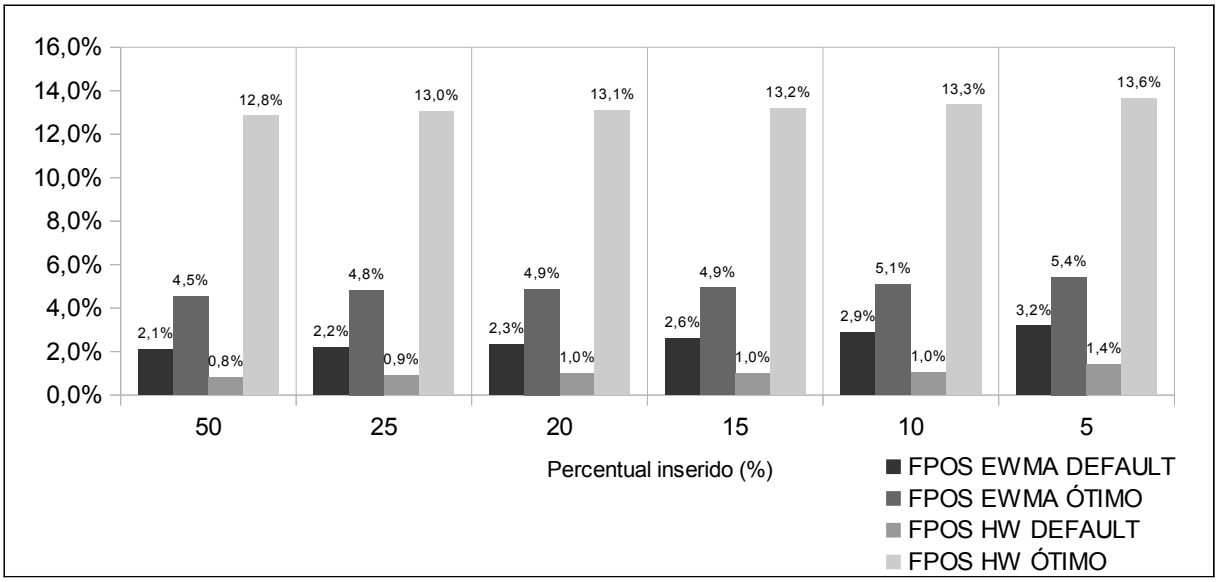

FIG. 4.13: Índice de FPOS para RNP com limites rigorosos

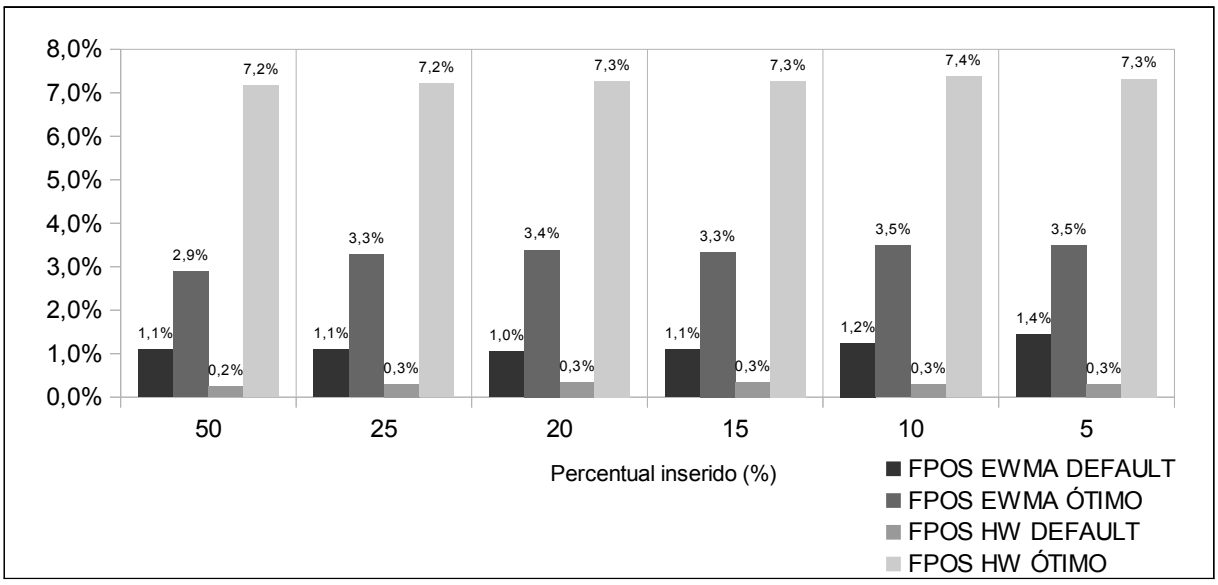

FIG. 4.14: Índice de FPOS para RNP com limites amplos

4.6.1.1.3 Resultados – DARPA

Pode-se observar na FIG. 4.15, que as taxas de FNEG obtidas com os *traces* do DARPA foram relativamente altas, mesmo com inserções mais volumosas. Numa das configurações testadas, por exemplo, o índice de FNEG chegou a 100%. Fica evidente, também, que os menores índices de FNEG foram obtidos com o emprego de parâmetros otimizados. Além disso, o HW mostrou-se mais adequado ao *trace* do que o EWMA.

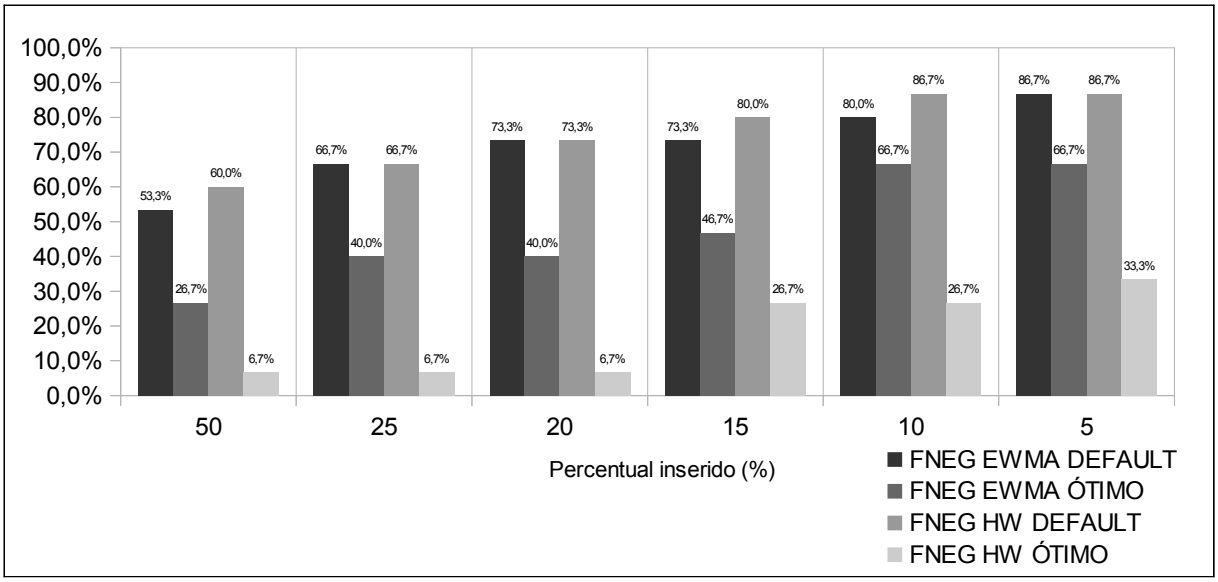

FIG. 4.15: Índice de FNEG para DARPA com limites rigorosos

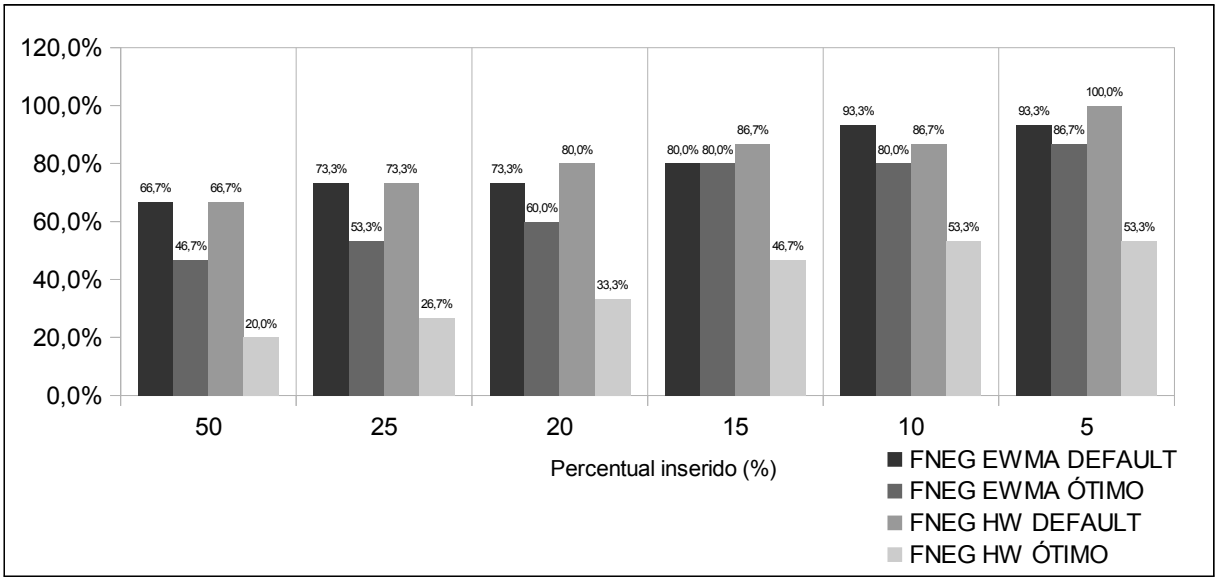

FIG. 4.16: Índice de FNEG para DARPA com limites amplos

Os índices de FPOS, por sua vez, permaneceram em patamares relativamente baixos, como pode ser visto na FIG. 4.17 e FIG. 4.18. O maior índice observado foi de 5,17%, quando foi empregado limites rigorosos com o HW.

