Universidade Federal de Santa Catarina
Departamento de Informática e Estatística
Curso de Ciências da Computação

Detecção e Classificação de Anomalias no Tráfego de Redes de Computadores

GUILHERME FERNANDES RAPHANELLI

Florianópolis – SC
Ano 2008 / 2

GUILHERME FERNANDES RAPHANELLI

Detecção e Classificação de Anomalias no Tráfego de Redes de Computadores

Trabalho de conclusão de curso apresentado como parte dos requisitos para obtenção do grau de Bacharel em Ciências da Computação

Orientador: Prof. Dr. Mario Antonio Ribeiro Dantas
Co-Orientador: Edison Tadeu Lopes Melo

Banca Examinadora
Prof. Dr. João Bosco Mangueira Sobral

Agradecimentos

Duas pessoas foram essenciais para tornar este trabalho possível: o pesquisador Philippe Owezarski e a minha amada mãe Melania. Devo agradecer imensamente ao Philippe por ter acreditado em mim, tornado viável meu estágio de pesquisa no LAAS e me orientado durante os seis meses que lá passei. À minha mãe agradeço todos os dias (mesmo que silenciosamente) por ter tornado possível que a minha vida fosse da maneira que foi: livre e cheia de oportunidades.

Agradeço muito aos meus orientadores, Dantas e Melo, pelos valiosos comentários e conselhos, tanto para a criação deste trabalho quanto para os mais variados assuntos.

Sendo este um trabalho de conclusão de curso, aproveito para agradecer também aos professores do curso de Ciências da Computação da UFSC pelo conhecimento repassado durantes todos estes anos, e aos meus amigos e familiares pelo apoio e momentos de descontração que perduraram por todos estes anos, tornando a caminhada viável e agradável.

Resumo

As redes de computadores são hoje infra-estrutura crítica e devem ter uma qualidade de serviço mínima garantida. Anomalias no tráfego de rede podem deteriorar seriamente esta qualidade de serviço, tornando essencial que os administradores de rede identifiquem e eliminem-nas o mais rápido possível. A área de detecção de anomalias no tráfego de rede tem como objetivo desenvolver métodos que facilitem a detecção destas anomalias. Apesar do estado avançado desta área, os administradores de redes de médio e grande porte não conseguem lidar manualmente com a quantidade de anomalias detectadas. Maiores informações a respeito de cada anomalia devem ser providas para que eles possam priorizar o tempo gasto com a análise manual. A área de classificação automatizada, que ainda está em seus passos iniciais, busca complementar as informações providas no alerta de uma anomalia a fim de dar a base necessária para esta priorização.

Este trabalho apresenta um algoritmo para fazer a detecção e classificação automatizada de anomalias no tráfego de rede, operando através de quatro etapas principais: (i) realizar a detecção de anomalias utilizando múltiplas métricas de volume de tráfego, (ii) (para cada anomalia) identificar os pacotes ou *flow records* responsáveis, (iii) usar estes pacotes ou *flow records* para derivar diversas novas métricas específicas da anomalia, e (iv) classificar a anomalia utilizando estas métricas e um módulo de classificação baseado em regras. Este método provê (i) a expressividade necessária para distinguir confiavelmente entre diferentes tipos de anomalias, (ii) uma descrição mais completa sobre a anomalia, e (iii) a simplicidade e flexibilidade necessária para que os administradores de rede entendam e manipulem facilmente o processo de classificação. Uma validação estatística é feita utilizando dados com características bem diferentes, retirados do projeto METROSEC e do repositório de tráfego MAWI.

Palavras-chave: Detecção de Anomalias, Classificação de Anomalias, Gerenciamento de Redes

Abstract

Computer networks are today critical infrastructure and must have a minimum quality of service guaranteed. Network traffic anomalies may disrupt this quality of service badly, thus network operators must identify and mitigate them quickly. The field of network traffic anomaly detection has the main goal of developing methods that facilitate the detection of these anomalies. In spite of the advanced state of this field, network operators of medium and large networks cannot handle the amount of anomalies detected. More information must be given regarding each anomaly so that these operators can efficiently prioritize the time spent in manual analysis. The field of automated classification, which is still incipient, aims to give more, meaningful information about a detected anomaly, giving network operators a better understanding of the anomaly prior to manual analysis.

In this paper, we present a new algorithm for automated classification of network traffic anomalies. The algorithm relies on four steps: (i) detect anomalies using traffic volume metrics, (ii) (for each anomaly) identify all (or most) related packets or flow records; (iii) use these packets or flow records to derive distinct metrics directly related to the anomaly; and (iv) classify the anomaly using these metrics in a signature-based approach. We show how this approach provides (i) the expressiveness necessary to reliably distinguish between different types of anomalies, (ii) a richer set of information about detected anomalies and (iii) the flexibility needed by network administrators to understand and manipulate the classification process. We validate our algorithm on two different datasets: the METROSEC project database and the MAWI traffic repository.

Keywords: Network Anomaly Detection, Network Anomaly Classification, Network Management.

Sumário

Lista de Abreviaturas e Siglas

BGP: Border Gatewar Protocol

DARPA: Defense Advanced Research Projects Agency

DoS: Denial of Service

DDoS: Distributed Denial of Service

FC: Flash Crowd

FPR: False Positive Rate

ICMP: Internet Control Message Protocol

ISO: International Organization for Standardization

LAAS: Laboratoire d'Architecture et d'Analyse des Systèmes

LRD: Long-Range Dependence

METROSEC: Metrology for Security and Quality of Service

MIT: Massachusetts Institute of Technology

NADA: Network Anomaly Detection Algorithm

NMS: Network Management System

RENATER: Le Réseau National de télécommunications pour la Technologie l'Enseignement et la Recherche

ROC: Receiver Operating Characteristic

TCP: Transmission Control Protocol

TPR: True Positive Rate

TTL: Time to Live

UDP: User Datagram Protocol

1 INTRODUÇÃO

A monitoração de redes de computadores é uma atividade essencial para um gerenciamento bem-sucedido das mesmas. Quanto mais as redes de computadores crescem, mais difícil é a tarefa de mantê-las seguras e eficientes. Isto se dá não apenas pela crescente diversidade da rede, que passou a servir as mais diversas aplicações, de celulares a serviços essenciais de grandes empresas, mas também pelo aumento no volume de tráfego, que torna a análise manual uma tarefa limitada. Os administradores de rede precisam garantir que o desempenho e a qualidade de serviço da rede sejam aceitáveis, e para isto devem lidar com diversos desafios.

Anomalias do tráfego de rede podem causar graves problemas no desempenho da rede, e devem ser identificadas e eliminadas pelos administradores de rede o mais rápido possível. Um tipo específico de anomalia, as chamadas anomalias de volume, é responsável por modificações inusitadas nas características de volume da rede, como a quantidade de pacotes. Estas anomalias podem ser causadas por diversos eventos: de problemas físicos ou técnicos da rede (e.g. quedas de energia, configurações problemáticas em roteadores), a um comportamento malicioso intencional (e.g. ataques de negação de serviço, tráfego relativo a *worms*), a mudanças abruptas causadas por tráfego legítimo (e.g. *flash crowds* [26], *alpha flows*). As anomalias causadas por ataques de serviço são especialmente importante por serem comuns e terem um forte impacto na qualidade de serviço da rede.

Este trabalho apresenta uma contribuição para a área de detecção e classificação automatizada destas anomalias. Apesar da área de detecção de anomalias do tráfego de rede estar bem avançada, mostramos as dificuldades apresentadas pela falta de métodos confiáveis para a classificação automatizada, principalmente no que diz respeito à análise manual feita pelos administradores de rede. O tema deste trabalho é apresentar um algoritmo novo para realizar a detecção e, principalmente, a classificação automatizada das anomalias do tráfego de rede. O algoritmo foi desenvolvido principalmente pelo autor durante um estágio

de pesquisa no Laboratório de Análise e Arquitetura de Sistemas (LAAS) do Centro Nacional de Pesquisa Científica da França, em Toulouse.

1.1 OBJETIVOS

O objetivo principal deste projeto é apresentar um algoritmo para a detecção e, principalmente, classificação automatizada de anomalias no tráfego de rede. Para o desenvolvimento do algoritmo, um algoritmo de detecção deve ser adaptado a um método de classificação automatizada. A caracterização de diversos tipos de anomalias de tráfego de rede deve ser feita a fim de se definir claramente as características utilizadas para se realizar uma classificação confiável. Pelo menos as anomalias causadas por ataques de negação de serviço, incluindo os ataques de baixa intensidade, devem ser estudadas e tratadas corretamente pelo algoritmo.

1.2 JUSTIFICATIVAS

As redes de computadores são componentes essenciais da sociedade moderna. As redes mais críticas, como as redes de empresas ou as de provedores de serviço de Internet, devem ter uma qualidade de serviço mínima garantida. Anomalias no tráfego de rede podem deteriorar seriamente esta qualidade de serviço, tornando essencial que os administradores de rede identifiquem-nas e eliminem-nas o mais rápido possível.

Diversos algoritmos de detecção de anomalias foram criados para auxiliar os administradores de rede na identificação de anomalias de tráfego de rede. Estes algoritmos evoluíram de forma a, além de detectar estas anomalias com uma precisão surpreendente, identificar os fluxos responsáveis pelas anomalias. Isto facilita bastante o trabalho dos administradores de rede que precisam então analisar manualmente o tráfego para entender a natureza da anomalia e quais medidas devem ser tomadas. Entretanto, a quantidade de anomalias presente no tráfego de uma rede de médio a grande porte torna proibitiva a análise manual de todas as anomalias detectadas.

Os administradores de rede devem então receber maiores informações a respeito de cada anomalia para que possam priorizar o tempo gasto com a análise manual. A classificação automatizada busca complementar as informações providas no alerta de uma anomalia a fim de dar a base necessária para esta priorização. O tipo da anomalia é uma das informações essenciais que devem ser providas. Este trabalho busca apresentar um algoritmo que realize a detecção e classificação automatizada das anomalias de tráfego de rede de forma a permitir que os administradores da rede consigam utilizar seu tempo da forma mais eficiente possível.

2 GERENCIAMENTO DE REDES DE COMPUTADORES

Este capítulo destina-se a esclarecer conceitos teóricos para entender-se o desenvolvimento do trabalho, e, principalmente, a realizar a análise de trabalhos correlatos. Uma pequena introdução é dada ao tema de gerenciamento de redes de computadores, seguida de uma explanação da área de monitoramento e caracterização do tráfego de rede. Em seguida, os métodos de se obter os dados para este monitoramento são discutidos. As anomalias de tráfego de rede e a área de detecção de anomalias são introduzidas depois. Finalmente, trabalhos correlatos da área de caracterização e classificação do tráfego são abordados.

As redes de computadores têm hoje um papel fundamental na sociedade. Da Internet a redes privadas de empresas, a infra-estrutura que permite esta comunicação é considerada de missão crítica, não podendo parar ou ter sua qualidade degradada abaixo de um determinado nível. O gerenciamento de redes de computadores pode ser definido como o conjunto de atividades que procuram manter a rede funcionando da melhor forma possível. O Modelo de Gerenciamento de Redes ISO [24] divide o gerenciamento de redes em cinco áreas: gerência de configuração, gerência de faltas, gerência de desempenho, gerência de segurança e gerência de contabilidade.

A gerência de configuração se encarrega de monitorar as informações de configuração da rede e dos dispositivos que a compõe. Isto abrange atividades como verificar a topologia atual da rede, descobrindo novos elementos e as interconexões entre eles, e a alteração da configuração dos elementos gerenciados, entre outras. O objetivo principal da gerência de faltas é a detecção, isolamento e conserto de falhas na rede [24]. A gerência de desempenho por sua vez tem como objetivo a monitoração e a análise de indicadores de desempenho da rede. Isto costuma ser realizado em três etapas [24]. Primeiro, os dados dos indicadores de desempenho são coletados. Estes dados são então analisados para determinar o seu nível normal (*baseline*). Finalmente, níveis limites são definidos para cada variável importante de forma que valores que os ultrapassem sejam indícios de problema no desempenho. A gerência de segurança é responsável pela proteção dos recursos da rede. Ela deve implementar e monitorar a política de segurança da

rede. A gerência de contabilidade tem como objetivo contabilização do uso dos recursos da rede feito pelos usuários ou grupos de forma a garantir que limites de utilização predefinidos não sejam ultrapassados.

Apesar de estas áreas estarem estritamente definidas, muitas das operações de gerência de redes atuam em diversas áreas de gerência ao mesmo tempo. Este é o caso da detecção e classificação de anomalias do tráfego de redes, ilustrado pelos seguintes pontos. As anomalias do tráfego de rede são importantes por afetarem diretamente o desempenho da rede. Estas anomalias podem ser causadas por falhas na rede, problemas de configuração ou ataques maliciosos, entre outras razões. Ainda mais, um sistema automatizado pode reconfigurar a rede após a detecção e classificação de uma anomalia para tentar minimizar seu impacto. Neste trabalho utilizaremos o termo *gerenciamento de redes* de forma genérica, abrangendo todas as áreas de gerência mencionadas.

