Universidade Federal do Paraná

Mateus Pelloso

Sistema de Predição de Ataques de Negação de Serviço Distribuídos

Curitiba PR
2018

Mateus Pelloso

Sistema de Predição de Ataques de Negação de Serviço Distribuídos

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática, no Programa de Pós-Graduação em Informática, setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Profa. Dra. Michele Nogueira Lima.

Curitiba PR
2018

Catalogação na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

P392s Pelloso, Mateus
Sistema de predição de ataques de negação de serviço distribuído / Mateus Pelloso. – Curitiba, 2018.
53 p. : il. color. ; 30 cm.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Informática, 2018.

Orientador: Michele Nogueira Lima .
Bibliografia: p. 51-53.

1. Redes de computadores – Medidas de segurança. 2. Ataques de negação de serviço. 3. Ataques DDoS. I. Universidade Federal do Paraná. II. Lima, Michele Nogueira. III. Título.

CDD: 005.8

Bibliotecário: Elias Barbosa da Silva CRB-9/1894

MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO INFORMÁTICA

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em INFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **MATEUS PELLOSO** intitulada: **Sistema de Predição de Ataques de Negacao de Servico Distribuído**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 23 de Março de 2018.

MICHELE NOGUEIRA LIMA
Presidente da Banca Examinadora (UFPR)

WAGNER HUGO BONAT
Avaliador Externo (UFPR)

ALEX BORGES VIEIRA
Avaliador Externo (UFJF)

MINISTÉRIO DA EDUCAÇÃO - UNIVERSIDADE FEDERAL DO PARANÁ - Setor de Ciências Exatas - Coordenação de Pós-Graduação em Informática

Aos meus pais, Luzia e Ismael, que me permitiram existir.

Agradecimentos

À Profa. Dra. Michele Nogueira Lima sou grato, não apenas pelo acompanhamento acurado no percurso de desenvolvimento da pesquisa, orientando-me com precisão e competência indiscutíveis, mas principalmente pelo apoio, estímulo e paciência dispensados no decorrer de todo o curso de Mestrado em Informática da Universidade Federal do Paraná (UFPR).

Ao Prof. Dr. Aldri Luiz dos Santos pelos ensinamentos e estímulos dispensados durante todo o percurso do curso de mestrado enquanto coordenador do Núcleo de Redes Sem Fio e Redes Avançadas (NR2) da UFPR. E pelos conteúdos ministrados durante sua disciplina Tópicos em Redes de Computadores.

Aos Professores Dr. Bruno Bogaz Zarpelão (UEL) e Dr. Wagner Hugo Bonat (UFPR), membros da banca de qualificação, pelas críticas e sugestões que tanto contribuíram para o amadurecimento e avanço do trabalho.

Aos Professores Dr. Alex Vieira (UFJF) e Wagner Hugo Bonat (UFPR), membros da banca de defesa, pelas críticas, sugestões e discussão promovida durante a banca que contribuíram imensamente para avanços significativos da versão final do trabalho, bem como o estímulo para a continuidade e evolução da pesquisa como um todo.

Aos meus pais, Luzia de Melo Pelloso e Ismael Pelloso, e meu irmão, Fernando Pelloso, por ouvir todos os desabafos em momentos de cansaço e tensão resultantes do árduo trabalho realizado e dedicação dispensada.

À Nanachara Carolina Sperb, minha namorada, pelo estímulo para participar do programa de mestrado da UFPR, pelas intermináveis conversas sobre o andamento da pesquisa e do trabalho e, mais importante, pela compreensão ao abrir mão de momentos de descanso e lazer enquanto estive comprometido com o mestrado.

Aos amigos do NR2, Adi Nascimento Marcondes, Agnaldo de Souza Batista, Andressa Vergütz, Arthur Emilio Garcete Ferreira, Benevid Felix Silva, Bruno Marquez Cremonezi, Cainã Passos, Carlos Alberto Pedroso Junior, Danilo Rodrigo Possati, Diego Milhomem Schmitt, Euclides Peres Farias Junior, Gustavo Henrique Carvalho de Oliveira, Igor Steuck Lopes, Ligia Francielle Borges, Marcos Antônio Dellazari, Nelson Gonçalves Prates Junior, Paulo Lenz Junior, Rafael Araújo da Silva e Ricardo Tombesi Macedo pela convivência, pelas argumentações e discussões técnicas e científicas e também pelo apoio durante o andamento das disciplinas e desenvolvimento da pesquisa.