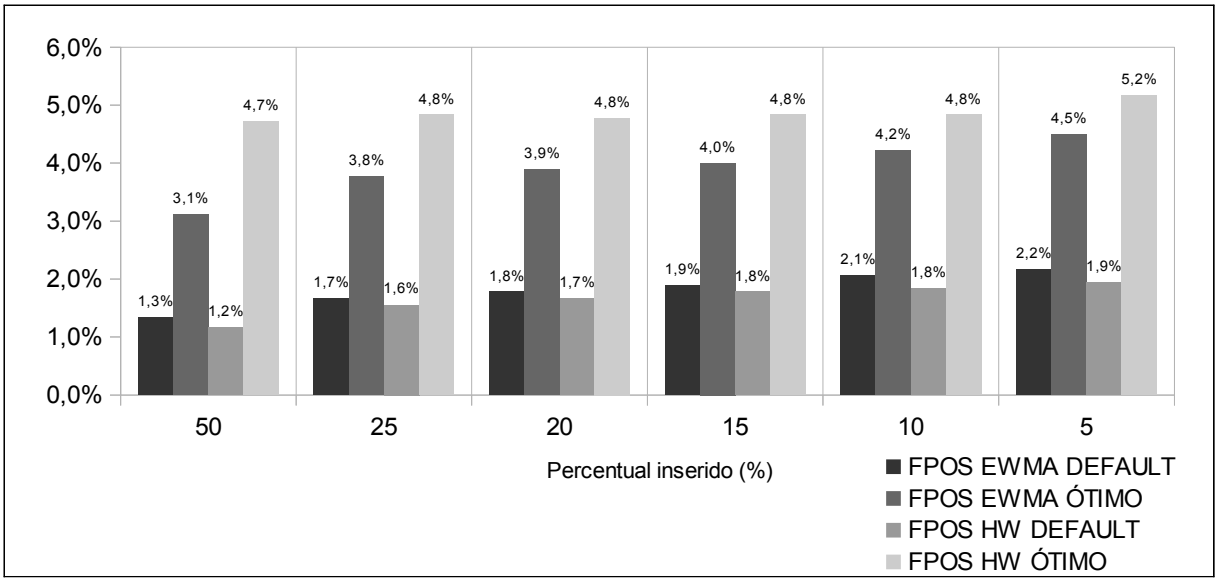

FIG. 4.17: Índice de FPOS para DARPA com limites rigorosos

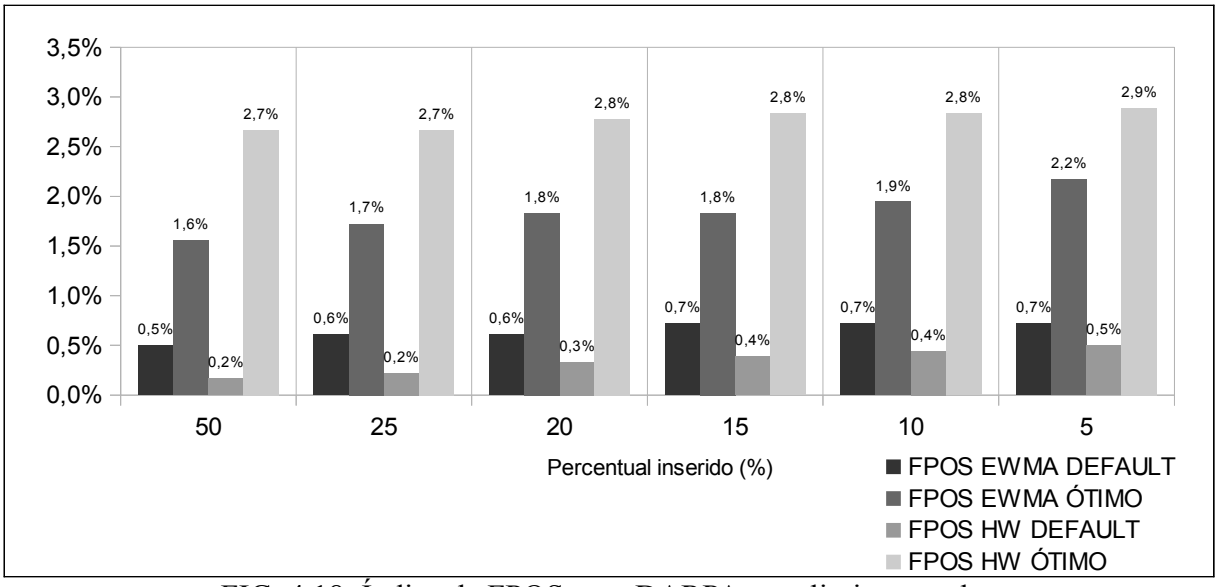

FIG. 4.18: Índice de FPOS para DARPA com limites amplos

Pode-se perceber que, em relação aos experimentos realizados com os outros *traces*, os índices de FNEG obtidos com os *traces* do DARPA são bem altos. Entretanto, nas inserções de 50, 25 e 20%, o HW com parâmetros otimizados e limites rigorosos apresentou uma taxa de FNEG de 6,67%, referente a 1 ataque não detectado dentre 15 existentes. Sendo assim, para este trace, o HW com parâmetros otimizados apresenta boa capacidade de detecção de ataques mais volumosos. Além disso, sabe-se que o alvo do ataque DDoS detectado estaria diretamente ligado a este enlace, de forma que o volume de inundação esperado seja alto.

Analisando a distribuição de volume de tráfego em todo período dos *traces* do DARPA, constata-se que a taxa média de tráfego no roteador é de 5.971 pacotes a cada 5 minutos, e que e o desvio-padrão nesta distribuição é de 5.738. Sendo assim, a dispersão estatística da distribuição permite que os valores inseridos representem diferentes percentuais do total avaliado, de acordo com o momento do ataque. Por conta disso, a variação da entropia causada pelo tráfego anômalo varia de acordo com o momento da inserção, apesar do volume de ataque ser o mesmo. Pode-se observar na FIG. 4.19 a série temporal contendo o número de pacotes a cada intervalo de 5 minutos num determinado período do *trace* do DARPA.

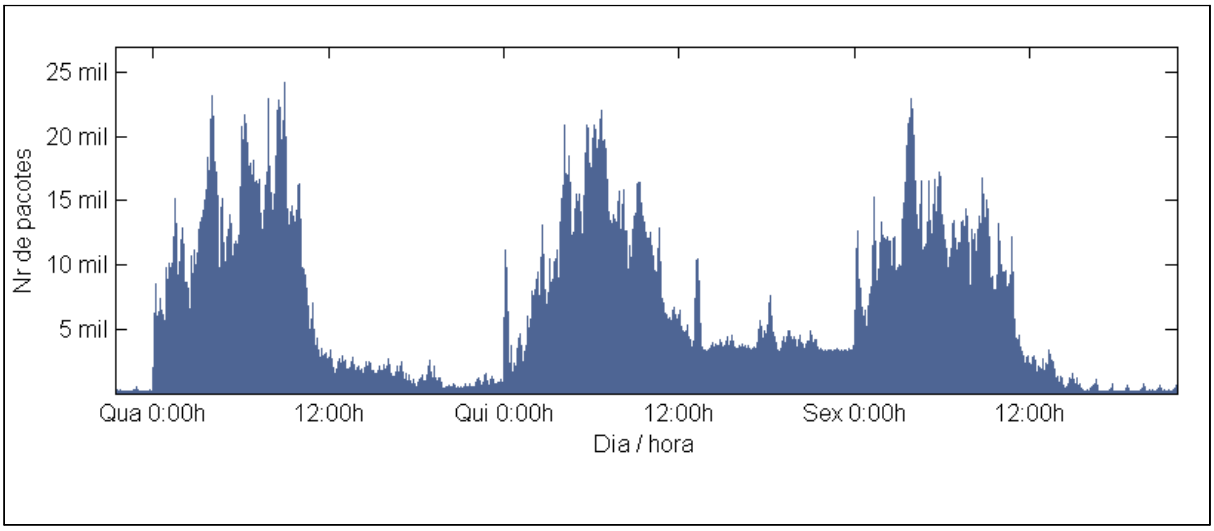

FIG. 4.19: Série temporal com número de pacotes a cada 5 minutos - DARPA

Além disso, sabe-se que esse *trace* foi gerado artificialmente e representa o tráfego de uma rede relativamente pequena e antiga. Como não há encaminhamento de pacotes para outras redes, a quantidade de destinos possíveis torna-se bastante reduzida. Dessa forma, a variação da entropia se torna mais limitada, uma vez que esse valor está relacionado ao número de diferentes ocorrências no espaço amostral, como pode ser visto na seção 3.2.1.

Pode-se observar na FIG. 4.20 e na FIG. 4.21 as séries temporais contendo o número de pacotes por intervalo de tempo em determinados períodos dos *traces* da RNP e do MAWILab, respectivamente. Assim como no *trace* do DARPA, verifica-se uma sazonalidade diária para a RNP. Já para o WAWILab, devido ao reduzido período de observação, essa sazonalidade não pode ser percebida.