2.1 MONITORAÇÃO E CARACTERIZAÇÃO DE TRÁFEGO

A monitoração e caracterização do tráfego de redes são atividades comumente praticadas no gerenciamento de redes. A monitoração de tráfego se preocupa com as características de fluxo da rede, sendo importante para o gerenciamento e planejamento de redes de computadores [6]. A monitoração pode ser feita de forma mais genérica, observando métricas que dizem respeito ao tráfego todo, como a quantidade de bytes que trafegaram por uma determinada interface de um roteador ou a quantidade de erros detectados nela, e de forma mais específica, realizando a caracterização dos fluxos. A definição clássica de fluxo (*flow*), criada por Claffy et. al., é relativamente abstrata [11]: um fluxo é uma seqüência de pacotes que se enquadram em um determinado critério, correspondente a uma comunicação entre duas entidades da rede. Estes critérios costumam estar relacionados a características de tráfego como o protocolo utilizado, os endereços IP de origem e destino, as portas de origem e destino, entre outros. Apesar do tráfego de rede poder ser caracterizado por diversos critérios, é mais fácil agregar o tráfego em apenas uma dimensão ao mesmo tempo [18]. Entretanto, este tipo de agregação resulta em perda de informações que podem ser interessantes ao administrador. Por

exemplo, agregando o tráfego através de uma visão de aplicação (e.g. usando o protocolo e as portas) um administrador de rede pode concluir que aplicações *peer-to-peer* de troca de arquivos estão sendo usadas em grande escala. Mas uma análise mais detalhada do tráfego, incluindo agora também endereços IP de origem e destino, pode mostrar que poucas máquinas são responsáveis pela grande maioria deste tráfego.

Outra característica da monitoração do tráfego de rede diz respeito ao método de captura da informação. A monitoração e caracterização do tráfego são normalmente feitas utilizando medições passivas. Medições passivas são aquelas que não introduzem tráfego adicional na rede, sendo assim não intrusivas [6]. Estas medições podem ser feitas pelos equipamentos onde o tráfego passa naturalmente, como os roteadores e switches, ou por equipamentos externos que recebem uma cópia dos pacotes ou informações dos fluxos que estão trafegando na rede. Apesar dos roteadores da rede serem pontos ideais para estas medições, eles não costumam estar equipados para caracterizar exaustivamente o tráfego, mas mantêm contadores genéricos (e.g. a quantidade de pacotes descartados por uma interface). Para se obter informações mais detalhadas do tráfego de rede, utiliza-se a captura de pacotes ou protocolos de monitoramento como o Cisco NetFlow.

O NetFlow [10] oferece uma maneira de agregar o tráfego em fluxos, normalmente utilizando a tupla (endereço IP de origem, porta de origem, endereço IP de destino, porta de destino e protocolo), diretamente nos roteadores, repassando apenas as informações estatísticas agregadas (e.g. quantidade de pacotes e de bytes do fluxo) para uma máquina coletora. A informação fornecida sobre cada fluxo pelo NetFlow é chamada de *flow record*. A vantagem deste método é a redução drástica da quantidade de informação que deve ser processada pela máquina coletora, algo de suma importância nas velocidades multi-gigabit de hoje. Entretanto, o processamento adicional incorrido ao roteador para realizar esta tarefa pode prejudicar seu funcionamento. Para minimizar este efeito, muitos roteadores também implementam amostragem do tráfego, selecionando apenas uma fração dos fluxos (e.g. 1 em cada 1000) para serem computados. Esta medição amostrada em nível de fluxo provê um balanço entre escalabilidade e detalhamento, já que restrições de desempenho podem ser tratadas reduzindo a taxa de amostragem [18].

A perda de expressividade e a distorção de características estatísticas do tráfego derivadas da medição amostrada em nível de fluxo podem requerer o uso do método de captura de pacotes. Algumas aplicações de monitoramento de tráfego, como sistemas de detecção de intrusão, muitas vezes necessitam analisar os dados contidos dentro dos pacotes que compõe o fluxo. Outras aplicações podem exigir uma precisão maior para a caracterização do tráfego do que a que pode ser obtida com as informações dos fluxos amostrados. A detecção e a classificação de anomalias no tráfego de rede podem ser feitas utilizando qualquer um dos dois métodos apresentados. Apesar da detecção de algumas anomalias poder ser feita utilizando as informações providas pelos contadores genéricos dos roteadores, este método não é recomendado para análises mais avançadas por possuir diversas limitações, como não conseguir identificar informações dos fluxos anômalos.

2.2 DETECÇÃO DE ANOMALIAS NO TRÁFEGO DE REDES

As anomalias de tráfego de rede são inerentes à maneira que a Internet funciona hoje. Estas anomalias, definidas como um desvio acentuado de uma determinada característica do tráfego relativo a um modelo de comportamento normal [30], podem ser causadas por diversos tipos de eventos: de problemas físicos ou técnicos da rede (e.g. quedas de energia, configurações problemáticas em roteadores), a um comportamento malicioso intencional (e.g. ataques de negação de serviço, tráfego relativo a *worms*), a mudanças abruptas causadas por tráfego legítimo (e.g. *flash crowds, alpha flows*). Além desta grande diversidade de anomalias, o próprio tráfego de Internet possui uma forte variabilidade de seu volume, graças a características como auto-similaridade [38], fractalidade múltipla [21] e dependência de longa distância (LRD) [16].

A área de detecção de anomalias de tráfego de rede busca desenvolver métodos que identifiquem quando o tráfego está se comportando de maneira anômala. A identificação destas anomalias é importante para se garantir o desempenho e a qualidade de serviço das redes de computadores. Mesmo com as dificuldades mencionadas acima, progresso constante tem sido feito na área de detecção de anomalias de tráfego de rede. Atualmente existe uma vasta literatura

abordando o tema de detecção. A maior parte dos métodos de detecção analisa variações estatísticas de métricas de volume de tráfego, como o número de pacotes, número de bytes ou número de fluxos da rede, métricas de características do tráfego, como distribuições de endereços IP ou portas, ou uma combinação destas. A detecção pode ser realizada de maneira temporal, quando se compara valores atuais com um comportamento histórico, ou espacial, quando se compara o valor em um ponto da rede com o valor de outros pontos (i.e. um ponto pode ser um link, por exemplo).

Os trabalhos de [5, 2, 28, 30, 29] são referências clássicas na área de detecção de anomalias do tráfego de rede, e trabalhos como [32, 14, 7, 39] são bons exemplos da evolução das técnicas do estado da arte e de técnicas inovadoras. Estes trabalhos se baseiam em uma enorme variedade de técnicas para definir e identificar comportamento anômalo. Por exemplo, [2] implementou uma análise de sinais baseada em *wavelets* para detectar e caracterizar anomalias que afetam o número de bytes do tráfego de um link da rede. Por outro lado, os autores de [29] buscam identificar e caracterizar anomalias de tráfego de rede com uma visão global da rede (i.e. não restritos a um único link) utilizando o método de subespaço aliado ao calculo de entropia das distribuições resultantes. As técnicas de detecção vão de análises estatísticas simples, como média e desvios padrões, às técnicas mencionadas acima, a métodos que utilizam técnicas de inteligência artificial e reconhecimento de imagens.

2.2.1 NETWORK ANOMALY DETECTION ALGORITHM – NADA

O *Network Anomaly Detection Algorithm* [19], ou NADA, é um algoritmo desenvolvido no contexto do projeto METROSEC [34] para detectar e caracterizar anomalias do tráfego de rede. Este algoritmo utiliza um método de tomografia com uma análise de múltiplos critérios, escalas e níveis. A idéia central do algoritmo para detecção é que qualquer anomalia será responsável por algum nível de variação em ao menos um dos critérios utilizados, em alguma escala de tempo e em algum nível de agregação IP. Os critérios utilizados costumam serem métricas de volume de tráfego, como o número de pacotes, mas podem ser qualquer métrica que varie em função do tempo. A fórmula 1 [20] representa o funcionamento do algoritmo de

detecção do NADA. Esta fórmula se baseia no conceito de *deltoids absolutos* desenvolvido por [12] para detectar mudanças significativas no tráfego. Um *deltoid* é basicamente uma variação significativa de um parâmetro.

$$X = \{x_1, x_2, \ldots, x_n\}, \qquad x_i = \{\#packets | \#bytes | \#syn\}/\Delta$$
$$P = \{p_1, p_2, \ldots, p_{n-1}\}, \; p_i = x_{i+1} - x_i \qquad (1)$$
$$\begin{cases} p_i \geq K\sigma_p, & anomalous \\ p_i < K\sigma_p, & not\ anomalous \end{cases}$$

A fórmula 1 pode ser explicada da seguinte forma. Dado arquivo de captura de tempo T e uma granularidade de escala de tempo Δ (e.g. 30 segundos), dividir os dados em N *slots* (i.e. seções de tempo) em que N ϵ [1, T/Δ]. Para cada *slot i* obter a série de tempo X para cada métrica sendo considerada. Calcular os *deltoids* absolutos P de X e calcular seu desvio padrão. Para qualquer *pi* sobre o valor limite *K*σp* marque o seu *slot* como anômalo. O uso de *deltoids* é essencial por considerar a variação sobre a amplitude da curva ao invés da variação do tráfego de rede, já que o último é insignificante devido à alta variabilidade natural do tráfego de rede.

No NADA, a Fórmula 1 é aplicada de forma recursiva. Cada nível de iteração utiliza um nível de agregação do espaço IP diferente. Na primeira iteração todo o espaço IP é considerado (i.e. o tráfego é agregado completamente), e os *slots* de tempo de duração Δ que possuem possíveis anomalias são identificados. A cada nova iteração, os *flows* de cada *slot* previamente identificado como anômalo são analisado com agregações mais específicas (i.e. de máscara /1 a máscara /32). É importante ressaltar que apenas os *slots* que foram identificados como anômalos para um nível de agregação maior (i.e. nível /0 no início) continuam sendo analisados em níveis mais específicos em busca de anomalias.

2.3 CARACTERIZAÇÃO DE ANOMALIAS DE TRÁFEGO

A caracterização das anomalias de tráfego de rede tem sido amplamente estudada. [2] utilizou uma técnica de análise de sinal baseada em *wavelets* nos dados de volume do tráfego de um link para caracterizar quatro classes gerais de

anomalias: falhas de rede, *flash crowds*, ataques e falhas de medição. Lakhina et. al. usou o método do subespaço para caracterizar diversos tipos distintos de anomalias em escala de rede global baseado em métricas de volume e métricas [29] de características do tráfego [31]. Esforço considerável tem se dado a fim de caracterizar tipos específicos de anomalias. Por exemplo, os ataques de negação de serviço distribuídos e não distribuídos receberam uma análise extensiva em [36, 23, 35]. Jung et. al. [26] estudou as diferenças de comportamento entre ataques de negação de serviço distribuídos e *flash crowds* sob a perspectiva de um servidor web.

Alguns trabalhos também propuseram maneiras de se obter maiores informações sobre o tráfego de rede, por exemplo, através de agrupamento em clusters [31, 18], e maneiras de se priorizar ocorrências, como selecionar *heavy-hitters* [41] ou usar heurísticas como *unexpectedness* [18]. O método não supervisado de [31] cria clusters baseados em como as anomalias estão representadas no espaço de entropia de suas características de tráfego (i.e. endereços IP e portas). Este método agrupa anomalias com características similares, mas não distingue entre diferentes tipos de anomalias. Os administradores de rede ainda precisam checar manualmente cada anomalia para determinar seu tipo, mas, se houverem anomalias já conhecidas suficientes dentro de cada cluster, os administradores tem uma forma melhor de priorizar a análise manual. O trabalho de [18] não é específico a anomalias, mas mostra como a caracterização automática do tráfego pode ajudar muito os administradores de rede em suas tarefas diárias. Selecionar os *heavy-hitters* é um método comumente utilizado no gerenciamento de redes (e.g. para monitoração do tráfego, contabilidade, etc.). Entretanto, este método pode não ser uma solução absoluta para a detecção de anomalias, considerando que as anomalias de baixa intensidade podem ser dominadas por fortes variações normais do tráfego de rede. O método de classificação apresentado neste trabalho complementa estes esforços de caracterização automatizada, provendo maiores informações sobre cada anomalia.

3 ALGORITMO PARA DETECÇÃO E CLASSIFICAÇÃO DE ANOMALIAS NO TRÁFEGO DE REDE

Nesta seção apresentamos um algoritmo que realiza a detecção e, principalmente, a classificação automatizada de anomalias no tráfego de rede. Primeiro fazemos uma revisão dos problemas atuais no estado da arte da área de detecção de anomalias de tráfego de rede e da necessidade de classificação automatizada. Depois abordamos como estas técnicas devem ser integradas para fornecer um conjunto melhor de informações para que os administradores de rede possam identificar e resolver rapidamente as anomalias encontradas. Em seguida, introduzimos as idéias principais do nosso algoritmo, que permite a integração das diferentes etapas, a identificação de uma gama maior de informações sobre a anomalia e finalmente a classificação automatizada de diferentes tipos de anomalias de tráfego de rede de forma confiável. O resto do capítulo abrange os detalhes de cada etapa do algoritmo.

A área de detecção de anomalias de tráfego de rede tem evoluído satisfatoriamente para resolver os desafios de se detectar os diversos tipos de anomalias de tráfego de rede. Mas existe um problema que é apenas parcialmente tratado pelas técnicas de detecção de anomalias atuais. Com o grande número de anomalias presentes no tráfego de rede de grandes redes corporativas, e um mecanismo de detecção com alta sensibilidade, apenas identificar que uma anomalia está ocorrendo, ou ocorreu em um determinado instante, não é o suficiente para que os administradores de rede possam-na identificar e eliminar com eficiência, uma vez que a análise manual de ocorrências despende uma quantidade de tempo considerável. Técnicas mais novas de detecção de anomalias evoluíram ao ponto de poder identificar quais são os fluxos (i.e. endereços IP e portas de origem e destino) que causam a anomalia, e quais variáveis estão sendo afetadas por eles. Entretanto, esta informação, apesar de ser um progresso considerável, ainda não é suficiente para que os administradores consigam lidar com o grande número de anomalias existentes no tráfego.

A classificação automatizada, área onde está a principal contribuição deste trabalho, vem em auxílio dos administradores para este problema. A idéia central da classificação automatizada de anomalias de tráfego de rede é prover o máximo de informação agregada possível sobre as anomalias detectadas, a fim de propiciar uma base de decisão maior para a priorização de anomalias na análise manual feita pelos administradores, e, em certos casos, permitir estratégias automáticas de tratamento destas anomalias (e.g. através da reengenharia do tráfego). Uma informação essencial é o tipo da anomalia em questão. A caracterização das anomalias de tráfego de rede recebeu atenção considerável da comunidade científica. Alguns trabalhos caracterizaram diversos tipos de anomalias ao mesmo tempo, utilizando métricas de volume de tráfego de rede, *features* de tráfego (i.e. endereços IP e portas), ou ambos (ver subcapítulo 2.3). Outros focalizaram em um tipo específico de anomalia, como falhas em enlaces ou problemas de configuração de BGP, e, em alguns casos, até subtipos foram identificados (e.g. os diversos tipos de DDoS).

A análise destes trabalhos de caracterização mostra claramente que utilizar apenas métricas de volume e distribuições de endereço IP e porta não é suficiente para distinguir entre diferentes tipos de anomalias sem ambigüidade. Apesar dos trabalhos de caracterização que trataram diversos tipos ao mesmo tempo não terem procurado resolver esta ambigüidade, pode-se recorrer aos estudos específicos de cada anomalia para obter-se uma visão mais completa das influências de cada tipo no tráfego de rede. A partir deste conhecimento, torna-se possível desenvolver um algoritmo que, além de detectar a anomalia, as métricas de volume afetadas, e os fluxos responsáveis, recupere informações mais específicas sobre cada anomalia a fim de se poder classificá-las de forma confiável.

Para atender estas necessidades, o algoritmo desenvolvido neste trabalho possui quatro etapas principais: (i) realizar a detecção de anomalias utilizando múltiplas métricas de volume de tráfego, (ii) (para cada anomalia) identificar os pacotes ou *flow records* responsáveis, (iii) usar estes pacotes ou *flow records* para derivar diversas novas métricas específicas da anomalia, e (iv) classificar a anomalia utilizando estas métricas e um módulo de classificação baseado em regras. A inovação do algoritmo está na busca dos pacotes ou *flow*

records específicos da anomalia, com a subseqüente derivação de métricas usando apenas o tráfego diretamente relacionado. Este método provê (i) a expressividade necessária para distinguir confiavelmente entre diferentes tipos de anomalias, (ii) uma descrição mais completa sobre a anomalia, e (iii) a simplicidade e flexibilidade necessária para que os administradores de rede entendam e manipulem facilmente o processo de classificação. Esta flexibilidade também permite que o processo de classificação seja utilizado como um filtro para os falsos positivos do algoritmo de detecção. Justificamos estas afirmações e explicamos em detalhes as etapas do algoritmo nas seções que seguem.

3.1 PRIMEIRA ETAPA: DETECÇÃO DE ANOMALIAS

A primeira etapa do nosso algoritmo se encarrega da detecção das anomalias no tráfego de rede. Para que esta etapa possa ser realizada, os dados (i.e. os pacotes ou *flow records* do tráfego) devem estar disponíveis. A análise pode ser realizada (semi-)online, ocorrendo em tempo real, ou offline, sendo feita com dados históricos. Neste trabalho utilizamos apenas arquivos de captura de pacotes e análise offline, mas resultados equivalentes seriam obtidos com o uso de *flow records*. Os arquivos de captura e os métodos que utilizamos para testar o algoritmo são discutidos no capítulo 4.

O algoritmo de detecção utilizado é baseado no algoritmo discutido no subcapítulo 2.2.1. Para a detecção, utilizamos *deltoids* absolutos (ver seção 2.2.1) de séries de tempo de métricas do volume de tráfego de rede na construção de um modelo estatístico para o comportamento padrão do tráfego de rede. Qualquer *deltoid* que desviar mais que um determinado valor deste modelo, neste caso K desvios padrões, é considerado anômalo. É importante ressaltar que o uso de *deltoids* é essencial por considerar a variação sobre a amplitude da curva ao invés da variação do tráfego de rede, já que o último é insignificante devido à alta variabilidade natural do tráfego de rede. Neste trabalho utilizamos uma granularidade fixa de 30 segundos e três métricas de volume: número de pacotes, número de bytes e número de syn. A métrica número de syn é uma aproximação para a métrica, comumente utilizada, número de novos fluxos. Ela consiste no

número de pacotes TCP que possuem apenas o flag SYN habilitado, e diminui consideravelmente o processamento necessário quando os dados estão na forma de pacotes de rede. Entretanto, mesmo quando utilizamos pacotes de rede como dados de entrada, o conceito de fluxo continua sendo essencial. Os fluxos podem ser agregados em uma ou mais de suas dimensões (i.e. IPs, portas, protocolo) e a análise dos algoritmos de detecção ocorre em cima desta agregação.

Nosso algoritmo de detecção é baseado em níveis de agregação do endereço IP de destino dos pacotes. Em contraste ao conceito de tomografia utilizado por [20, 19] (ver seção 2.2.1), onde cada nível de agregação é analisado em seqüência, com apenas as anomalias de cada nível sendo analisadas no próximo, nosso algoritmo faz a análise de todos os níveis ao mesmo tempo, correlacionando-os através de diferentes atributos (o conceito de atributos de uma anomalia será formalizado na seção 3.2). Um nível de agregação é definido pela quantidade de bits do endereço IP de destino que estão sendo considerados, equivalentemente ao conceito de *bitmask*. A figura 1 ilustra como os níveis de agregação estão divididos no espaço IP. Por exemplo, o nível 1 possui dois grupos (i.e. os que possuem o primeiro bit do endereço IP igual a 1 e os que o tem igual a 0), sendo que todos os endereços IPs de destino com o primeiro octeto entre 0 e 127 estarão no primeiro grupo, e os demais no segundo. O nível zero (0) é o equivalente à agregação total do tráfego, e é sempre considerado. Considerar todos os outros níveis de agregação demanda um enorme poder de processamento e uma grande quantidade de memória, com um ganho de informação muito pequeno entre um nível e o seguinte. Assim, é preferível que poucos níveis sejam selecionados baseando-se na topologia da rede, nos dados e na análise pretendida. Seguindo o exemplo de [20], neste trabalho utilizamos apenas três dos outros 32 níveis possíveis nesta primeira etapa do algoritmo: 8, 16 e 24. Estes valores foram definidos para análises genéricas, e devem ser revistos para situações particulares. Por exemplo, se uma rede possui seu espaço de endereçamento divido em sub-redes de *bitmask* /26, seria mais apropriado utilizar o nível 26 ao invés do nível 24.

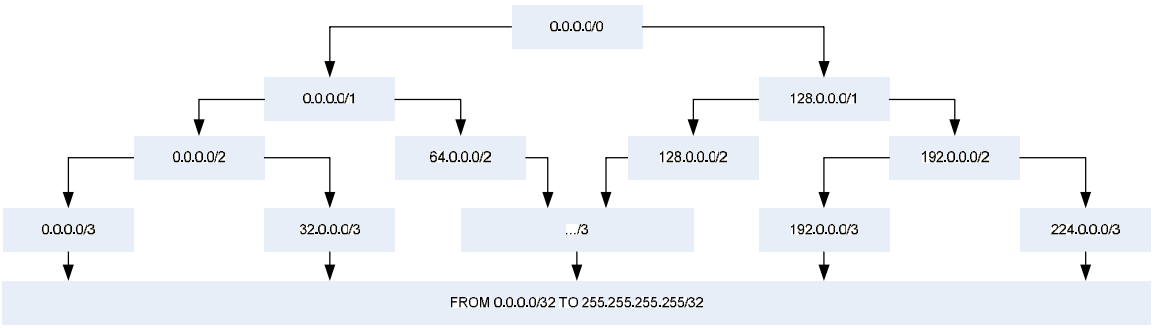

Figura 1 - Níveis de agregação de endereços IP

O algoritmo foi modificado para considerar todos os níveis de agregação ao mesmo tempo para aumentar as chances de detecção de anomalias de baixa intensidade. O método de tomografia utilizado por [20] permite a identificação dos fluxos responsáveis e destrincha o impacto causado por cada nível de agregação, mas apenas detecta anomalias que geram uma variação considerável na agregação total do tráfego (i.e. nível 0). A figura 2 mostra graficamente como uma anomalia de baixa intensidade pode ser encolhida pela magnitude do tráfego agregado e passar despercebida por algoritmos de detecção que consideram apenas o tráfego agregado. A anomalia mostrada na figura 2 é causada por um ataque de negação de serviço distribuído (DDoS , ver seção 3.4.1) e, dos níveis escolhidos, pode ser detectada apenas no nível 24. Detectar anomalias de baixa intensidade é importante por diversas razões. No caso particular de ataques DDoS, se a detecção da anomalia for feita perto da vítima, a degradação do desempenho e da qualidade de serviço da rede já é alta. Assim, é necessário detectar estes ataques perto das origens, onde a intensidade das anomalias é baixa. Mesmo quando os ataques DDoS de baixa intensidade não esgotam os recursos da rede, eles impactam diretamente a dependência de longa distância (LRD) do tráfego causando a degradação da qualidade de serviço da rede [37]. Outra situação onde é importante detectar anomalias de (relativa) baixa intensidade é quando se está analisando um link onde os dados de diversas redes trafegam, como no caso de um link de backbone de um provedor de serviço. Neste caso, teremos naturalmente o tráfego de redes grandes que quando agregadas podem tornar as anomalias no tráfego de clientes menores imperceptíveis.

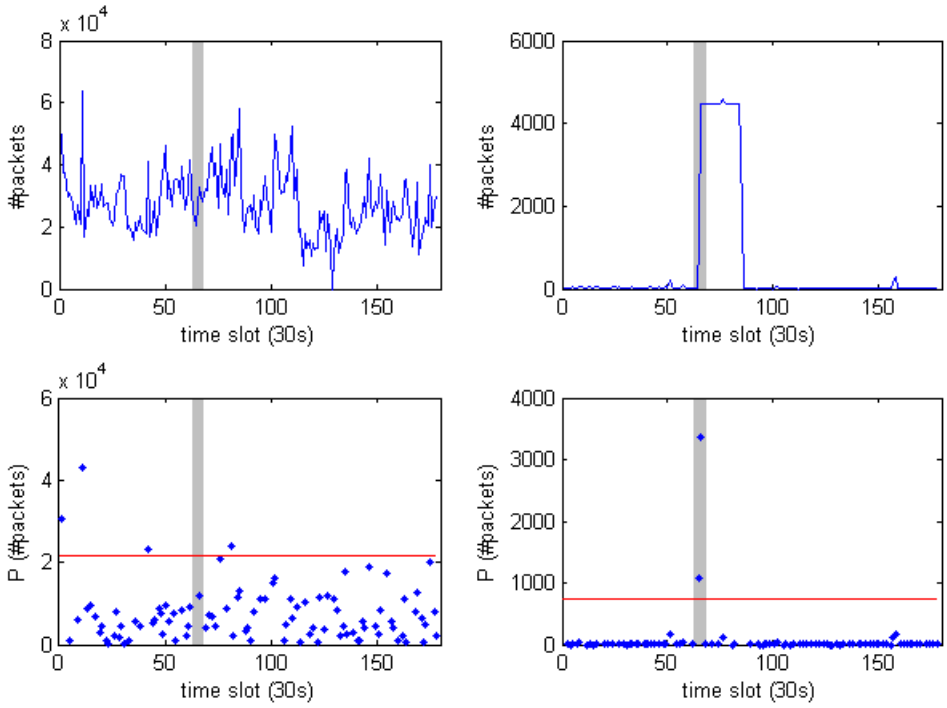

Figura 2 - Gráficos das séries de tempo do número de pacotes em dois níveis de agregação diferentes: nível 0 à esquerda, e nível 24 (para uma rede específica) à direita. A região destacada é o momento em que ocorreu uma anomalia.