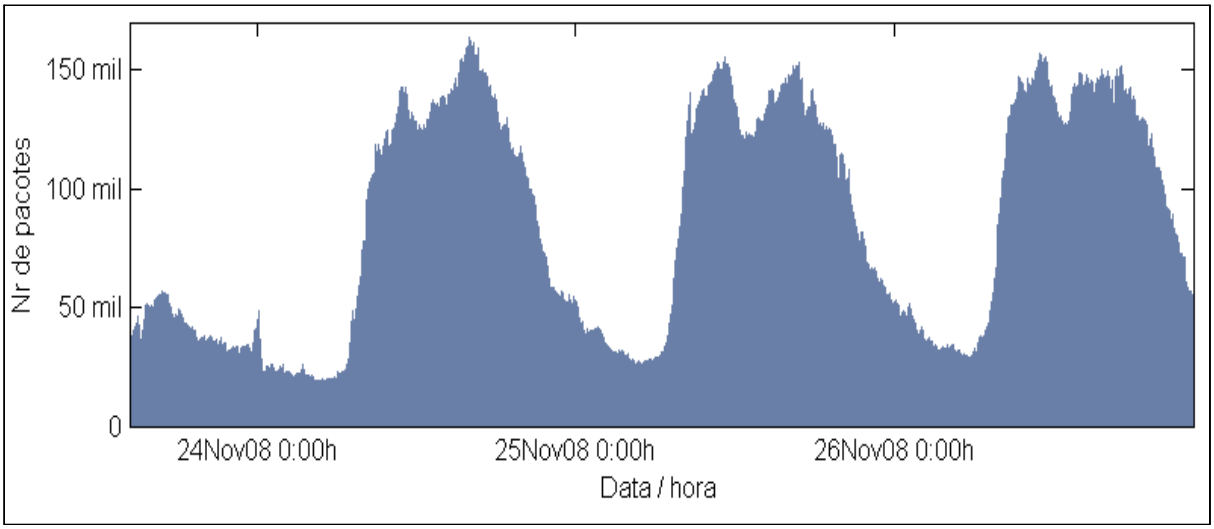

FIG. 4.20: Série temporal com número de pacotes a cada 5 minutos - RNP

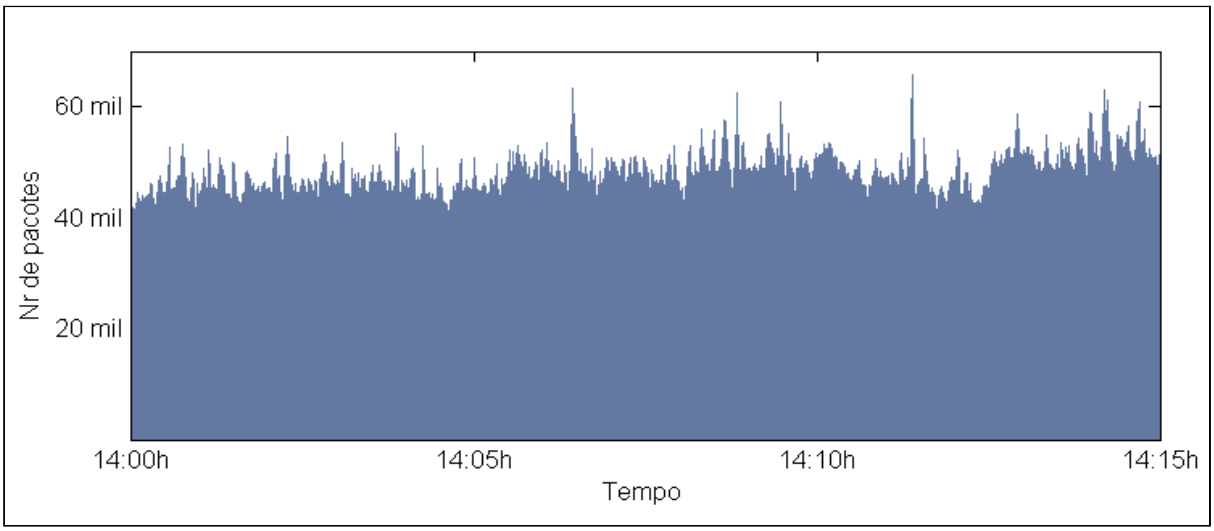

FIG. 4.21: Série temporal com número de pacotes a cada 1 segundo - MAWI

4.6.1.2 Resultados para Divergências de Hellinger e *Chi-Square*

Para os *traces* do MAWILab, a Divergência de Hellinger apresentou índices competitivos de FNEG apenas na detecção dos ataques mais volumosos, com inserções de 50%, 25% e 20%. A partir deste ponto do gráfico, os índices de FNEG começam a crescer linearmente até apresentar 80% de erro. Já com o *Chi-Square*, verificou-se uma taxa de FNEG de 6,67% para todos os cenários testados, constituindo o melhor resultado alcançado no corrente trabalho (FIG. 4.22).

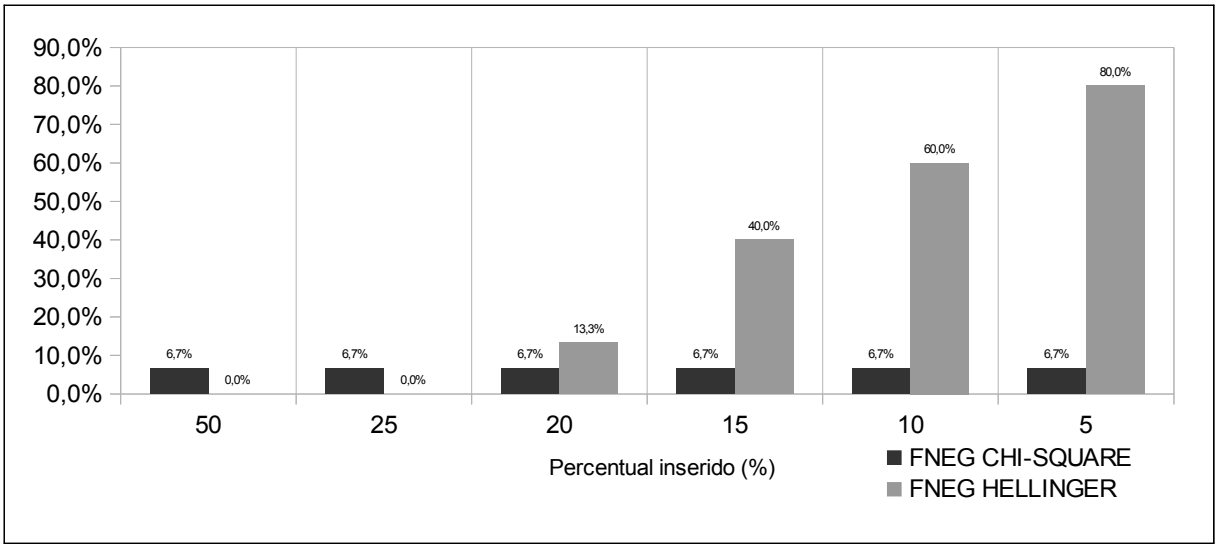

FIG. 4.22: Índice de FNEG para MAWILab com Divergência de Hellinger e *Chi-Square*

Para os *traces* da RNP, os resultados obtidos foram diferentes, como pode ser verificado na FIG. 4.23. A divergência de Hellinger apresentou os melhores resultados em todos os casos simulados, igualando-se ao *Chi-Square* apenas com 5% de inserção. Vale salientar que, embora os índices de FNEG tenham sido mais altos, os valores observados com o *Chi-Square* para inserções de 20%, 15%,10% e 5% permaneceram em 33,33%, que equivale a 5 erros de detecção, demonstrando uma certa constância para diferentes volumes de ataque.

Para os *traces* do DARPA, os índices de FNEG foram relativamente altos, quando comparados àqueles obtidos com os outros dois *traces*, mesmo com ataques mais volumosos. O melhor resultado foi obtido com a divergência de Hellinger, que apresentou uma taxa de 53,33% de FNEG com a inserção de 50% (FIG. 4.24).