Considerando todos os níveis de agregação ao mesmo tempo, podemos detectar qualquer anomalia que apareça em pelo menos um dos níveis. Mas este aumento de sensibilidade possui um preço considerável com relação a tempo de processamento, quantidade de memória necessária e a quantidade de falsos positivos gerados. Quando trabalhamos com dados históricos é possível termos uma quantidade enorme de dados para processar. Por exemplo, capturando 64 bytes por pacote em um link de 622 Mb/s (OC-12) que está sendo utilizado 100% teremos 42GB de dados em uma hora de captura (uma análise interessante dos requisitos de banda e disco para capturas de links de alta velocidade pode ser vista em [22]). Isto levanta uma restrição grave com relação ao tempo de processamento e a quantidade de memória necessária para a análise. O algoritmo deste trabalho foi implementado de forma a diminuir ao máximo o tempo de processamento (em

detrimento da quantidade de memória necessária), realizando apenas uma passagem por todo o arquivo de captura durante esta etapa. As conseqüências desta decisão são analisadas no subcapítulo 4.3.

Como com qualquer algoritmo de detecção de anomalias, um aumento na sensibilidade ocasiona naturalmente um aumento no número de falsos positivos (i.e. variações normais do tráfego são consideradas anomalias). Entretanto, utilizando esta variação da fórmula 1 que considera múltiplos níveis, surge um novo fenômeno que gera uma enorme quantidade de falsos positivos. Este fenômeno se deve a comunicações infreqüentes com determinadas redes. Por exemplo, se estamos buscando anomalias em um arquivo com uma hora de captura de pacotes de um determinado link com tráfego bidirecional, e usando uma granularidade de 30 segundos, temos 120 *slots* (ver seção 2.2.1) para análise. Para cada um destes 120 *slots*, os *Ps* correspondentes de cada rede e níveis de agregação presentes no arquivo serão analisados na busca de anomalias. Caso um destes *Ps* estiver acima de *K* desvios padrões da distribuição de *P*, a tupla (*slot*, rede, nível de agregação) será marcada como anômala. Quando a comunicação com uma determinada rede ocorre pouquíssimas vezes no arquivo de captura, usualmente em *slots* seqüenciais, esta comunicação tende a ser marcada como anômala no *slot* inicial, já que para todos os outros *slots* o valor de *p* será zero (i.e. seu desvio padrão será muito pequeno). Isto ocorre com freqüência no comportamento normal da Internet, com usuários acessando servidores externos por alguns minutos e não ocorrendo nenhuma outra comunicação com estes durante o resto da captura. Este fenômeno explica a enorme quantidade de anomalias encontradas nas análises da seção 4.3.

Apesar de estes problemas tornarem o algoritmo de detecção indesejável, a simplicidade do algoritmo torna fácil a descoberta dos pacotes responsáveis pela anomalia (ver seção 3.2), e, conseqüentemente, também a derivação dos atributos da anomalia. Assim foi possível concentrarmo-nos na classificação automatizada das anomalias, que é o objetivo principal deste trabalho. Além disto, o fenômeno explicado acima nos permitiu mostrar com clareza como a classificação automatizada pode ser utilizada para filtrar os falsos positivos, ao passo que mantém a detecção e classificação das anomalias de baixa intensidade (ver seção 3.4.3). Espera-se que o método apresentado neste trabalho para a classificação

automatizada possa ser aplicado a outros algoritmos de detecção de anomalias (e.g. [14]), e isto fica definido como trabalho futuro.

3.2 SEGUNDA ETAPA: IDENTIFICAÇÃO DOS PACOTES RESPONSÁVEIS

Analisando o trabalho de caracterização de anomalias do tráfego de rede feito previamente pela comunidade científica [2, 29, 31], podemos ver que diferentes tipos de anomalias podem afetar métricas de volume e *features* de tráfego, como endereços IPs e portas, da mesma maneira. Isto demonstra que não é possível classificar estas anomalias de forma confiável utilizando apenas esta informação, sendo necessário identificar outros dados relativos à anomalia. Neste trabalho, apresentamos esta informação adicional através de atributos da anomalia. Os atributos representam informações obtidas analisando os pacotes diretamente relacionados com a anomalia. Os atributos podem ser simples e diretamente obtidos (e.g. o impacto da anomalia em uma determinada métrica), ou podem ser obtidos através de uma derivação mais complexa (e.g. a proporção de origens por destinos responsáveis). O módulo de classificação utiliza assinaturas baseadas nos atributos derivados. Um processo de classificação confiável necessita de atributos significativos que proporcionem informação suficiente a respeito das anomalias para que se possa distinguir entre os diferentes tipos de anomalias. Por exemplo, o método de distinção entre ataques DDoS e *flash crowds* baseado em clusters desenvolvido por [19] poderia ser implementado e abstraído como um atributo para ajudar na classificação destes dois tipos.

O primeiro passo no processo de derivação destes atributos é encontrar os pacotes responsáveis pela anomalia. O algoritmo de detecção usado neste trabalho torna esta tarefa relativamente simples. Uma anomalia detectada na etapa anterior é identificada pela tupla (*slot*, rede, nível de agregação). Também temos a informação de quais métricas foram significativamente afetadas pela anomalia (i.e. para quais métricas o respectivo *p* foi considerado anômalo). Usando esta informação passamos a ler todos os pacotes no *slot* correspondente, que estão destinados para aquela rede naquele nível de agregação, a fim de encontrar os

fluxos responsáveis. Primeiramente, identificamos os endereços IPs de destino responsáveis pela anomalia (i.e. nível 32), num conceito de responsabilidade similar à noção de faixa de endereços IP e portas dominantes apresentada por [29]. Em nosso trabalho, o conjunto de destinos responsáveis é composto de todos os endereços IPs de destino que aparecem em qualquer um dos possíveis conjuntos mínimos de endereços que se retirados do tráfego deixariam os *deltoids* da anomalia abaixo de uma fração do valor limite original. De forma simplificada, encontramos os destinos que contribuíram significativamente ao impacto da anomalia. Após a identificação dos endereços IPs de destino responsáveis, os endereços IPs de origem, as portas de origem e destino e os protocolos dominantes são definidos com métodos equivalentes. Potencialmente, esta etapa do algoritmo pode ser realizada com qualquer algoritmo de detecção de anomalias que identifique o tempo de início e os fluxos anômalos das anomalias (e.g. [32, 14]).

3.3 TERCEIRA ETAPA: DERIVAÇÃO DE MÉTRICAS

Apesar de termos separado logicamente a derivação dos atributos como uma terceira etapa, alguns atributos são derivados na primeira e na segunda etapa. A tabela 1 mostra os atributos que são utilizados neste trabalho. Os atributos *found*, *impactlevel*, *duration* e *decrease* são derivados durante a etapa de detecção das anomalias (i.e. primeira etapa). O atributo *found* indica as métricas que foram consideradas anômalas, possuindo o valor correspondente do *deltoid* em caso positivo (i.e. considerado anômalo), e zero em caso negativo. O atributo *impactlevel* corresponde ao nível de impacto causado por uma anomalia. O nível de impacto é definido como a quantidade de níveis de maior agregação em que uma anomalia também foi encontrada devido às variações causadas por esta anomalia. O *impactlevel* é sempre calculado dos níveis de baixo (maior especificidade) aos de cima (maior agregação), já que os níveis mais específicos compõem os níveis gerais. Por exemplo, se a rede x.y.z/24 foi considerada anômala no *slot* w para a métrica número de pacotes, pegamos o valor do p correspondente e verificamos se a rede x.y/16 também teve uma anomalia detectada para esta métrica no *slot* w. Caso positivo, tiramos a proporção entre o p do nível 24 e o p do nível 16. Se o valor

for maior que 0.45, aumentamos o *impactlevel* da anomalia do nível 24. Este valor é então multiplicado à proporção entre o p dos dois próximos níveis (i.e. neste caso entre o 16 e o 8), e a mesma comparação é feita. Quando o valor resultante fica abaixo de 0.45, ou uma anomalia não foi encontrada, o algoritmo pára e o *impactlevel* está definido. Todas as constantes utilizadas na derivação dos atributos foram definidas empiricamente.

Nome do Atributo	**Tipo**	**Descrição**
found{p,b,s}	integer	Se a métrica correspondente for considerada anômala, o valor do P, caso contrário, zero.
impactlevel{p,b,s}	integer	O nível de impacto da anomalia.
duration{p,b,s}	integer	A quantidade de slots que a métrica permaneceu anômala.
decrease{p,b,s}	float	O maior deltoid negativo que ocorreu durante a anomalia em proporção ao valor limite.
#respdest	integer	O número de destinos responsáveis.
#rsrc/#rdst	integer	A proporção entre endereços IPs de origem e destinos responsáveis.
avg#rdstports	integer	O número médio de portas de destino responsáveis.
avg#rsrcports	integer	O número médio de portas de origem responsáveis.
#rpkt/#rdstport	integer	A proporção do número de pacotes por portas de destino responsáveis.
#rpkt/#rsrc	integer	O número médio de pacotes por origem responsável.
bpprop	integer	Tamanho médio dos pacotes da anomalia.
spprop	float	Proporção entre o número de pacotes SYN e o número total de pacotes da anomalia.
samesrcpred	boolean	Se existe um único endereço IP de origem responsável pela maioria dos destinos.
samesrcportspred	boolean	Se a maioria dos endereços IPs de origem utilizam as mesmas portas de origem.
oneportpred	boolean	Se apenas uma porta de destino é dominante.
invalidpred	boolean	Se a anomalia consiste principalmente de pacotes inválidos (e.g. cabeçalhos mal formados, tamanho insuficiente).
invprotopred	boolean	Se a anomalia é dominada por pacotes que usam números ou tipos de protocolo inválidos.
landpred	boolean	Se a maioria dos pacotes possui o mesmo valor para o endereço de origem e o endereço de destino.
echopred	boolean	Se a maioria dos pacotes é do tipo ICMP Echo, ICMP Echo Reply, ou os dois.
icmppred	boolean	Se a maioria dos pacotes são ICMP de outro tipo.
rstpred	boolean	Se a maioria dos pacotes são TCP com a flag RST habilitada.

Tabela 1 - Atributos derivados durante as primeiras fases do algoritmo. As letras p, b e s represetam pacotes, bytes e syn respectivamente

O atributo *duration* corresponde à duração da anomalia em *slots*. Neste trabalho, definimos que uma anomalia acaba quando a soma dos *deltoids* subseqüentes resulta em um valor abaixo de 90% do valor limite. O atributo *decrease* define a maior queda, após o início da anomalia, do *deltoid* correspondente em proporção ao valor limite. Estes atributos não foram utilizados na caracterização das anomalias feita neste trabalho, mas trabalhos prévios de caracterização (e.g. [19]) sugerem que podem ser úteis. Os demais atributos são obtidos direta ou indiretamente do processo de identificação dos fluxos responsáveis (i.e. segunda etapa). Os atributos *#respdest*, *#rsrc/#rdst*, *avg#rdstports*, *avg#rsrcports*, *#rpkt/#rdstport* e *#rpkt/#rsrc* são obtidos diretamente e não carecem maior explicação (ver sua descrição na tabela 1). Os atributos *bpprop* e *spprop* são calculados usando apenas os dados referentes às origens e destinos responsáveis, onde o primeiro é uma aproximação do tamanho médio dos pacotes responsáveis pela anomalia, e o segundo é a proporção entre o número de pacotes TCP com o flag SYN habilitado e o número total de pacotes da anomalia.

Os demais atributos são booleanos que expressam se uma determinada condição é predominante. No caso dos atributos *echopred* e *icmppred* a predominância é definida caso os pacotes dos tipos correspondentes foram os que mais contribuíram para a anomalia, e eles contribuíram no mínino o dobro do segundo tipo que mais contribuiu. Para os atributos *invalidpred*, *invprotopred* e *rstpred*, a contribuição de cada situação deve ser no mínimo 75% do total. O atributo *oneportpred* é obtido diretamente e definido quando apenas uma porta de destino é responsável. Os atributos *samesrcpred* e *samesrcportspred* utilizam um valor de 85% do valor total, significando, por exemplo, que para considerarmos que a mesma origem é dominante para todos os destinos (*samesrcpred*), a quantidade de vezes que ela aparece deve ser maior que 85% do total de origens responsáveis. O atributo *landpred* é um caso especial onde sabemos que a ocorrência de um pacote com os endereços IP de origem e destino iguais se deve a um ataque (ver seção 3.4.1), então basta que este fluxo (i.e. mesmo endereço de origem e destino) esteja na lista de responsáveis para que este atributo seja habilitado. Esta lista de atributos pode ser estendida com qualquer outro tipo de derivação que possa ser feita tendo os pacotes diretamente responsáveis pela anomalia.

3.4 QUARTA ETAPA: CLASSIFICAÇÃO

A contribuição principal deste trabalho esta na área de classificação automatizada das anomalias do tráfego de rede já detectadas. Esta classificação é importante para que os administradores de rede possam utilizar seu escasso tempo de forma eficiente ao ter uma gama maior de informações sobre cada anomalia, permitindo uma priorização adequada para análise manual. Neste contexto, é necessário que os administradores da rede possam facilmente entender e manipular o processo de classificação. Um processo de classificação automatizado e confiável (i.e. assinaturas com uma baixíssima taxa de falsos positivos) também pode ser utilizado em um sistema totalmente automatizado que reage às anomalias sem a necessidade de um operador humano, por exemplo, mudando a política de QoS dos roteadores. Desafortunadamente, este objetivo de classificação automatizada das anomalias do tráfego de rede com uma taxa baixa de classificações errôneas permanece um problema aberto e difícil. O grande número de tipos de anomalias [29] e as grandes variações de cada tipo individual (e.g. para DDoS [35]) tornam necessária a criação de assinaturas extremamente especializadas para se obter baixas taxas de falsos positivos. Neste trabalho utilizamos o esforço prévio de caracterização destas anomalias (ver seção 2.3) para dar um passo adiante na área de classificação automatizada, obtendo um grau menor de ambigüidade através do uso de uma quantidade maior de informações sobre cada anomalia.

Durante o curso deste trabalho identificamos três tipos de assinaturas para a classificação automatizada: assinaturas universais, assinaturas fortes e assinaturas locais. Assinaturas universais são regras que, caso satisfeitas, nunca erram a classificação da anomalia. Assinaturas deste tipo são raras, normalmente se apoiando em especificações de protocolo (exemplos serão dados mais adiante). Estas assinaturas não costumam ser afetadas pelas características da rede, podendo ser aplicadas universalmente. Por outro lado, as assinaturas fortes são aquelas que possuem uma baixa taxa de falsos positivos que pode variar (pouco) dependendo das características da rede. As assinaturas fortes normalmente possuem um ou mais valores limite para determinados atributos, e estes valores são difíceis de definir. A taxa de falsos positivos destas assinaturas deve ser baixa o

suficiente para que os falsos positivos não dominem o tempo de análise manual dos administradores de rede. Assinaturas locais são específicas para cada domínio administrativo, sendo definidas localmente pelos administradores de rede. Apesar dos administradores poderem manipular e modificar todo o processo de classificação (e.g. escolhendo quais regras ou valores limite utilizar), definimos algumas regras universais e fortes que podem ser utilizadas de forma global, em redes com características bem diferentes (suportamos esta afirmação com os resultados apresentados na seção 4.3). No restante deste subcapítulo, caracterizamos diferentes tipos de anomalias utilizando os atributos identificados e mostramos algumas das assinaturas que utilizamos para verificar o desempenho do algoritmo. Concentramo-nos principalmente na classificação dos ataques de negação de serviço distribuídos, mas abordamos também as anomalias causadas por varreduras de rede, varreduras de porta, *flash crowds*, *alpha flows* e por respostas a ataques. Algumas assinaturas locais e sua importância para os administradores de rede são discutidas no final.

3.4.1 ATAQUES DE NEGAÇÃO DE SERVIÇO

Os ataques de negação de serviço, ou DoS, são tentativas maliciosas de negar acesso aos recursos da rede. Estes recursos podem variar de largura de banda da rede a estruturas internas usadas por um servidor, e os ataques podem prevenir acesso a um serviço por usuários legítimos. Ataques DoS podem ser divididos amplamente em ataques lógicos e ataques de inundação [36]. Os ataques lógicos usam falhas de software para degradar o desempenho ou desabilitar completamente um serviço. Os ataques de inundação aprisionam os recursos utilizando ou requisitando mais do que a vítima suporta. Apesar de ambos os tipos serem importantes, estudamos apenas os ataques de inundação, por apenas estes terem a tendência de causar anomalias no tráfego de rede (ataques lógicos não causam impacto no tráfego de rede e no seu desempenho). Ataques de negação de serviço distribuídos (DDoS) são ataques DoS (de inundação) onde múltiplas origens são utilizadas para aumentar a capacidade de dano e dificultar as ações contra o ataque. Além de serem utilizados normalmente para atividades maliciosas (e.g. extorsão [40], guerras políticas virtuais [13]), os ataques DDoS podem reduzir

consideravelmente a qualidade de serviço de uma rede, mesmo quando esta possui recursos suficiente para lidar com o ataque [37].

A natureza agregada destes ataques torna a prevenção, ou até reação, muito difícil após terem chegado à vítima. Isto torna essencial que a detecção e prevenção deste tipo de ataque sejam feitas o mais próximo possível das origens. Para detectar estes ataques próximos às suas origens, é necessário detectar anomalias de baixa intensidade. Entretanto, anomalias causadas por ataques DoS de baixa intensidade e com um número pequeno de atacantes aproximam-se em comportamento ao tráfego normal e a outros tipos de anomalias, sendo necessário analisar extensivamente suas características para construir assinaturas fortes correspondentes. As anomalias de tráfego de rede causadas por ataques DDoS tem sido caracterizadas com variações acentuadas nas séries de tempo das métricas de volume número de pacotes, número de fluxos, ou em ambas [29, 2], e também afetando as distribuições estatísticas das *features* de tráfego endereços IP e portas de origem e destino [31]. Infelizmente, apenas estas características não são o suficiente para criar assinaturas robustas o suficiente para sua classificação automatizada, uma vez que diversas ambigüidades entre tipos diferentes permanecem presentes. Para diminuir consideravelmente esta ambigüidade utilizamos os estudos específicos sobre ataques DDoS e seus diferentes tipos [23, 35], junto com análises de como diferentes ferramentas utilizadas por atacantes funcionam (e.g. [15]).

Id	**Tipo**	**Assinatura**
1	DDoS Invalido	invalidpred ou invprotopred ou landpred
2	ICMP Echo DDoS	#respdest == 1 e echopred e (#rpkt/#rdstport > 30*Gr ou #rsrc/#rdest > 15)
3	TCP SYN DDoS	#respdest == 1 e founds e spprop > 0.9 e oneportpred e #rpkt/#rdstport > 10*Gr
4	Varredura de Rede	#respdest > 200 e samesrcpred
5	Varredura de Porta SYN	#respdest == 1 e #rsrc/#rdest == 1 e spprop > 0.8 e avg#rdstports > 5
6	Resposta a Ataque	#respdest == 1 e (rstpred ou icmppred) e foundp > 20*Gr e (não (impactlevelp == 3)) e (#rsrc/#rdest == 1 ou samesrcportspred)

Tabela 2 - Exemplos de assinaturas utilizadas neste trabalho. Apenas a primeira é uma assinatura universal, as demais são assinaturas fortes.

Assinaturas universais para anomalias DDoS podem ser definidas baseando-se na análise dos ataques DDoS que utilizam pacotes que não seguem as especificações dos protocolos utilizados. Por exemplo, muitos ataques DDoS de inundação têm sido vistos em redes de produção utilizando pacotes IP de tamanho mínimo (i.e. 40 bytes) [23], protocolos inválidos (e.g. protocolo IP 0 ou 255 [36, 23], ou pacotes com o mesmo endereço IP para origem e destino). Como estes tipos de pacote nunca são utilizados em uma comunicação legítima, assinaturas que identifiquem quando estas situações estão ocorrendo podem ser consideradas universais. Isto pode ser estendido para a camada de transporte, como ataques que utilizam uma combinação inválida de flags TCP ou tipos ICMP inválidos [23]. Identificar os pacotes que compõe a anomalia torna simples a derivação de atributos que podem ser usados para classificas estes ataques específicos. Por exemplo, os atributos *invalidpred*, *invprotopred* e *landpred* (ver tabela 1) foram definidos para cobrir os casos mencionados. Assim, uma regra universal pode ser definida como: *se invalidpred ou invprotopred ou landpred então classifique como DoS*. A tabela 2 mostra algumas das regras utilizadas neste trabalho, com a regra número 1 representando a assinatura descrita anteriormente. Todas as outras informações obtidas durantes as etapas anteriores (e.g. origens, destinos, protocolo, etc.) também são mostradas como parte do alerta. A tabela 3 mostra um exemplo de como estas informações podem ser mostradas ao usuário. A anomalia detectada foi classificada pela regra 1 por usar tanto pacotes inválidos (código de porta -4) quanto usar o mesmo endereço de origem e destino.

Criar assinaturas universais para anomalias causadas por ataques DDoS que utilizam pacotes que não possuem erros estruturais é extremamente difícil (senão impossível). Para este tipo de ataques, assinaturas fortes devem ser definidas utilizando uma grande variedade de atributos. A assinatura para ataques DDoS do tipo ICMP Echo mostrada na tabela 2 se caracteriza por uma anomalia possuir apenas um destino responsável (*#respdest*), sendo composta principalmente de pacotes ICMP Echo ou ICMP Echo Reply (*echopred*), e tendo ou uma taxa maior do que 30 pacotes por segundo (aproximação feita por *#rpkt/#rdstport*) ou mais que 15 origens responsáveis (*#rsrc/#rdst*). Estes valores limites foram definidos para detectar ataques de baixa intensidade, mas apresentaram uma baixa taxa de falsos

positivos (ver seção 4.3). A terceira assinatura da tabela 2 classifica ataques do tipo TCP SYN destinados a um serviço específico (*oneportpred*) com uma média de 10 ou mais pacotes por segundo. Esta assinatura também utiliza os atributos *founds* e *spprop* para verificar que a maioria dos pacotes responsáveis pela anomalia possui a flag SYN do TCP habilitada (e apenas ela). Outras assinaturas para anomalias DDoS de baixa intensidade utilizam as informações sobre a quantidade de endereços de origem (*#rsrc/#rdst*), quantidade de portas de origem (*avg#rsrcports*) e destino (*avg#rdstports*), ou combinações destes para identificar situações que sejam muito atípicas para o tráfego normal. Assinaturas para anomalias de alta intensidade possuem maior precisão por utilizarem os indicadores fortes de *impactlevel* (i.e. 3) para número de pacotes ou número de syn. O tamanho médio dos pacotes (*bpprop*) também pode ser utilizado como limitador, tendo em vista que é raro o uso de pacotes grandes em um ataque DDoS [36].

```
Network: 1.81.90.0

-----ANOMALY DETECTED-------
Type: ddos
Destination: 1.81.90.0
Slot: 11
----------------------------

Anomaly slot 11:
 Flag      Point     Deviation     Level     Duration    Decrease
 packet    583       552           1         6           0.336190
 bytes     19974     18954         1         6           0.338190
 syn       0         0             0         0           0.000000

#Responsible Destinations: 1

Destination IP: 1.81.90.192
 P Sources: 0   Ports: 0
 Number of Ports: 4
  Port -4: 459
  Port -6: 106
  Port -1: 17
  Port -7: 1

 #Responsible Sources: 1

  Source IP: 1.81.90.192
```

Number of Ports: 1 Port -4: 459 Total Number of Sources: 1

Tabela 3 - Classificação de uma anomalia como DDoS por uma assinatura universal, feita por uma implementação do algoritmo.

Algumas observações preliminares podem ser feitas com relação a este tipo de caracterização e às assinaturas correspondentes. Nosso algoritmo utiliza estes atributos para classificar as anomalias do tráfego de rede, e não para detectar comportamento anômalo. Isto torna os valores limites mais robustos com relação a mudanças nas características da rede. Por exemplo, a assinatura de ataques ICMP Echo poderia ser acionada pelas respostas ao gerenciamento de falhas feito por sistemas de gerenciamentos de rede (Network Management Systems, NMS), mas o modo de operação em intervalos regulares destes sistemas não seria detectado como anômalo.

3.4.2 OUTROS TIPOS DE ANOMALIAS

Os outros tipos de anomalias que caracterizamos neste trabalho são: varreduras de rede, varreduras de porta, *flash crowds*, *alpha flows* e respostas a ataques. As varreduras de rede, ou *network scans*, são tentativas de sondagem para identificar a disponibilidade de um serviço específico em muitas máquinas diferentes [35]. Este tipo de anomalia é importante porque são normalmente precursoras da subversão de máquinas vulneráveis por um atacante – que eventualmente as usará como origens para um ataque DDoS –, ou por worms. As varreduras de rede podem ser confiavelmente caracterizadas como uma única origem se comunicando com muitos destinos. Atacantes podem utilizar estratégias distribuídas de varreduras de rede, com várias origens se encarrega de uma porção da rede, ou ainda realizarem a varredura com pacotes bem espaçados no tempo. Estas estratégias mais elaboradas de varredura de rede são difíceis de serem detectadas e não as abordamos neste trabalho. Para o caso mais simples (e muito mais freqüente) de uma única origem se comunicando com diversas máquinas de uma rede, os atributos mais interessantes para caracterização são os que expressam o número de destinos responsáveis (*#respdest*) e se a anomalia possui uma origem dominante

(*samesrcpred*). Durante a análise realizada neste trabalho, identificamos que um valor limite de 20 para o atributo *#respdest* é eficiente e possui uma baixa taxa de falsos positivos para as redes testadas. Esta regra pode classificar erroneamente atualizações em massa de software (e.g. sistemas operacionais, antivírus) quando existem muitos clientes numa mesma rede atualizando ao mesmo tempo. Assinaturas mais fortes podem ser desenvolvidas utilizando os atributos *bpprop*, *foundsyn*, *spprop*, *oneportpred*, e *#rpkt/#rdstport* para identificar tipos mais específicos de varreduras de rede (e.g. varreduras que usam TCP SYN) e melhorar a precisão. A assinatura que utilizamos neste trabalho para classificar este tipo de anomalia pode ser vista na tabela 2. Escolhemos utilizar o valor limite de 200 para o número de destinos responsáveis pela enorme quantidade de anomalias encontradas para este tipo (ver seção 4.3).