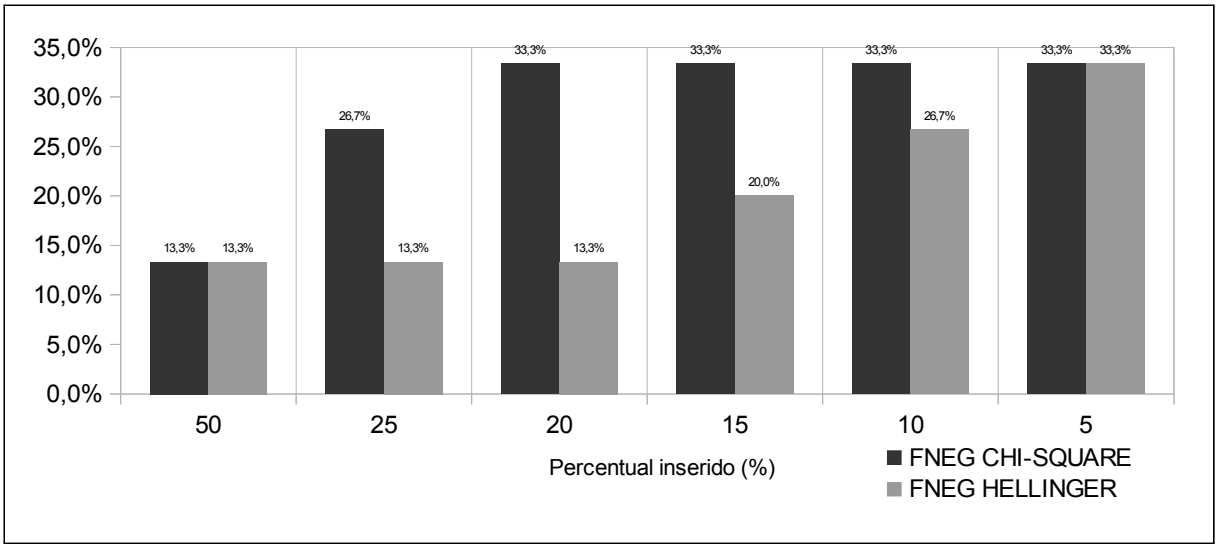

FIG. 4.23: Índice de FNEG para RNP com Divergência de Hellinger e *Chi-Square*

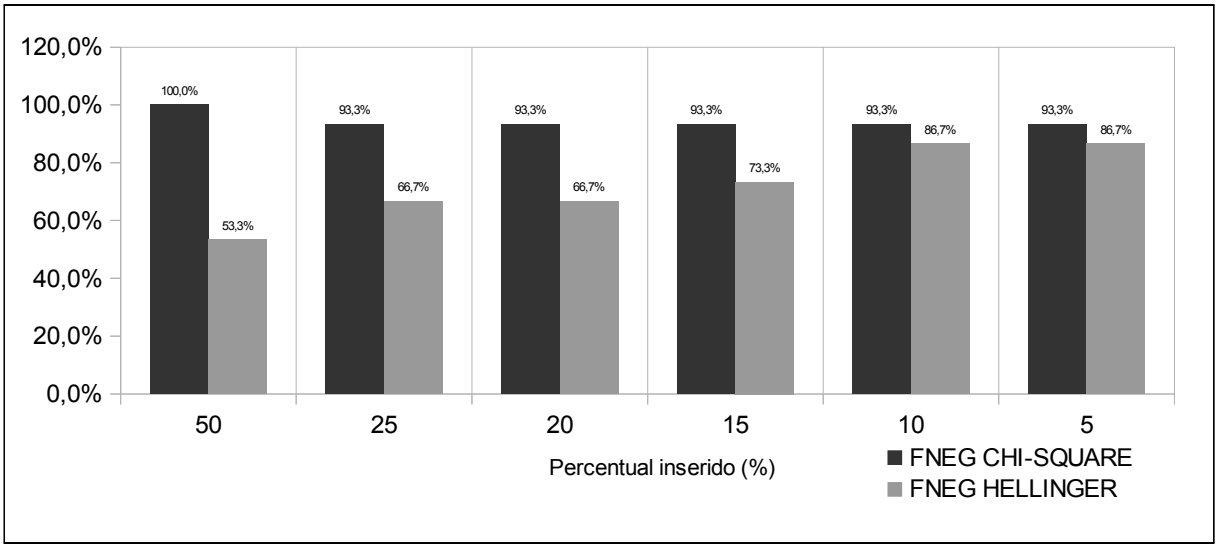

FIG. 4.24: Índice de FNEG para DARPA com Divergência de Hellinger e *Chi-Square*

Verifica-se que os valores dos índices de FNEG são bem diferentes em cada caso. As características de cada *trace*, tais como taxa de amostragem, intervalos de tempo considerados e outros inerentes ao tráfego, podem ser as principais causas das distinções observadas. Entretanto, em todos os casos estudados, os índices de FPOS são bem reduzidos e apresentam resultados bem semelhantes em todos os *traces*, como pode ser observado nas FIG. 4.25, 4.26 e 4.27. Todas as taxas foram abaixo de 3% e, apenas para os *traces* do DARPA, a divergência de *Chi-Square* resultou em maiores índices de FPOS do que a divergência de Hellinger.

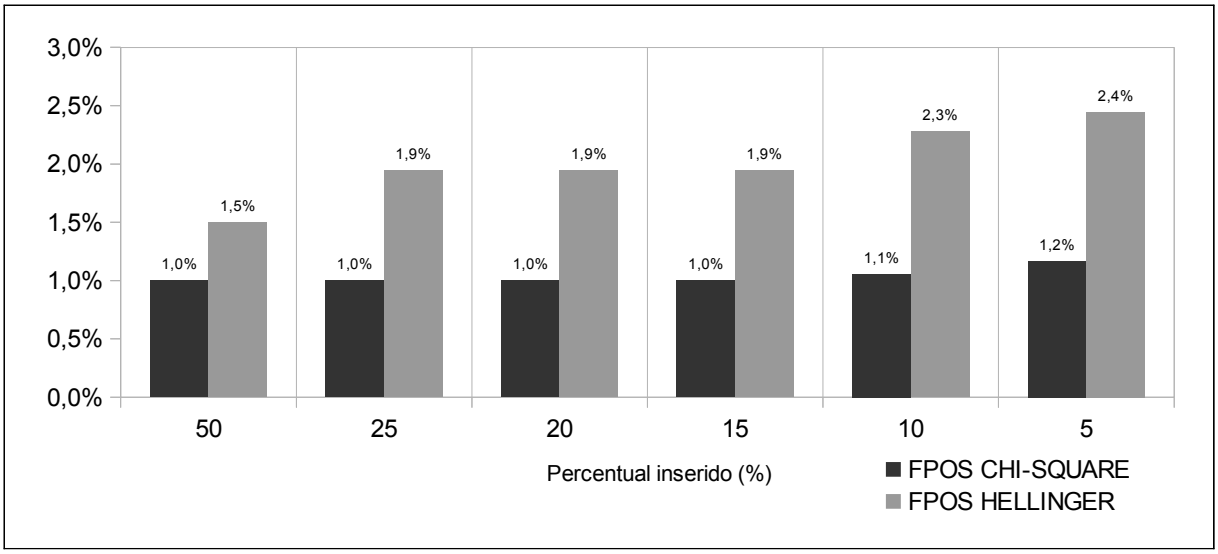

FIG. 4.25: Índice de FPOS para MAWILab com Divergência de Hellinger e *Chi-Square*

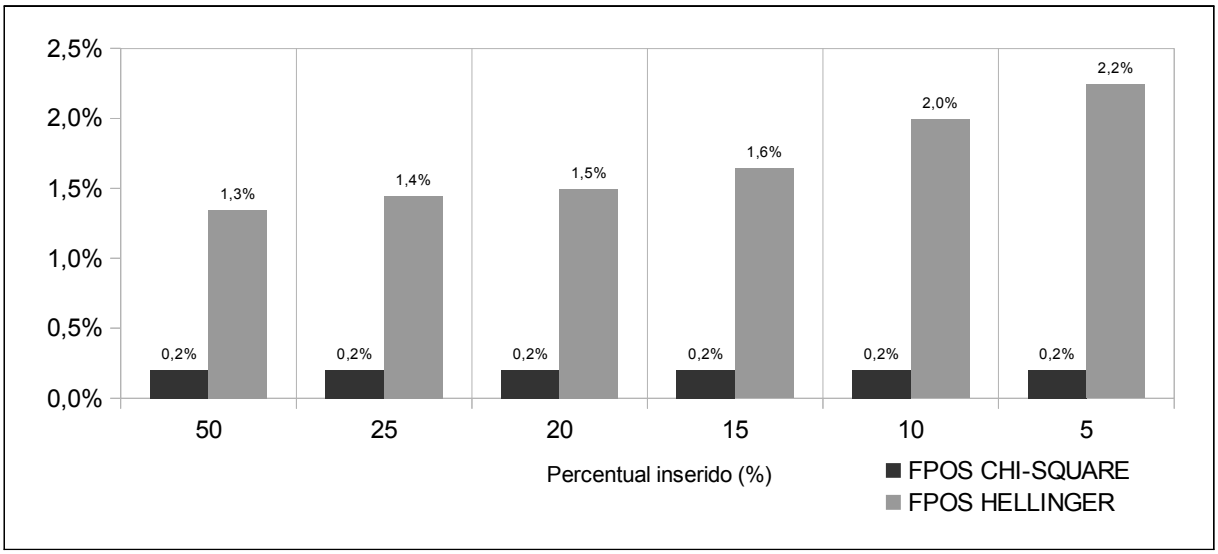

FIG 4.26: Índice de FPOS para RNP com Divergência de Hellinger e *Chi-Square*

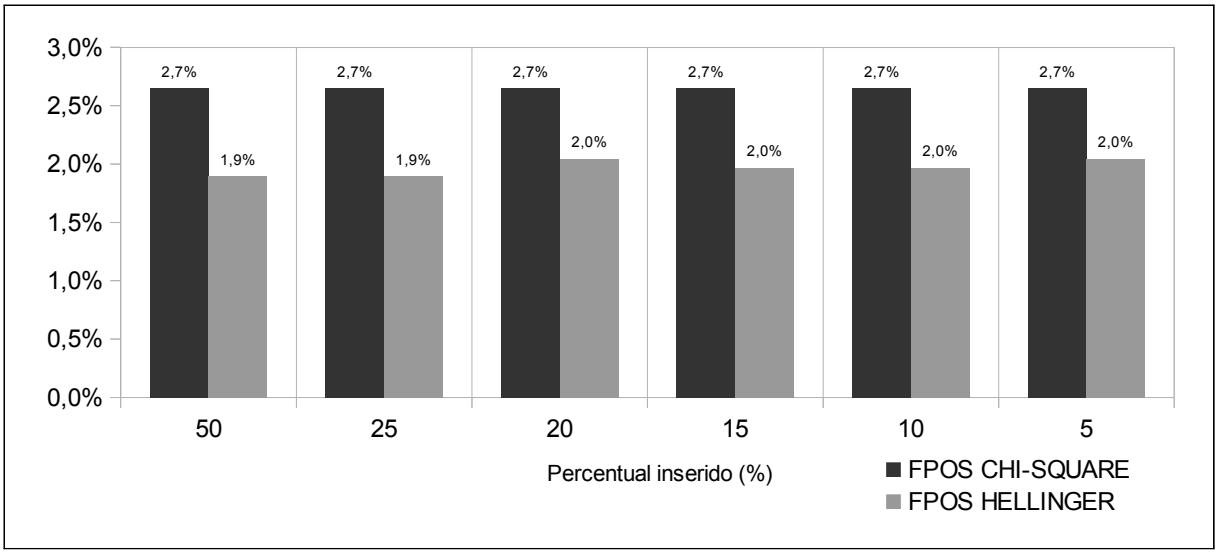

FIG 4.27: Índice de FPOS para DARPA com Divergência de Hellinger e *Chi-Square*

4.6.2 Análise de Resultados

De maneira geral, pode-se observar em todos os gráficos que o desempenho dos estimadores diminui a medida em que o volume do ataque é reduzido (de 50% até 5%). Pode-se verificar o mesmo resultado nos gráficos relacionados à divergência. Embora não haja alterações significativas nos índices de FPOS, os índices de FNEG aumentam a medida em que o volume inserido diminui.

No que diz respeito ao estudo de configuração dos estimadores, os resultados indicam que a utilização de parâmetros ÓTIMOS com limites rigorosos aumenta expressivamente a sensibilidade dos dois estimadores estudados, trazendo melhores taxas de detecção em todos os experimentos realizados, apesar do pequeno aumento de FPOS. Para os *traces* da RNP e do DARPA, os menores índices de FNEG foram obtidos com o estimador HW, devido ao comportamento sazonal presente neles, como visto na FIG. 4.19 e FIG. 4.20. Já para o MAWILab, as menores taxas foram alcançadas com o EWMA.

No que diz respeito ao desempenho das métricas estudadas, percebe-se que, para os os *traces* da RNP e do DARPA, os resultados obtidos com o HW são melhores do que aqueles obtidos através da divergência Hellinger e do *Chi-Square*. Já para o *trace* do MAWILab, as menores taxas de FNEG para inserções de 5% e 10% foram obtidas com o *Chi-Square*, que

demonstrou maior sensibilidade para menores tráfegos maliciosos. Por outro lado, para volumes maiores de inserção, as taxas de FNEG obtidas no MAWILab com o *Chi-Square* foram maiores do que as obtidas com a divergência de Hellinger, com o HW e com o EWMA. Ainda assim, os valores são muito próximos, já que representa a identificação de apenas 1 ataque a mais.

Dessa forma, verifica-se que há uma métrica mais adequada para cada *trace* estudado, como pode ser observado na TAB. 4.4, que traz os melhores resultados obtidos em cada métrica com os ataques mais volumosos (50 % do volume) e menos volumosos (5 % do volume). Para o EWMA e o HW, observa-se nesta tabela as taxas de FNEG obtidas com limites rigorosos e parâmetros otimizados.

TAB. 4.4: Comparativo entre índices de FNEG das métricas nos *traces* estudados

MÉTRICA	**Ataque com 50% do volume**			**Ataque com 5% do volume**		
	DARPA	**RNP**	**MAWILab**	**DARPA**	**RNP**	**MAWILab**
EWMA	26,7 %	0 %	0 %	66,7 %	20,0 %	20,0 %
Holt-Winters	6,7 %	0 %	0 %	33,3 %	0 %	40,0 %
Divergência de Hellinger	53,3 %	13,0 %	0 %	86,7 %	33,33 %	80,0 %
Chi-Square	100 %	13,0 %	6,67 %	93,3 %	33,33 %	6,67 %

De maneira geral, pode-se perceber que, em relação aos experimentos realizados com os outros *traces*, os índices de FNEG obtidos com os *traces* do DARPA são bem altos. Entretanto, nos ataques mais volumosos observa-se a ocorrência de 1 erro de detecção (6,67%) com o HW. Sendo assim, pode-se utilizar essa métrica num enlace deste porte, que estaria diretamente ligado ao alvo do ataque de inundação. Como não há encaminhamento de pacotes para outras redes, a variedade de endereços de destinos possíveis torna-se reduzida. Dessa forma, a variação da entropia se torna mais limitada, uma vez que esse valor está relacionado ao número de diferentes ocorrências no espaço amostral, como pode ser visto na seção 3.2.1.

Os resultados obtidos com o *trace* MAWILab divergem um pouco daqueles obtidos com os outros *traces*, particularmente com os ataques menos volumosos (TAB. 4.4). Embora o

tráfego seja volumoso, o período disponibilizado foi de apenas 15 minutos a contar das 14:00h, sendo necessário avaliar intervalos de 1 segundo, ao invés de 5 minutos. Tais características impossibilitaram a verificação das mudanças ocorridas em outros horários e, consequentemente, a determinação do comportamento do tráfego neste enlace.

4.7 Construção de um Sistema Colaborativo de Detecção

No presente trabalho, propõe-se a avaliação do desempenho de algumas métricas de detecção de ataques DDoS, considerando-se diferentes volumes de ataque. Nesta seção, será apresentada um sistema colaborativo que pode aproveitar as peculiaridades identificadas no corrente trabalho, bem como os aspectos que influenciam na escolha das melhores métricas.

Neste sistema, o compartilhamento de dados entre diferentes Sistemas Autônomos e a correlação de alertas pode reduzir os índices de falso positivos e permite uma detecção mais eficaz. Além disso, com o emprego de limites mais sensíveis, a detecção pode ocorrer antecipadamente, em pontos mais afastados da vítima.

No sistema aqui sugerido, a detecção do ataque DDoS ocorre em três etapas. Na primeira, os roteadores de borda enviam para um ou mais servidores pacotes do tipo *NetFlow* referentes aos fluxos que passam pelo Sistema Autônomo. A partir desses dados, podem ser montadas séries temporais que representam o histórico do tráfego avaliado. Nesta etapa, a cada atualização desta série, verifica-se a ocorrência de anomalias nas interfaces, de acordo com as métricas empregadas. Sob condições normais, espera-se que os valores calculados a partir desse extrato respeitem as margens de segurança, que também são atualizadas dinamicamente. Os pacotes *NetFlow*, gerados periodicamente, fornecem dados que são suficientes para a criação e atualização dessas séries temporais e limites.

Para se garantir o mínimo possível de ataques não detectados, critérios suficientemente rigorosos devem ser implementados, uma vez que a taxa de pacotes maliciosos pode ser relativamente baixa em Sistemas Autônomos que estejam mais afastados da vítima. Entretanto, a incidência excessiva de falsos positivos tornaria a técnica inviável. A seleção de métricas e parâmetros mais adequados ao tráfego monitorado pode atender a essas exigências,

tornando os limites mais confiáveis. Pode-se embasar essas escolhas a partir de experimentos que caracterizem o tráfego, como os realizados no corrente trabalho.

Verificar a presença de anomalias nos roteadores de borda dos Sistemas Autônomos facilita, de certa forma, o monitoramento dos tráfegos, pois esses pontos são concentradores de fluxos. Entretanto, avaliações individualizadas das violações de limites ocorridos nos roteadores não garantem bons resultados, devido ao alto índice de falsos positivos ocasionados pelo emprego de limites mais rigorosos.

Na segunda etapa, as suspeitas identificadas na etapa anterior são avaliadas como um todo. Nesse momento, busca-se por afunilamentos de fluxo malicioso, caracterizados pela alteração no padrão de tráfego destacada na FIG. 4.28 (c), quando duas ou mais entradas e uma saída apresentam alertas. Essa configuração constitui um indício da passagem de ataques DDoS pelo SA avaliado. Vale salientar que essa identificação é realizada antes do fluxo chegar no SA da vítima.

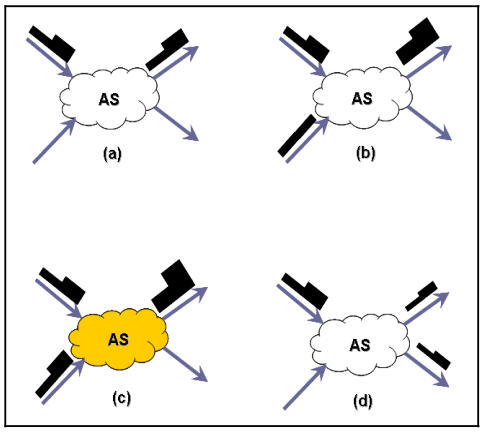

FIG. 4.28: Padrões de Tráfego em roteadores de borda de Sistemas Autônomos

Os endereços mais frequentes num mesmo intervalo podem ser facilmente extraídos a partir da leitura de pacotes *NetFlow*. Através dessa análise, pode-se identificar o endereço da vítima, já que o número de pacotes direcionados a esta máquina tende a aumentar. Espera-se que essa identificação seja mais evidente na saída do SA, devido ao acúmulo do tráfego malicioso.

Uma vez identificada a vítima, um alerta pode ser montado e compartilhado com com outros SA. O padrão IDMEF trata da transmissão e do armazenamento de alertas e é bastante flexível (DEBAR, 2007), podendo ser adaptado para que atenda às necessidades desse sistema.

Na terceira etapa, o SA envia esse alerta para um ou mais SA participantes do sistema colaborativo. Pode-se utilizar o protocolo BGP para viabilizar essa comunicação, uma vez que este protocolo permite a troca de mensagens entre SA, mesmo quando eles não são vizinhos (CASTELÚCIO, 2009). Desta maneira, forma-se uma uma rede sobreposta, como pode ser observado na FIG. 4.29.