As varreduras de portas são similares às varreduras de rede, mas se concentram em apenas um destino para descobrir quais serviços estão disponíveis nele. Estas varreduras costumam criar uma variação muito pequena no tráfego, mas alguns tipos podem ser detectados por utilizarem tipos específicos de pacotes, como os pacotes TCP SYN. Estas anomalias são caracterizadas por uma única origem (#rsrc/rdst), um único destino (*#respdest*) e múltiplas portas de destino (*avg#rdstports*) com poucos pacotes sendo utilizados (*#rpkt/#rdstport*). Para o tipo específico de varreduras de portas TCP SYN podemos utilizar os atributos *foundsyn* ou *spprop* como exemplificado pela assinatura 5 da tabela 2.

As anomalias respostas a ataques são causadas pelo tráfego gerados pelas vítimas de ataques (e.g. ataques DDoS, varreduras). Os pacotes que compõe estas anomalias são normalmente ou pacotes TCP com os flags RST ACK, RST ou SYN ACK definidos, ou pacotes ICMP de controle (e.g. Destination Unreachable, TTL Exceeded) [36]. As respostas a ataques costumam ter pacotes com um tamanho bem pequeno, já que não costumam portar dados (i.e. em alguns casos os cabeçalhos dos pacotes originais são enviados de volta como referência). As respostas a ataques DoS são caracterizadas por uma única origem responsável, enquanto que as respostas a varreduras de rede possuem muitas origens responsáveis. A linha entre uma anomalia causada por respostas a ataques e uma anomalia causada por um ataque DDoS de baixa intensidade é tênue,

especialmente pelo fato de existirem ataques DDoS, ditos refletores [23], que utilizam este comportamento de resposta de máquinas legitimas para atacar uma vítima. As assinaturas fortes para este tipo de anomalia são difíceis de definir e um cuidado maior deve ser tomado na priorização das anomalias classificadas pelas assinaturas correspondentes. Os atributos que caracterizam este tipo de ataque são: *#respdest*, *impactlevelpkts*, *bpprop*, *icmppred*, *rstpred*, *samesrcportspred* e *#rsrc/#rdst* e. A assinatura 6 da tabela 2 mostra uma assinatura unificada para classificar respostas a ataques de inundação e a tentativas de varredura. Neste trabalho decidimos considerar as anomalias de alta intensidade (i.e. as anomalias que possuem um impacto visível no nível 0) com estas características como ataques DDoS.

Os dois últimos tipos de anomalias que caracterizamos, *flash crowds* e *alpha flows*, não foram analisados empiricamente por não termos ocorrências suficientes registradas nos arquivos de captura documentados (ver seção 4.1). Flash *crowds* (FC) podem ser definidos como um evento onde ocorre um aumento súbito do número de requisições legítimas de clientes por um recurso (e.g. uma página web). Estes eventos podem ser esperados (e.g. transmissões de eventos pela web) ou inesperados (e.g. os acessos recebidos por um site pequeno porque foi referenciado por um site popular). A natureza distribuída das anomalias FCs torna-as muito parecidas às anomalias causadas por ataques DDoS [26]. Os atributos que podem ajudar a caracterizar este tipo de anomalia são: *#rsrc/#rdst*, *oneportpred*, *foundsyn*, *foundpkts*, *#rpkt/#rsrc* e *decrease*. As anomalias FCs não costumam serem detectadas em granularidades pequenas (como a que utilizamos neste trabalho, 30 segundos), necessitando o uso de granularidades maiores, como 5 minutos. *Alpha flows* são transferências com uma alta taxa de dados de uma única origem para um único destino, causando um forte impacto nas métricas número de bytes e número de pacotes [29]. Estes fluxos também tendem a possui um tamanho médio de pacote maior que os utilizados para ataques DoS (i.e. isto pode ser usado para diferenciar entre estes dois tipos). Normalmente os *alpha flows* são identificados pelas portas utilizadas que são características de operações de gerenciamento de rede que costumam gerá-los (e.g. backups agendados, medições de vazão de rede). Os atributos que identificamos para este tipo de anomalia são:

impactlevelbytes, *impactlevelpkts*, *#respdest*, *#rsrc/#rdst*, *bpprop* e *foundsyn*. Os administradores da rede podem definir regras utilizando portas específicas para filtrar operações conhecidas, o que caracterizamos como assinaturas locais.

3.4.3 ASSINATURAS LOCAIS

Tendo atributos entendíveis que representem informações significativas a respeito das anomalias, os administradores da rede podem definir suas próprias assinaturas baseadas em sua preciosa experiência. Estas regras podem não ser portáveis para outras redes, mas a flexibilidade de ser capaz de entender e manipular a maneira como as anomalias são classificadas é uma funcionalidade essencial para a aplicabilidade de sistemas de detecção e classificação automatizadas de anomalias de tráfego de rede em redes reais. As assinaturas fortes podem ser modificadas (i.e. mudando seus valores limites) ou totalmente desabilitadas, enquanto que assinaturas locais também podem ser definidas. Os administradores da rede podem preferir definir assinaturas para diferentes intensidades de anomalias ou utilizar suas próprias convenções para a marcação dos tipos das anomalias. Por exemplo, ao invés de tentar separar as anomalias respostas a ataques das anomalias DDoS que utilizam pacotes TCP RST, uma assinatura pode ser definida como *se #respdest == 1 e rstpred e impactlevelpackets > 2 então classifique como AnomaliaRSTForte*. Esta assinatura poderia ainda ser especializada pra dois casos, um com origem única e outro com múltiplas origens. Utilizando assinaturas customizadas como esta, os administradores da rede podem fugir das classificações de termos gerais e passar a agregar maior informação diretamente na identificação dada a cada anomalia.

A flexibilidade provida por este método de classificação também pode ser utilizada para reduzir a quantidade de falsos positivos dos algoritmos de detecção. O raciocínio é que uma grande variedade de assinaturas podem ser definidas para potencialmente cobrir a maioria das anomalias verdadeiras, e uma regra padrão – aplicada a todas as anomalias que não forem classificadas por uma regra anterior – poderia marcar as anomalias com um identificador que teria a menor prioridade possível para eventualmente serem descartadas. Isto reduz a taxa de detecção de anomalias verdadeiras, mas troca a taxa de falsos positivos dos algoritmos de

detecção pela taxa de falsos positivos do processo de classificação (que pode ser muito mais preciso). Por exemplo, a sensibilidade adicionada ao nosso algoritmo de detecção de anomalias pelo método multi-nível pode ser filtrada aplicando as diferentes assinaturas mencionadas neste trabalho, tendo uma última assinatura padrão que seleciona todas as anomalias com ao menos um atributo *impactlevel* igual a 3 (estas seriam todas as anomalias detectadas pelo algoritmo de detecção não modificado), e descartando todas as outras anomalias não classificadas por estas regras

4 AMBIENTE E RESULTADOS EXPERIMENTAIS

Neste capítulo apresentamos uma avaliação do desempenho do algoritmo desenvolvido neste trabalho para a classificação automatizada de anomalias de tráfego de rede. Validamos estatisticamente o algoritmo para anomalias DDoS e realizamos uma análise de resultados dos tipos de anomalia DDoS, varredura de rede, varredura de porta e respostas a ataques. A diversidade de ataques DDoS nos permite verificar a expressividade do nosso algoritmo. Se é possível distinguir diferentes tipos de anomalias DDoS de variações normais do tráfego e de outros tipos de anomalias, segue que a classificação automatizada deve ser possível para outros tipos de anomalias de tráfego de rede. Primeiro apresentamos os dados utilizados para a avaliação, seguidos da metodologia utilizada na validação estatística, e terminamos com a análise dos resultados.

4.1 DADOS

Uma validação estatística apropriada de algoritmos de detecção (e classificação) de anomalias requer o uso de dados históricos com anomalias conhecidas e bem documentadas. Estes dados podem ser coletados de uma rede real e depois catalogados por administradores de redes experientes. Desta forma se obtém um conjunto de dados com anomalias reais (i.e. anomalias que não foram sintetizadas para avaliação) conhecidas, mas que pode conter erros causados pelos administradores (e.g. uma classificação manual errônea). Este procedimento também não permite o controle sobre as características das anomalias, como variar sua intensidade, dificultando algumas formas de validação. Além disto, criar este tipo de conjunto de dados é caro, pois demanda muito esforço de administradores de redes experientes, e não existe algum que esteja publicamente disponível. A outra maneira de gerar um conjunto de dados com anomalias catalogadas é produzir artificialmente as anomalias em uma rede real ou simulada. Com este procedimento as anomalias podem ser completamente documentadas e não estão sujeitos a erros de interpretação. As características das anomalias também podem ser controladas (e.g. variar sua intensidade, duração, etc.), permitindo uma avaliação sobre diversas

situações. A desvantagem é que as anomalias podem não ser muito representativas das ocorrências de redes em produção. Nossa validação utiliza ambos os tipos de dados: os arquivos de captura do projeto METROSEC, que possui anomalias criadas artificialmente, e o conjunto de dados do repositório MAWI, com anomalias reais.

id	Tt	At	Tipo	#att	Intensidade	Vazão
tN	7200	600	UDP	3	22.9%	365 kb/s
tT	7200	600	UDP	4	86.8%	18,9 mb/s
1, tM	7200	600	UDP	2	7%	200 kb/s
2	3600	600	UDP	4	4%	388 kb/s
3	5400	600	UDP	4	7%	388 kb/s
4	7200	600	TCP SYN	2	12%	166 kb/s
5	5400	1800	ICMP Echo	4	8%	288 kb/s
6	3600	600	ICMP Echo	4	10%	388 kb/s
7	3600	600	Mixed	4	27%	250 kb/s
8	3600	600	Smurf	4	4%	250 kb/s
9	3600	600	TCP SYN	3	33%	252 kb/s
10	7200	660	UDP	4	92%	25,3 mb/s
11	10800	160 420 500	TCP SYN	1 1 1	90.5% 70.8% 45.62%	não documentado
12	1800	300	TCP RST	1	91.7%	não documentado

Tabela 4 - Características dos arquivos de captura selecionados do banco de dados do projeto METROSEC para a avaliação do algoritmo. Tt é a duração da captura em segundos, At é a duração do ataque em segundos, #att é o número de atacantes, Intensidade é referente ao número de pacotes por segundo (ataque/total) e a vazão é o volume de dados do ataque.

Os arquivos de captura do projeto METROSEC consistem de pacotes de tráfego real capturados na Rede Nacional de Ensino e Pesquisa da França (RENATER) com ataques simulados utilizando ferramentas de ataques DDoS reais. O conjunto de dados foi criado no contexto do projeto de pesquisa METROSEC [34] para, entre outros objetivos, estudar a natureza das anomalias de tráfego de rede e o seu impacto na qualidade de serviço das redes. Este conjunto de dados já foi utilizado na validação de diferentes estudos em detecção de anomalias (e.g. [39, 1, 19]). Os dados foram capturados utilizando placas de captura DAG, entre o final de 2004 e o final de 2006, e possuem anomalias que vão de intensidades muito baixas (i.e. menos de 4% do volume do tráfego agregado) a intensidades muito altas (i.e. mais de 80%). De um a quatro laboratórios franceses de pesquisa, localizados em

Mont-De-Marsan, Lyon, Nice e Paris, foram usados para gerar ataques DDoS realísticos e complexos. O destino destes ataques é outro laboratório de pesquisa francês localizado em Toulouse, o LAAS.

Os arquivos de captura deste projeto estão documentados com o tempo de começo e fim de captura, o tempo de começo e fim do ataque, a intensidade, tipo e número de atacantes do ataque. A tabela 4 mostra os arquivos de captura utilizados para a validação do nosso algoritmo. Concentramo-nos em ataques de baixa intensidade para diferentes tipos de ataques DDoS. Destes tipos, UDP, TCP SYN, TCP RST e ICMP Echo são ataques de inundação básicos que utilizam este tipo de pacotes, Mixed é uma combinação destes métodos, e Smurf é um ataque refletor que envia pacotes ICMP Echo Request para um endereço IP de broadcast apontando como origem o endereço da vítima. Os primeiros três ataques da tabela 4 foram feitos usando a ferramenta Trin00 [15] e os demais utilizando a ferramenta TFN2K [3].

Para complementar os dados (com anomalias simuladas) do METROSEC, utilizamos os dados disponibilizados pelo projeto MAWI. O conjunto de dados do projeto MAWI consiste em arquivos de captura de 15 minutos coletados diariamente às 2PM em uma rede japonesa chamada WIDE desde 1999 até o presente. Estes arquivos são disponibilizados publicamente após os pacotes serem anonimizados (i.e. os endereços IPs são mudados mantendo a consistência) e terem seus dados removidos (i.e. apenas os cabeçalhos até a camada de transporte são mantidos). Os autores de [14] começaram um esforço de detecção e documentação das anomalias encontradas neste conjunto de dados. Esta documentação foi feita a partir de uma seleção inicial de anomalias através do algoritmo de detecção desenvolvido pelos autores, e a análise manual destas anomalias. Este método pode resultar em anomalias não sendo detectadas (como mostraremos na análise dos resultados) ou em erros de classificação cometidos pelos autores. Dos arquivos de captura documentados por [14], selecionamos aleatoriamente um total de 30 arquivos obtidos entre 2001 a 2006, entre os quais alguns possuíam anomalias DDoS. Os arquivos de captura escolhidos pertencem a dois links chamados *samplepoint*-B e *samplepoint*-F, que representam links entre o Japão e os Estados Unidos. O tráfego destes links consiste principalmente em dados trocados entre

universidades japonesas e provedores de serviço comerciais, e uma grande diversidade de anomalias pode ser vista consistentemente durante os sete anos de dados [4]. Usamos este segundo conjunto de dados para verificar que nosso algoritmo (i.e. atributos e assinaturas utilizadas) não se restringe a uma única rede ou a ataques artificiais.