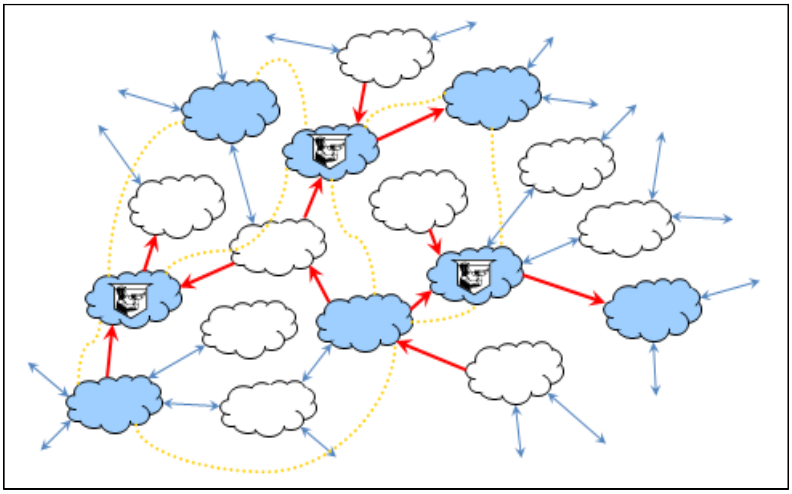

FIG. 4.29: SA participantes do Sistema Colaborativo de Detecção

Após o recebimento dos alertas, os mesmos podem ser armazenados numa tabela, para fins de correlação. Espera-se que sejam gerados vários alertas, mas a confirmação da presença do ataque só se dará quando, num dado intervalo de tempo, um mesmo endereço estiver presente em alertas de vários SA. Desta forma, os falso positivos gerados na primeira e na segunda etapa podem ser descartados, aumentando a eficácia do sistema. À medida em que os ataques se aproximam da vítima, o volume de tráfego malicioso aumenta e, consequentemente, torna-se mais evidente a confirmação e identificação do ataque, como pode ser visto na FIG. 4.30. Tais considerações são válidas para ataques DDoS que se originam em vários SA.

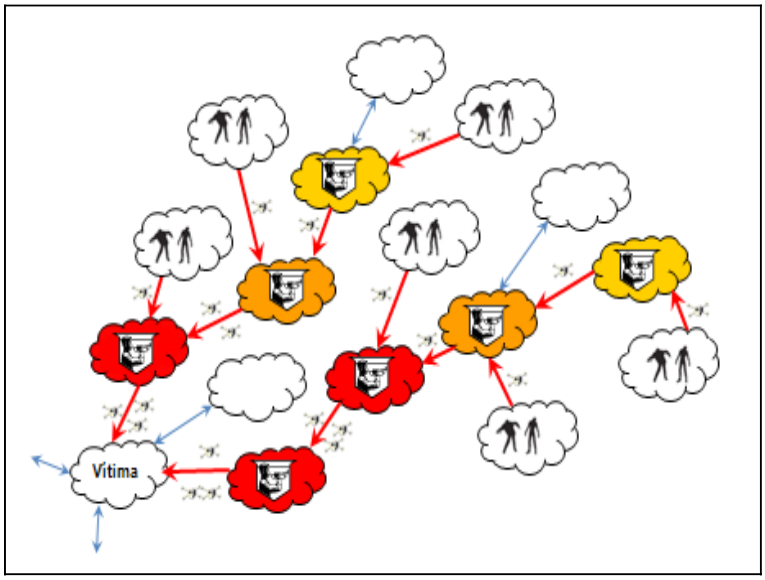

FIG. 4.30: Compartilhamento de alertas confirmando a existência de ataque

Esta sugestão apenas representa uma maneira de realizar a detecção de um ataque DDoS de forma antecipada seguindo um esquema colaborativo, e que ainda precisa ser melhor estudada para verificar sua viabilidade.

De acordo com os resultados obtidos no corrente trabalho, desde que se conheça as características do tráfego monitorado e dos enlaces observados, pode-se selecionar as métricas mais eficientes para cada cenário.

5 CONSIDERAÇÕES FINAIS

Neste trabalho, foram avaliados dois métodos de detecção de anomalias de tráfego de rede, baseados na extração de entropias de endereços IP e no uso de estimadores (EWMA e HW). Foram verificados dois tipos de configuração (PADRÃO e ÓTIMA), com o intuito de encontrar qual a combinação mais adequada para a detecção de ataques em pontos afastados da vítima, onde o volume de tráfego malicioso ainda é baixo, considerando-se um cenário colaborativo. Verificou-se, também, o desempenho dessas métricas com o uso de limites AMPLOS e RIGOROSOS. Verificou-se, ainda, o desempenho de duas métricas baseadas em divergência (divergência de Hellinger e *Chi-Square*) para verificar se poderiam ser mais adequados a esse tipo de cenário.

Foi utilizada uma metodologia própria, baseada na inserção artificial de pacotes em *traces* de origens bem distintas. A performance das métricas estudadas foi avaliada em termos dos índices de FPOS e FNEG obtidos em cada cenário.

Diante do exposto, conclui-se que a utilização de parâmetros otimizados e limites RIGOROSOS aumenta expressivamente a sensibilidade das duas métricas estudadas, trazendo melhores taxas de detecção em ambos os *traces*. Foi observado um aumento nos índices de FPOS, devido ao maior ajuste dos limites de segurança, embora essa alteração tenha sido relativamente pequena. Por outro lado, os índices de FNEG diminuíram significativamente com parâmetros ÓTIMOS. Por essa razão, a utilização dos estimadores com parâmetros ÓTIMOS em ambientes colaborativos, ou mesmo naqueles onde se deseja identificar volumes menos expressivos de ataque, apresentam uma relação custo-benefício mais vantajosa.

Vale salientar que o estimador HW trouxe melhores índices de FNEG apenas para os *traces* do RNP e DARPA, devido ao comportamento sazonal identificado nestes, como visto na FIG. 4.19 e FIG. 4.20. A divergência de *Chi-Square* trouxe melhor desempenho apenas para o *trace* do MAWILab, demonstrando maior sensibilidade para tráfegos maliciosos de menor volume.

5.1 Trabalhos Futuros

A arquitetura de avaliação de desempenho de métricas de detecção de ataques DDoS aqui proposta permite ao pesquisador uma certa liberdade para realizar diversos tipos de experimentos. Para trabalhos futuros, pode ser considerado o estudo da sensibilidade de outras métricas não abordadas no corrente trabalho, bem como o estudo de ataques mais sofisticados, cujo comportamento, relacionado ao volume ou endereçamento, dificulta a sua identificação.

A detecção de um ataque de inundação não se resume apenas na escolha da métrica mais adequada, de forma que se faz necessário o estudo de outras etapas de detecção. Outra sugestão para trabalhos futuros seria o estudo de um sistema de detecção, como aquele citado na Seção 4, que realiza uma abordagem colaborativa e hierárquica e que, devido à correlação de alertas, pode utilizar métricas mais sensíveis, como as estudadas no presente trabalho, permitindo uma identificação antecipada do ataque em pontos mais afastados da vítima.

Os dados gerados a partir das detecções realizadas pode servir para alimentar uma base de dados que pode ser compartilhada entre vários integrantes desse sistema colaborativo. Estudar a forma de utilização dessas informações constitui uma outra maneira de dar continuidade a esta pesquisa.

Verificou-se nos experimentos que os parâmetros otimizados para o HW e EWMA foram diferentes em cada *trace* estudado, e, de certa forma, podem representar o tráfego, constituindo uma assinatura do mesmo. Tais informações podem ser úteis para a construção de um gerador de tráfego, por exemplo.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AHNLAB, INC. **Analytical report on 3.4 DDoS attack.** White paper, April 2011. Disponível: http://www.ahnlab.com [capturado em 10 de julho de 2011]

BASSEVILLE, M. e NIKIFOROV, I. **Detection of Abrupt Changes: Theory and Applications**. Em Englewood Cliffs, NJ: Prentice Hall, Inc., 1993.

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

CABRERA, J. B. D., LEWIS, L., QIN, X., LEE, W., PRASANTH, R. K., RAVICHANDRAN, B., e MEHRA, R. K. **Proactive Detection of Distributed Denial of Service Attacks using MIB Traffic Variables - A Feasibility Study,** 7th IFIP/IEEE International Symposium on Integrated Network Management, Seattle, WA-May 14-18, 2001.

CASTELÚCIO, A. O., SALLES, R. M. e ZIVIANI, A. **An AS-Level Overlay Network for IP Traceback** - IEEE Network, Vol. 23, pp. 36-41, 2009.

CERT.BR. **Cartilha de segurança para Internet**. Disponível: http://cartilha.cert.br [capturado em 10 de novembro de 2011].

CHEN, Y. e HWANG, K. **Collaborative Change Detection of DDoS Attacks on Community and ISP Networks** - IEEE Networks, pp. 401 - 410, 2006.

CHEN, Y., HWANG, K., e KU, W. S. **Collaborative Detection of DDoS Attacks over Multiple Network Domains** - IEEE Transactions on Parallel and Distributed Systems, Vol. 18, Issue 12, pp. 1649 - 1662, 2007.

CISCO SYSTEMS, INC. **NetFlow Services Solution Guide**, Disponível: http://www.cisco.com/en/US/docs/ios/solutions_docs/netflow/nfwhite.pdf. 2008 [capturado em 22 de novembro de 2010.]

CISCO SYSTEMS, INC. **NetFlow Version 9 Flow-Record Format.** white paper Disponível: http://www.cisco.com/en/US/technologies/tk648/tk362/technologies_white_paper09186a00800a3db9_ps6601_Products_White_Paper.html, May 2011 [capturado em 2 de junho de 2011].

DEBAR, H., CURRY, D. e FEINSTEIN, B. **The Intrusion Detection Message Exchange Format (IDMEF).** RFC 4765, March 2007, URL: http://www.ietf.org/rfc/rfc4765.txt.

DEMIR, O. e KHAN, B. **Quantifying Distributed System Stability through Simulation: A Case Study of an Agent-Based System for Flow Reconstruction of DDoS Attacks** - IEEE, 2010 ISMS, Liverpool, England, January 2010.

DITTRICH, D. e DIETRICH, S**. P2P as botnet command and control: a deeper insight** - IEEE Network, 3rd International Conference on Malicious and Unwanted Software pp. 41-48, 2008.

ESTEVEZ-TAPIADOR, J. M., GARCIA-TEODORO, P. e DIAZ-VERDEJO, J. E. **Anomaly Detection Methods in Wired Networks: a Survey and Taxonomy**, Computer Communications, Vol. 27, 1569-1584, 2004.