4.2 METODOLOGIA

Para mostrar os resultados da validação estatística feita nos arquivos de captura do METROSEC, utilizamos uma técnica chamada de *Receiver Operating Characteristic*, ou curva ROC. Usamos um gráfico ROC que pode ser definido como a representação gráfica da probabilidade de se detectar sinais verdadeiros (i.e. a taxa de verdadeiros positivos, ou TPR) em função da probabilidade de se gerar um alarme falso (i.e. a taxa de falsos positivos, ou FPR). Esta técnica tem sido usada extensivamente para avaliar sistemas de detecção de intrusão e de anomalias, tenso sido introduzida na área pelo famoso estudo do Grupo de Tecnologia de Sistemas da Informação do MIT, a 1998 DARPA *Intrusion Detection Evaluation*. Entretanto, alguns problemas foram levantados para uma análise significativa de sistemas de detecção de intrusão utilizando esta técnica [33]. Uma explicação cuidadosa da metodologia utilizada é necessária para que os resultados possam ser utilizados em análise subseqüente de outros estudos. Por esta razão, explicamos com detalhe a metodologia usada para validar nosso algoritmo nesta seção.

O sucesso do algoritmo desenvolvido neste trabalho depende da detecção da anomalia, da identificação dos pacotes responsáveis, da derivação dos atributos e das assinaturas usada para classificação. A validação feita considera todas estas etapas como um todo para se ter uma visão mais realista do desempenho do algoritmo em uma rede em produção. A validação estatística foi feita apenas para anomalias DDoS, já que estas são as únicas que temos disponíveis em um conjunto de arquivos de captura bem documentado. Este novo contexto, considerando outros fatores além da detecção, levanta algumas sutilezas para determinar uma unidade de análise [33] apropriada. A unidade de análise base de nosso algoritmo é qualquer anomalia detectada que entra o módulo de

classificação (i.e. as anomalias do último nível, ou nível 24 neste caso). Para uma anomalia ser passada ao módulo de classificação, ela deve ter sido detectada por nosso algoritmo de detecção (ver seção 3.1), que utiliza *deltoids* para encontrar comportamento anômalo. Isto significa que o algoritmo de detecção garante apenas a detecção do começo da anomalia, já que os *slots* subseqüentes podem não possuir grandes variações. Como a validação estatística será realizada apenas com anomalias DDoS, o processo de classificação também precisa ser reduzido a uma decisão binária com relação a apenas este tipo de anomalia.

Solucionamos estes problemas da seguinte maneira. Definimos uma anomalia de ataque como uma anomalia detectada por nosso algoritmo de detecção devido principalmente aos pacotes de um ataque. Estas anomalias foram todas identificadas e documentadas previamente por análise manual utilizando a documentação dos ataques. O algoritmo deve então identificar e classificar corretamente as anomalias de ataque para considerarmos como verdadeiro positivo. A identificação é considerada correta se o endereço IP de destino dito responsável for o da vítima e a lista de endereços IPs de origem tiver apenas endereços IPs dos pacotes de ataques, mas não necessariamente todos eles. Consideramos que uma anomalia deve ser detectada em até dois *slots* de tempo (i.e. 60 segundos) após o começo do ataque, principalmente devido a ataques que possuem um começo com aumento gradativo. Isto significa que, se uma anomalia não é detectada em 1 minuto do começo do ataque (documentado), um falso negativo será considerado. Por outro lado, se uma anomalia de ataque é detectada durante a duração do ataque, normalmente devido a uma nova variação de intensidade do ataque, ela deve ser corretamente classificada. A conversão binária do processo de classificação foi feita considerando duas classes: anomalias DDoS e todo o resto (variações de tráfego normal inclusive). Assim, anomalias detectadas que não foram causadas pelos ataques são consideradas apenas para falsos positivos (i.e. quando são classificadas como DDoS) e verdadeiros negativos (i.e. quando não é classificada como DDoS). Por exemplo, caso uma anomalia causada por uma varredura de rede seja classificada como DDoS, conta-se um falso positivo, caso esta mesma anomalia seja classificada como uma varredura de rede, varredura de porta, etc., conta-se um verdadeiro negativo.

Como os arquivos do projeto MAWI não foram totalmente documentados, não é possível utilizarmos a mesma metodologia neles. Para estes dados fazemos apenas uma análise dos resultados, sem utilizá-los como parte da validação estatística. Usando o trabalho de documentação de [14] temos uma primeira aproximação das anomalias que existem nestes arquivos, mas muitas outras permanecem escondidas e então não podemos determinar a taxa de verdadeiros positivos do algoritmo. Nossa análise foi realizada inspecionando cuidadosamente (de forma manual) os alertas gerados pelo algoritmo para descrever os falsos e verdadeiros positivos e qualquer falso negativo conhecido. Para ambos os conjuntos de dados (i.e. METROSEC e MAWI), realizamos também um estudo dos resultados para entender como as diferentes etapas do algoritmo afetam o desempenho. O mesmo conjunto de assinaturas foi utilizado na análise de ambos os conjuntos de dados. Os seguintes parâmetros foram utilizados para o algoritmo de detecção: granularidade de 30 segundos, níveis de agregação 0, 8, 16 e 24. Usamos diferentes valores de K para o algoritmo de detecção nos dados do METROSEC para ilustrar as sutilezas levantadas para a unidade de análise para nosso algoritmo, e usamos um K fixo de 2 (escolhido de uma curva ROC anterior [19]) para os dados do projeto MAWI e para a análise das assinaturas de outros tipos de anomalias.

4.3 RESULTADOS E DISCUSSÃO

Nem todo sistema de detecção de intrusão ou anomalia possui um parâmetro de entrada ou um resultado apropriado para se aplicar um valor limite deslizante a fim de que uma curva possa ser plotada no espaço ROC [33]. Para estes sistemas, plotar um único ponto no espaço ROC é mais apropriado, sem conectar este ponto por linhas às coordenadas (0,0) e (1,1). Nosso algoritmo se encaixa neste grupo por causa das sutilezas discutidas na seção anterior. A Figura 3 mostra o gráfico do espaço ROC com os resultados obtidos para a análise dos dados do METROSEC. Cada ponto do gráfico representa o resultado (i.e. a probabilidade de alarmes verdadeiros em função da probabilidade de alarmes falsos) obtido para cada valor de K testado. O gráfico mostra que, neste caso, aumentar este parâmetro pode na verdade aumentar a taxa de falsos positivos e

reduzi-lo pode fazer o contrário. Apesar de contra-intuitivo, isto pode ser explicado pela filtragem feita pelo módulo de classificação. Novas anomalias detectadas pelo aumento de sensibilidade, normalmente apenas variações normais do tráfego, dificilmente serão classificadas como DDoS, e ao mesmo tempo aumentam a quantidade total de anomalias que estão sendo testadas (i.e. o número de verdadeiros negativos).

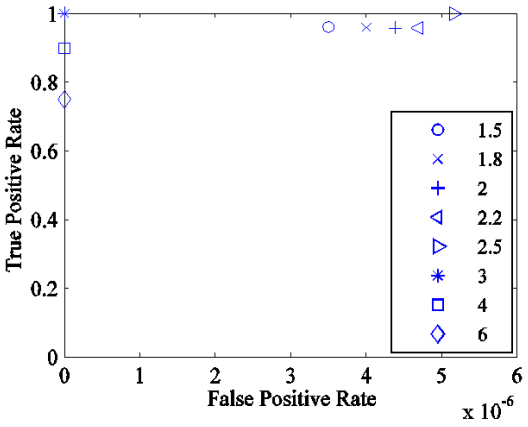

Figura 3 – Gráfico do espaço ROC com os resultados obtidos para os dados do METROSEC. A escala do eixo X é e-06.

K	TP	FP	FN	TN	TPR	FPR
1.5	25	2	1	571715	0.961	3.5e-06
1.8	24	2	1	500448	0.96	4.0e-06
2	23	2	1	455731	0.958	4.4e-06
2.2	23	2	1	425752	0.958	4.7e-06
2.5	23	2	0	387719	1	5.2e-06
3	22	0	0	346656	1	0
4	19	0	2	288446	0.9	0
6	12	0	4	192247	0.75	0

Tabela 5 - Resultados do algoritmo para anomalias DDoS nos dados do METROSEC.

Os resultados podem ser vistos mais claramente na Tabela 5. Apesar de termos apenas 16 ataques nos arquivos escolhidos do METROSEC, o número total de anomalias de ataques para os valores de K considerados chega a 26. Estas anomalias “extras” significam que alguns ataques sofreram uma variação positiva considerável em sua intensidade enquanto estava em andamento, e esta variação foi detectada como uma nova anomalia de ataque. Estes resultados mostram uma precisão surpreendente das assinaturas utilizadas para classificar anomalias DDoS. Considerando o pior resultado (i.e. K igual a 2.2), dois alarmes falsos foram emitidos e uma (1) anomalia de ataque foi considerada normal. Analisando manualmente estes alarmes, identificamos que um alarme falso era uma resposta a ataque de pacotes TCP RST ACK de alta intensidade em resposta ao ataque de inundação TCP SYN do arquivo de captura 11 (ver tabela 4). O outro alarme falso era uma

ataque DDoS real, não simulado pelos pesquisadores, que estava acontecendo na mesma hora da simulação. O falso negativo se deve ao ataque do arquivo 8 começar nos últimos 7 segundos de um *slot* e criar uma anomalia inicial de baixíssima intensidade seguida de uma anomalia com intensidade mais forte no próximo *slot*. A primeira anomalia não se enquadra em nenhuma de nossas assinaturas por ser muito fraca, mas a anomalia seguinte, que está dentro dos dois *slots* previstos, é classificada corretamente. Nenhuma nestas anomalias problemáticas são detectadas pelo algoritmo de detecção utilizando um K igual a 3, enquanto que todas as anomalias de ataque são detectadas. É claro que o desempenho perfeito destas assinaturas para um K igual a 3 é um artefato do conjunto de dados testado, e usar este algoritmo em redes de produção com anomalias reais eventualmente geraria alguns (mas poucos) falsos positivos (como mostramos em seguida na análise dos dados japoneses).

Avaliamos ainda a expressividade do nosso algoritmo para classificar outros tipos de anomalias utilizando as assinaturas não-DDoS da Tabela 2 nos dados do METROSEC. Com estas assinaturas foram encontrados 16 varreduras de porta, 13 respostas a ataques e 2471 varreduras de rede. A análise manual destes alarmes mostrou que todas as varreduras de porta eram verdadeiros positivos. Das 13 respostas a ataques, nove (9) eram respostas aos ataques DDoS documentados e uma (1) era resposta a uma varredura de rede. Não foi possível identificar a natureza das outras três anomalias, mesmo com a ajuda dos experientes administradores de rede do LAAS. Estas anomalias indefinidas são geradas por um súbito tráfego multi-origem ICMP, normalmente proveniente de um mesmo prefixo de rede, e se assimila a respostas a tentativas de varreduras de rede, mas são destinadas a endereços IPs inexistentes. É pouco provável que estas anomalias sejam de tentativas de varredura do tipo *idlescan* [8], considerando-se que elas são encontradas em vários meses diferentes de 2006 e é improvável que um atacante conseguisse sniffar a rede comutada do LAAS para ver as respostas. As varreduras de rede não foram analisadas manualmente devido ao seu grande número, mas a assinatura utilizada (ver tabela 2) possui uma taxa de classificação errônea extremamente baixa (senão inexistente).

Utilizar mais assinaturas ou baixar os valores limites das assinaturas da tabela 2 certamente revelaria novas anomalias verdadeiras, mas também poderia aumentar a taxa de falsos positivos do algoritmo. Por exemplo, se mudássemos o valor limite para *#rpkt/#rdstport* da assinatura ICMP Echo para > 15*Gr, os falsos negativos do arquivo 8 seriam corretamente classificados, mas o algoritmo classificaria 3 respostas a ataques como DDoS (i.e. teríamos em troca mais 3 falsos positivos). De maneira similar, se definíssemos o valor limite de *foundp* como > 10*Gr para a assinatura de resposta a ataque da tabela 2, encontraríamos uma (1) nova anomalia verdadeira de resposta a varredura de rede e quatro (4) novas anomalias indefinidas. Apesar destas não serem todas as anomalias que existem nestes arquivos, esta análise preliminar de assinaturas não-DDoS sugere fortemente que é possível realizar classificação automatizada confiável de diferentes tipos de anomalia de tráfego de rede.