FEINSTEIN, L., SCHNACKENBERG, D., BALUPARI, R. e KINDRED, D. **Statistical approaches to DDoS attack detection and response.** Em: DARPA Information Survivability Conference and Exposition, 2003. Proceedings, Vol. 1 , pp. 303 – 314, April 2003.

FEITOSA, E. L., SOUTO, E. J. P. e SADOK, D. **Tráfego Internet não Desejado: Conceitos, Caracterização e Soluções.** Livro-texto dos Minicursos do VIII Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais, pp 91-137, 2008.

FENG, J. e LIU, Y. **The Research of DDoS Attack Detecting Algorithm Based on the Feature of the Traffic** - Networking and Mobile Computing 5th International Conference on Wireless Communications (WiCom'09), 2009.

FERRER, Z. e FERRER, M. C. **In-depth Analysis of Hydraq,** URL: http://www.ca.com/files/securityadvisornews/in-depth_analysis_of_hydraq_final_231538.pdf [capturado em novembro de 2010].

HAINES, J. W., LIPPMANN, R. P., FRIED, D. J., ZISSMAN, M. A., TRAN, E., e BOSWELL, S. B. **1999 DARPA Intrusion Detection Evaluation: Design and Procedures.** Lincoln Laboratory. 26 de fevereiro de 2001.

HAWKINSON, J. **Guidelines for creation, selection, and registration of an Autonomous System (AS).** RFC 1930, March 1996, URL: http://tools.ietf.org/html/rfc1930.

ISC, **internet host count history**. Disponível: https://www.isc.org/solutions/survey/history [capturado em 7 de novembro de 2011].

JORNALNH. **Anonymous convoca ataque ao Facebook no dia 28**. Janeiro 2012, Disponível: http://www.jornalnh.com.br/tecnologia/369601 [capturado em 22 de maio de 2012]

KALEKAR, P. S., e REKHI, K. **Time series Forecasting using Holt-Winters Exponential Smoothing.** School of Information Technology, December 6, 2004.

KAUR, G., SAXENA V. e GUPTA, J. P**. Anomaly Detection in Network Traffic and Role of Wavelets** - IEEE, 2nd International Conference on Computer Engineering and Technology (ICCET), 2010.

KLINE, J., NAM, S., BARFORD, P., PLONKA, D. e RON, A. **Traffic Anomaly Detection at Fine Time Scales with Bayes Nets** - The Third International Conference on Internet Monitoring and Protection, 2008. ICIMP'08, PP. 37-46, 2008.

LABOVITZ, C. **The Internet Goes to War**. Em ARBOR SERT, Disponível: http://ddos.arbornetworks.com/2010/12/the-internet-goes-to-war/ December 2010 [capturado em 22 de maio de 2012].

LAKHINA, A., CROVELLA, M. e DIOT, C. **Mining Anomalies Using Traffic Feature Distributions** - Proceedings of the **ACM SIGCOMM'05**, Philadelphia, Pennsylvania, USA, 2005.

LAUFER, R. P., VELLOSO, P. B., CUNHA, D. O., MORAES, I. M., BICUDO, M. D. D., e DUARTE, O. C. M. B. **A new IP traceback system against denial-of-service attacks**. Em 12th International Conference on Telecommunications - ICT'2005, Capetown, South Africa, May 2005.

LAW, K.T., LUI, J. C. S. e YAU, D. K. Y. **An Effective Methodology to Traceback DDoS Attackers** - X IEEE Int'l Symp, MASCOTS'02, 2002.

LI, T. **Robust Divergence Measures for Time Series Discrimination**. Tech. Rept. 237, Department of Statistics, Texas A&M Univ., College Station, 1995.

LIN, B. P. e UDDIN, M. S. **Synmon Architecture for Source-based SYN-flooding Defense on Network Processor** - IEEE, 2005 Asia-Pacific Conference on Communications, Perth, Western Australia, 2005.

LINCOLN LABORATORY. **DARPA Intrusion Detection valuation.** Disponível: http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html 1999, [capturado em 22 de novembro de 2010].

LUCENA, S. C. e MOURA, A. S. **Detecção de Anomalias Baseada em Análise de Entropia no Tráfego da RNP** - XIII WGRS, pp. 163-176, 2008.

LUCENA, S. C. e MOURA, A. S. **Estimativa de Holt-Winters para Detecção de Ataques em Redes WAN.** Em X Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais, SBSeg'10, Fortaleza-CE, 2010.

MAWILAB DATABASE. 2011. Disponível: http://www.fukuda-lab.org/mawilab/ [capturado em 22 de novembro de 2011].

MIRKOVIC, J. e REIHER P. **D-WARD: A Source-End Defense against Flooding DoS Attacks** - IEEE Trans. Dependable and Secure Computing, pp. 216-232, 2005.

MOURA, A. S. **Detecção de Anomalias em Redes WAN usando Estimativa de Holt-Winters Aplicada a Medidas de Entropia**, Dissertação de Mestrado – UFRJ-2009.

PAPADOPOULOS, C., LINDELL, R., MEHRINGER, J., HUSSAIN, A. e GOVINDAN, R. **COSSACK: Coordinated Suppression of Simultaneous Attacks** - Proc.Third DARPA Information Survivability Conf. and Exposition (DISCEX-III'03), pp.2-13, 2003.

PARK, K. e LEE, H. **A Proactive Approach to Distributed DoS Attack Prevention using Route-Based Packet Filtering**, Technical Report CSD-TR-00-017, Purdue University, Dept. of Computer Sciences, 2000.

PASCHALIDIS, I. C. e SMARAGDAKIS, G. **Spatio-Temporal Network Anomaly Detection by Assessing Deviations of Empirical Measures** - IEEE/ACM Transactions on Networking, vol.17, no.3, 2009.

PHAAL, P., PANCHEN, S. e MCKEE, N. **InMon Corporation's sFlow: A Method for Monitoring Traffic in Switched and Routed Networks.** RFC 3176 (Informational), setembro 2001. URL: http://www.ietf.org/rfc/rfc3176.txt.

QUITTEK, J., ZSEBY, T., CLAISE, B. e ZANDER, S. **Requirements for IP Flow Information Export (IPFIX)**. RFC 3917 (Informational), outubro 2004. URL: http://www.ietf.org/rfc/rfc5286.txt.

RNP. **Rede Nacional de Ensino e Pesquisa**. 2011. Disponível: http://www.rnp.br/backbone/index.php [capturado em 22 de maio de 2011]

SANTOS, A. F. P. E SILVA, R. S. **Detecting Bandwidth DDoS Attack with Control Charts.** Em Network, 2007, 15th IEEE International Conference on, ICON 2007.

SANTOS, A. F. P. **Identificação e Análise de Comportamentos Anômalos**. 2009. Tese (Doutorado em Modelagem Computacional) - Laboratório Nacional de Computação Científica, LNCC, 2009.

SARDANA, A. e JOSHI, R. C. **Dual-Level Attack Detection and Characterization for Networks under DDoS** - International Conference on Availability, Reliability, and Security, ARES '10 , pp. 9 - 16, 2010.

SARRAUTE, C., MIRANDA, F. e ORLICKI, J. I. **Simulation of Computer Network Attacks** - Symposium on Computing Technology, Argentine, 2007.

SENGAR, H, WANG H., WIJESEKERA, D., e JAJODIA, S. **Detecting VoIP Floods Using the Hellinger Distance**. Em IEEE Transactions on Parallel and Distributed Systems, Journal Vol. 19, Issue 6, pp. 794 - 805, June 2008.

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

SILVEIRA, F., DIOT, C., TAFT, N. e GOVINDAN, R. **Empirical Evaluation of Network-wide Anomaly Detection, Thomsom Technical Report**, http://www.thlab.net/~fernando/papers/CR-PRL-2008-09-0004.pdf. Acessado em

29/11/2010.

SNOEREN, A. C., PARTRIDGE, C., SANCHEZ, L. A., JONES, TCHAKOUNTIO, C. E., F., SCHWARTZ, B., KENT, S. T., e STRAYER, W. T. **Single-packet IP traceback.** *IEEE/ACM Trans. Netw.*, 10(6):721–734, December 2002.

SPECHT, S., e LEE. R. **Distributed Denial of Service: Taxonomies of Attacks, Tools, and Countermeasures.** Proceedings of the 17th International Conference on Parallel and Distributed Computing Systems, 2004 International Workshop on Security in Parallel and Distributed Systems, pp. 543-550, September 2004.

SYMANTEC. **Anonymous Supporters Tricked into Installing Zeus Trojan.** Março 2012, Disponível: http://www.symantec.com/connect/blogs/anonymous-supporters-tricked-installing-zeus-trojan [capturado em 22 de maio de 2012]

WALFISH, M., VUTUKURU, M., BALAKRISHNAN, H., KARGER, D. R., e SHENKER, S. **DDoS defense by offense.** ACM Transactions on Computer Systems (TOCS), Journal Vol. 28, Issue 1, 2010.

WANG, L., WU, Q. e LIU, Y**. Design and Validation of PATRICIA for the Mitigation of Network Flooding Attacks** - IEEE CSE'09, Vanoucer, BC, Canada, 2009.

WIDE PROOJECTS. 2011. Disponível: http://www.wide.ad.jp/project/operation.html [capturado em 22 de novembro de 2011].

XIANG, Y., LI, K., e ZHOU, W. **Low-Rate DDoS Attacks Detection and Traceback by Using New Information Metrics**. IEEE Transactions on Information Forensics and Security, Vol. 6, No. 2, June 2011.