Para termos certeza que as assinaturas utilizadas não estavam especializadas nas características da rede e das anomalias dos dados do METROSEC, avaliamos as mesmas assinaturas no conjunto de dados do projeto MAWI. Em relação às anomalias DDoS, os testes resultaram em um total de 19 verdadeiros positivos, 3 falsos positivos e 9 falsos negativos conhecidos, em um total de mais de 2.5 milhões de anomalias detectadas. Dos 19 verdadeiros positivos, 6 não haviam sido identificados previamente pelo trabalho de [14] e consistiam principalmente de pacotes IP de tamanho mínimo (i.e. 34 bytes, sem cabeçalhos da camada de transporte) e pacotes com o mesma origem e destino. Os 3 falsos positivos foram classificados pela assinatura de ICMP Echo, eram de um único arquivo de captura e do mesmo endereço IP de origem. Como são os endereços de destino (cada falso positivo é para um destino diferente) que enviam pacotes ICMP Echo e a anomalia está sendo detectada pelos pacotes ICMP Echo Reply de resposta, não está claro as anomalias são de tráfego normal ou se está ocorrendo um ataque refletor. A quantidade de falsos negativos expõe uma limitação conhecida do algoritmo de detecção utilizado neste trabalho. Todos estes falsos positivos estão relacionados com anomalias que já haviam começado antes do começo do processo de captura e não tiveram grandes variações de intensidade enquanto ocorriam ou durante os 15 minutos de captura. Nosso algoritmo de detecção falha em detectar

estas anomalias por operar apenas com *deltoids*. Isto não apresenta uma limitação direta do processo de classificação ou das assinaturas, uma vez que a análise manual mostrou que estas anomalias seriam corretamente classificadas caso detectadas. Com relação aos outros tipos de anomalias, o algoritmo encontrou 4429 varreduras de rede, 5233 varreduras de porta e 72 respostas a ataques. Uma análise manual preliminar destes alarmes mostrou que uma grande parte destas anomalias é causada por atividades de worms (e respostas a estas), com variações das worms Sasser e Dabber sendo particularmente freqüentes. Uma análise manual completa destes alarmes fica como trabalho futuro.

5 CONCLUSOES E TRABALHOS FUTUROS

Com a crescente dependência em redes de computadores cada vez mais complexas, é essencial garantir o correto funcionamento das redes com um nível mínimo de qualidade de serviço. Anomalias do tráfego de rede podem causar graves problemas no desempenho da rede, e devem ser identificadas e eliminadas pelos administradores de rede o mais rápido possível. Este trabalho apresentou um algoritmo para realizar a detecção e classificação automatizada das anomalias do tráfego de rede. A contribuição real do trabalho foi feita especificamente na área de classificação automatizada de anomalias, que ainda está dando seus primeiros passos na comunidade acadêmica. Utilizando um algoritmo de detecção conhecido, com uma pequena alteração para aumentar sua sensibilidade a anomalias de baixa intensidade, apresentamos um novo método para a classificação automatizada de anomalias de tráfego de rede. Foi mostrado como a capacidade de identificar os fluxos anômalos dos algoritmos de detecção do estado da arte pode ser utilizada para se obter um conjunto de informações mais completo sobre as anomalias de forma a permitir uma classificação automatizada confiável. Esta informação é representada por uma variedade de atributos da anomalia (i.e. métricas que representam algum tipo de informação), e estes atributos são utilizados em um módulo de classificação baseado em assinaturas. O método apresentado ainda permite o total controle sobre o processo de classificação por parte dos administradores de rede.

Um conjunto inicial de atributos de anomalias foi definido e utilizado para caracterizar diferentes tipos de anomalias, como DDoS e varreduras de rede. Também foi mostrado como um processo de classificação automatizada pode ser utilizado como um filtro para reduzir enormemente a quantidade de falsos positivos de algoritmos de detecção de anomalias. Algumas assinaturas foram definidas a partir do esforço de caracterização para classificar uma diversidade de tipos de anomalias. O algoritmo e as assinaturas foram avaliados usando dois conjuntos de arquivos de captura de pacotes de redes reais com diferentes anomalias. Os resultados obtidos mostram a expressividade do nosso método de conseguir diferenciar entre diferentes tipos de anomalias (incluindo variações normais do

tráfego). Apesar de termos nos concentrado nas importantes anomalias DDoS, uma análise preliminar de outros tipos de anomalias indica que a classificação automatizada geral (i.e. de muitos tipos ao mesmo tempo) pode ser feita de maneira confiável. Os seguintes temas para trabalhos futuros foram identificados:

- Identificar novos atributos e assinaturas que possam ser utilizadas de maneira confiável para classificar os tipos de anomalias considerados neste trabalho e outros.
- Fazer uma análise detalhada da viabilidade de se utilizar este tipo de algoritmo em tempo real em redes de produção.
- Testar outros algoritmos de detecção de anomalias mais complexos e eficientes com o método de classificação apresentado.

6 REFERÊNCIAS

[1] ABRY, P.; BORGNAT, P.; DEWAELE, G. Invited talk: Sketch based anomaly detection, identification and performance evaluation. In: *SAINT-W '07: Proceedings of the 2007 International Symposium on Applications and the Internet Workshops*. Washington, DC, USA: IEEE Computer Society, 2007.

[2] BARFORD, P. et al. A signal analysis of network traffic anomalies. In: *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement*. New York, NY, USA: ACM, 2002.

[3] BARLOW, J.; THROWER, W. *TFN2K - An Analysis*. Disponível em: <http://packetstormsecurity.org/distributed/TFN2k%5FAnalysis-1.3.txt>. Acesso em: 01 Dezembro 2008.

[4] BORGNAT, P. et al. Seven years and one day: Sketching the evolution of internet traffic. HAL - CCSD, 2008. Disponível em: <http://prunel.ccsd.cnrs.fr/ensl-00290756/en/>. Acesso em: 01 Dezembro 2008.

[5] BRUTLAG, J. D. Aberrant behavior detection in time series for network monitoring. In: *LISA '00: Proceedings of the 14th USENIX conference on System administration*. Berkeley, CA, USA: USENIX Association, 2000.

[6] CHEN, T. M. Increasing the observability of internet behavior. *Commun. ACM*, ACM, New York, NY, USA, v. 44, n. 1, pp. 93-98, 2001.

[7] CHHABRA, P. et al. Distributed spatial anomaly detection. *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pp. 1705-1713, abril, 2008.

[8] CHMIELARSKI, T. *Intrusion Detection FAQ: Reconnaissance Techniques using Spoofed IP Addresses*. The SANS Institute, abril, 2001. Disponível em: <http://www.sans.org/resources/idfaq/spoofed%5Fip.php>. Acesso em: 01 Dezembro 2008.

[9] CHO, K.; MITSUYA, K.; KATO, A. Traffic data repository at the wide project. In: *ATEC '00: Proceedings of the annual conference on USENIX Annual Technical Conference*. Berkeley, CA, USA: USENIX Association, 2000.

[10] CISCO IOS NetFlow. Cisco Systems. Disponível em: <www.cisco.com/web/go/netflow>. Acesso em: 01 Dezembro 2008.

[11] CLAFFY, K.; BRAUN, H.-W.; POLYZOS, G. A parameterizable methodology for internet traffic flow profiling. *Selected Areas in Communications, IEEE Journal on*, v. 13, n. 8, pp. 1481-1494, out, 1995.

[12] CORMODE, G.; MUTHUKRISHNAN, S. What's new: finding significant differences in network data streams. *IEEE/ACM Trans. Netw.*, IEEE Press, Piscataway, NJ, USA, v. 13, n. 6, pp. 1219-1232, 2005.

[13] DANCHEV, D. *Coordinated Russia vs Georgia cyber attack in progress*. ZDNet, ago, 2008. Disponível em: <http://blogs.zdnet.com/security/?p=1670>. Acesso em: 01 Dezembro 2008.

[14] DEWAELE, G. et al. Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedures. In: *LSAD '07: Proceedings of the 2007 workshop on Large scale attack defense*. New York, NY, USA: ACM, 2007. pp. 145-152.

[15] DITTRICH, D. *The DoS Project's "trinoo" distributed denial of service attack tool*. University of Washington, out, 1999. Disponível em: <http://staff.washington.edu/dittrich/misc/trinoo.analysis>. Acesso em: 01 Dezembro 2008.

[16] ERRAMILLI, A.; NARAYAN, O.; WILLINGER, W. Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Trans. Netw.*, IEEE Press, Piscataway, NJ, USA, v. 4, n. 2, pp. 209-223, 1996.

[17] ESTAN, C. et al. Building a better netflow. *SIGCOMM Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 34, n. 4, pp. 245-256, 2004.

[18] ESTAN, C.; SAVAGE, S.; VARGHESE, G. Automatically inferring patterns of resource consumption in network traffic. In: *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM, 2003. pp. 137-148.

[19] FARRAPOSO, S.; OWEZARSKI, P.; MONTEIRO, E. Detection, classification et identification d'anomalies de trafic. In: *Colloque Francophone d'Ingenierie des Protocoles (CFIP)*. Les Arcs, France: [s.n.], 2007.

[20] FARRAPOSO, S.; OWEZARSKI, P.; MONTEIRO, E. A multi-scale tomographic algorithm for detecting and classifying traffic anomalies. *Communications, 2007. ICC '07. IEEE International Conference on*, pp. 363-370, June 2007.

[21] FELDMANN, A.; GILBERT, A. C.; WILLINGER, W. Data networks as cascades: investigating the multifractal nature of internet wan traffic. In: *SIGCOMM '98: Proceedings of the ACM SIGCOMM '98 conference on Applications, technologies, architectures, and protocols for computer communication*. New York, NY, USA: ACM, 1998. pp. 42-55.

[22] FRALEIGH, C. et al. Design and deployment of a passive monitoring infrastructure. In: *IWDC '01: Proceedings of the Thyrrhenian International Workshop on Digital Communications*. London, UK: Springer-Verlag, 2001. pp. 556-575.

[23] HUSSAIN, A.; HEIDEMANN, J.; PAPADOPOULOS, C. A framework for classifying denial of service attacks. In: *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM, 2003. pp. 99-110.

[24] Internetworking Technology Handbook: Network Management Basics. Cisco Systems. Disponível em: <http://www.cisco.com/en/US/docs/internetworking/technology/handbook/NM-Basics.html>. Acesso em: 01 Dezembro 2008.

[25] ITU-T Recommendation M.3400: TMN Management Functions. [S.l.]: International Telecommunication Union (ITU), feb, 2000. Disponível em: < http://www.itu.int/rec/T-REC-M.3400-200002-I/en>. Acesso em: 01 Dezembro 2008.

[26] JUNG, J.; KRISHNAMURTHY, B.; RABINOVICH, M. Flash crowds and denial of service attacks: characterization and implications for cdns and web sites. In: *WWW '02: Proceedings of the 11th international conference on World Wide Web*. New York, NY, USA: ACM, 2002. pp. 293-304.

[27] KIM, M.-S. et al. A flow-based method for abnormal network traffic detection. *Network Operations and Management Symposium, 2004. NOMS 2004. IEEE/IFIP*, v. 1, pp. 599-612 Vol.1, April 2004.

[28] KRISHNAMURTHY, B. et al. Sketch-based change detection: methods, evaluation, and applications. In: *IMC '03: Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2003. pp. 234-247.

[29] LAKHINA, A.; CROVELLA, M.; DIOT, C. Characterization of network-wide anomalies in traffic flows. In: *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2004. pp. 201-206.

[30] LAKHINA, A.; CROVELLA, M.; DIOT, C. Diagnosing network-wide traffic anomalies. In: *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM, 2004. pp. 219-230.

[31] LAKHINA, A.; CROVELLA, M.; DIOT, C. Mining anomalies using traffic feature distributions. In: *SIGCOMM '05: Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM, 2005. pp. 217-228.

[32] LI, X. et al. Detection and identification of network anomalies using sketch subspaces. In: *IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2006. pp. 147-152.

[33] MCHUGH, J. Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory. *ACM Trans. Inf. Syst. Secur.*, ACM, New York, NY, USA, v. 3, n. 4, pp. 262-294, 2000.

[34] METROSEC. Disponível em: <http://www.laas.fr/METROSEC>. Acesso em: 01 Dezembro 2008.

[35] MIRKOVIC, J.; REIHER, P. A taxonomy of ddos attack and ddos defense mechanisms. *SIGCOMM Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 34, n. 2, pp. 39-53, 2004.

[36] MOORE, D. et al. Inferring internet denial-of-service activity. *ACM Trans. Comput. Syst.*, ACM, New York, NY, USA, v. 24, n. 2, pp. 115-139, 2006.

[37] OWEZARSKI, P. On the impact of dos attacks on internet traffic characteristics and qos. *Computer Communications and Networks, 2005. ICCCN 2005. Proceedings. 14th International Conference on*, pp. 269-274, out, 2005.

[38] PARK, K.; WILLINGER, W. *Self-Similar Network Traffic and Performance Evaluation*. New York, NY, USA: John Wiley & Sons, Inc., 2000.

[39] SCHERRER, A. et al. Non-gaussian and long memory statistical characterizations for internet traffic with anomalies. *IEEE Trans. Dependable Secur. Comput.*, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 4, n. 1, pp. 56-70, 2007.

[40] SWARTZ, J.; ACOHIDO, B. *Botnets can be used to black-mail targeted sites*. USA TODAY, mar, 2008. Disponível em: <http://www.usatoday.com/tech/news/computersecurity/2008-03-16-bot-side%5FN.htm>. Acesso em: 01 Dezembro 2008.

[41] ZHANG, Y. et al. Online identification of hierarchical heavy hitters: algorithms, evaluation, and applications. In: *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2004. pp. 101-114.