7 APÊNDICES

7.1 Apêndice 1: Índice de falsos positivos e falsos negativos no MAWILab

MAWILab		δ = 2 (limites rigorosos)		
Inserção (%)	**Estimador**	**Parâmetros**	**FPOS**	**FNEG**
50	EWMA	DEFAULT	2,06%	0,00%
		ÓTIMOS	4,17%	0,00%
	HW	DEFAULT	1,61%	0,00%
		ÓTIMOS	1,78%	0,00%
25	EWMA	DEFAULT	2,11%	0,00%
		ÓTIMOS	4,56%	0,00%
	HW	DEFAULT	1,61%	0,00%
		ÓTIMOS	1,78%	0,00%
20	EWMA	DEFAULT	2,22%	0,00%
		ÓTIMOS	4,72%	0,00%
	HW	DEFAULT	1,61%	0,00%
		ÓTIMOS	1,78%	0,00%
15	EWMA	DEFAULT	2,28%	20,00%
		ÓTIMOS	4,83%	0,00%
	HW	DEFAULT	1,61%	20,00%
		ÓTIMOS	1,78%	0,00%
10	EWMA	DEFAULT	3,06%	33,33%
		ÓTIMOS	5,00%	0,00%
	HW	DEFAULT	1,61%	53,33%
		ÓTIMOS	1,78%	6,67%
5	EWMA	DEFAULT	3,72%	40,00%
		ÓTIMOS	5,11%	20,00%
	HW	DEFAULT	1,61%	73,33%
		ÓTIMOS	1,78%	40,00%

MAWILab		δ = 3 (limites amplos)		
Inserção (%)	**Estimador**	**Parâmetros**	**FPOS**	**FNEG**
50	EWMA	DEFAULT	0,61%	0,00%
		ÓTIMOS	2,33%	0,00%
	HW	DEFAULT	0,94%	0,00%
		ÓTIMOS	0,00%	0,00%
25	EWMA	DEFAULT	0,67%	20,00%
		ÓTIMOS	2,50%	0,00%
	HW	DEFAULT	0,94%	6,67%
		ÓTIMOS	0,00%	0,00%
20	EWMA	DEFAULT	0,72%	20,00%
		ÓTIMOS	2,67%	0,00%
	HW	DEFAULT	0,94%	26,67%
		ÓTIMOS	0,00%	0,00%
15	EWMA	DEFAULT	0,72%	40,00%
		ÓTIMOS	2,78%	6,67%
	HW	DEFAULT	0,94%	46,67%
		ÓTIMOS	0,00%	0,00%
10	EWMA	DEFAULT	0,83%	53,33%
		ÓTIMOS	2,89%	13,33%
	HW	DEFAULT	0,94%	66,67%
		ÓTIMOS	0,00%	26,67%
5	EWMA	DEFAULT	1,33%	60,00%
		ÓTIMOS	3,11%	46,67%
	HW	DEFAULT	0,94%	80,00%
		ÓTIMOS	0,00%	93,33%

7.2 Apêndice 2: Índice de falsos positivos e falsos negativos no *trace* da RNP

RNP		δ = 2 (limites rigorosos)		
Inserção (%)	**Estimador**	**Parâmetros**	**FPOS**	**FNEG**
50	EWMA	DEFAULT	2,09%	13,33%
		ÓTIMOS	4,53%	0,00%
	HW	DEFAULT	0,80%	0,00%
		ÓTIMOS	12,84%	0,00%
25	EWMA	DEFAULT	2,19%	20,00%
		ÓTIMOS	4,83%	0,00%
	HW	DEFAULT	0,90%	6,67%
		ÓTIMOS	13,04%	0,00%
20	EWMA	DEFAULT	2,34%	20,00%
		ÓTIMOS	4,88%	0,00%
	HW	DEFAULT	1,00%	13,33%
		ÓTIMOS	13,09%	0,00%
15	EWMA	DEFAULT	2,59%	26,67%
		ÓTIMOS	4,93%	0,00%
	HW	DEFAULT	1,00%	20,00%
		ÓTIMOS	13,19%	0,00%
10	EWMA	DEFAULT	2,89%	53,33%
		ÓTIMOS	5,08%	0,00%
	HW	DEFAULT	1,05%	40,00%
		ÓTIMOS	13,34%	0,00%
5	EWMA	DEFAULT	3,19%	53,33%
		ÓTIMOS	5,43%	20,00%
	HW	DEFAULT	1,39%	73,33%
		ÓTIMOS	13,64%	0,00%

RNP		δ = 3 (limites amplos)		
Inserção (%)	**Estimador**	**Parâmetros**	**FPOS**	**FNEG**
50	EWMA	DEFAULT	1,10%	20,00%
		ÓTIMOS	2,89%	0,00%
	HW	DEFAULT	0,25%	6,67%
		ÓTIMOS	7,17%	0,00%
25	EWMA	DEFAULT	1,10%	33,33%
		ÓTIMOS	3,29%	0,00%
	HW	DEFAULT	0,30%	13,33%
		ÓTIMOS	7,22%	0,00%
20	EWMA	DEFAULT	1,05%	46,67%
		ÓTIMOS	3,38%	0,00%
	HW	DEFAULT	0,35%	20,00%
		ÓTIMOS	7,27%	0,00%
15	EWMA	DEFAULT	1,10%	53,33%
		ÓTIMOS	3,33%	0,00%
	HW	DEFAULT	0,35%	53,33%
		ÓTIMOS	7,27%	6,67%
10	EWMA	DEFAULT	1,24%	60,00%
		ÓTIMOS	3,48%	13,33%
	HW	DEFAULT	0,30%	73,33%
		ÓTIMOS	7,37%	13,33%
5	EWMA	DEFAULT	1,44%	73,33%
		ÓTIMOS	3,48%	33,33%
	HW	DEFAULT	0,30%	80,00%
		ÓTIMOS	7,32%	20,00%

7.3 Apêndice 3: Índice de falsos positivos e falsos negativos no *trace* da DARPA

DARPA		δ = 2 (limites rigorosos)		
Inserção (%)	**Estimador**	**Parâmetros**	**FPOS**	**FNEG**
50	EWMA	DEFAULT	1,33%	53,33%
		ÓTIMOS	3,11%	26,67%
	HW	DEFAULT	1,17%	60,00%
		ÓTIMOS	4,72%	6,67%
25	EWMA	DEFAULT	1,67%	66,67%
		ÓTIMOS	3,78%	40,00%
	HW	DEFAULT	1,56%	66,67%
		ÓTIMOS	4,83%	6,67%
20	EWMA	DEFAULT	1,78%	73,33%
		ÓTIMOS	3,89%	40,00%
	HW	DEFAULT	1,67%	73,33%
		ÓTIMOS	4,78%	6,67%
15	EWMA	DEFAULT	1,89%	73,33%
		ÓTIMOS	4,00%	46,67%
	HW	DEFAULT	1,78%	80,00%
		ÓTIMOS	4,83%	26,67%
10	EWMA	DEFAULT	2,06%	80,00%
		ÓTIMOS	4,22%	66,67%
	HW	DEFAULT	1,83%	86,67%
		ÓTIMOS	4,83%	26,67%
5	EWMA	DEFAULT	2,17%	86,67%
		ÓTIMOS	4,50%	66,67%
	HW	DEFAULT	1,94%	86,67%
		ÓTIMOS	5,17%	33,33%

DARPA		δ = 3 (limites amplos)		
Inserção (%)	**Estimador**	**Parâmetros**	**FPOS**	**FNEG**
50	EWMA	DEFAULT	0,50%	66,67%
		ÓTIMOS	1,56%	46,67%
	HW	DEFAULT	0,17%	66,67%
		ÓTIMOS	2,67%	20,00%
25	EWMA	DEFAULT	0,61%	73,33%
		ÓTIMOS	1,72%	53,33%
	HW	DEFAULT	0,22%	73,33%
		ÓTIMOS	2,67%	26,67%
20	EWMA	DEFAULT	0,61%	73,33%
		ÓTIMOS	1,83%	60,00%
	HW	DEFAULT	0,33%	80,00%
		ÓTIMOS	2,78%	33,33%
15	EWMA	DEFAULT	0,72%	80,00%
		ÓTIMOS	1,83%	80,00%
	HW	DEFAULT	0,39%	86,67%
		ÓTIMOS	2,83%	46,67%
10	EWMA	DEFAULT	0,72%	93,33%
		ÓTIMOS	1,94%	80,00%
	HW	DEFAULT	0,44%	86,67%
		ÓTIMOS	2,83%	53,33%
5	EWMA	DEFAULT	0,72%	93,33%
		ÓTIMOS	2,17%	86,67%
	HW	DEFAULT	0,50%	100,0%
		ÓTIMOS	2,89%	53,33%

7.4 Apêndice 4: Índice de FPOS e FNEG no MAWILab e na RNP para Divergências

MAWI			
Inserção (%)	**Métrica**	**FPOS**	**FNEG**
50	CHI-SQUARE	1,00%	6,67%
	HELLINGER	1,50%	0,00%
25	CHI-SQUARE	1,00%	6,67%
	HELLINGER	1,94%	0,00%
20	CHI-SQUARE	1,00%	6,67%
	HELLINGER	1,94%	13,33%
15	CHI-SQUARE	1,00%	6,67%
	HELLINGER	1,94%	40,00%
10	CHI-SQUARE	1,06%	0,00%
	HELLINGER	2,28%	60,00%
5	CHI-SQUARE	1,17%	6,67%
	HELLINGER	2,44%	80,00%

RNP			
Inserção (%)	**Métrica**	**FPOS**	**FNEG**
50	CHI-SQUARE	0,20%	13,33%
	HELLINGER	1,34%	13,33%
25	CHI-SQUARE	0,20%	26,67%
	HELLINGER	1,44%	13,33%
20	CHI-SQUARE	0,20%	33,33%
	HELLINGER	1,49%	13,33%
15	CHI-SQUARE	0,20%	33,33%
	HELLINGER	1,64%	20,00%
10	CHI-SQUARE	0,20%	33,33%
	HELLINGER	1,99%	26,67%
5	CHI-SQUARE	0,20%	33,33%
	HELLINGER	2,24%	33,33%