Aos amigos do Instituto Federal Catarinense Prof. Msc. Jolcemar Ferro, Profa. Msc. Maribel Barbosa da Cunha e em particular o Prof. Dr. Eduardo Silva, que tive a grata surpresa de descobrir, após o início do mestrado, que também fora membro do NR2 onde cursou seu mestrado e doutorado.

A Deus, pela oportunidade e discernimento.

Resumo

Os ataques de Negação de Serviço Distribuídos (*Distributed Denial of Service* - DDoS) crescem significativamente em volume, sofisticação e impacto. Exemplos são os ataques DDoS contra a empresa francesa OVN e o provedor de nomes DYN, os quais atingiram volumes inéditos de tráfego malicioso. Em geral, esses ataques são detectados ou mitigados apenas quando se encontram em estágios avançados. Até então, estudos apresentam abordagens e técnicas focadas principalmente na detecção e mitigação desses ataques. Recentemente surgiram pesquisas que expõem artefatos focados na predição de ataques DDoS por meio de redes neurais atuando na predição da matriz de tráfego; ou através de ferramentas estatísticas, como exemplo, Markov que predizem as etapas de um ataque, ou ainda avaliam a estabilidade de séries temporais aplicando ARIMA, entre outras. Tais abordagens requerem o treinamento prévio das redes neurais ou dos respectivos algoritmos, assim sendo, exigem histórico de ataques DDoS no fluxo da rede ou assinaturas dos ataques. Desta forma, as abordagens expostas ficam limitadas a ataques previamente conhecidos. Em geral, a sobrecarga da vítima consequente de um ataque DDoS ocorre em um intervalo de tempo muito curto (*milisegundos*). Assim, quando as técnicas propostas pelas abordagens anteriores conseguem identificar a aproximação de um ataque na rede, a sobrecarga já está em andamento e portanto muito próxima, resultando na indisponibilidade dos serviços. Diferente de outros trabalhos, este estudo defende o prognóstico precoce de ataques DDoS a fim de evitar custos e perdas provenientes do ataque. Este trabalho apresenta o STARK, um sistema autoadaptativo de predição de ataques DDoS, que identifica indícios do ataque na rede antes deste alcançar estágios avançados. Com base na teoria da metaestabilidade, o sistema STARK provê um aprendizado estatístico não supervisionado e identifica a iminência de ataques DDoS. Isso implica em dizer que o sistema não necessita de conhecimento prévio das características do fluxo da rede, de assinaturas do ataques DDoS ou de treinamento prévio de algoritmos para que possa predizer o ataque. A avaliação do sistema STARK segue uma abordagem orientada a traços, em que três bases de dados são utilizadas. Dessa forma, são extraídas características dessas bases de dados com a finalidade de submeter aos indicadores estatísticos e assim avaliar a tendência de comportamento dos dados. De acordo com a tendência exposta é possível identificar a aproximação de uma transição crítica, neste caso a iminência de um ataque DDoS. Nas avaliações realizadas, com os diferentes traços, o sistema STARK demonstra capacidade de predizer os respectivos ataques DDoS com minutos ou horas de antecedência.

Palavras-chave: Ataques, DDoS, Metaestabilidade, Predição, Segurança de redes.

Abstract

Distributed Denial of Service (DDoS) attacks grow significantly in volume, sophistication and impact. Examples are the DDoS attacks against the OVN French company and the DYN name provider which reached unprecedented volumes of malicious traffic in 2016. In general, these attacks are detected or mitigated only when they are in advanced stages. So far the studies show approaches and techniques focused mainly on the detection and mitigation of these attacks. Recently researches have emerged which expose artifacts focused on the DDoS attack prediction by means of neural networks acting on the traffic matrix prediction; or through statistical tools, for example, Markov's which predict the steps of an attack, or still evaluate the stability of temporal series applying ARIMA, among others. Such approaches require the prior training of neural networks or the respective algorithms. Hence, they demand DDoS attack history in the network flow or attack subscriptions. The exposed approaches are limited to previously known attacks. In general, the overload of a DDoS attack victim occurs in a very short interval time (milliseconds). Thus, when the proposed techniques by the previous approaches can identify the closeness of an attack in the network, the overload is already in progress and so very close, resulting in the service unavailability. Different from other works, this study defends the early prognosis of DDoS attacks in order to avoid costs and losses from the attack. This work presents STARK, a self-adaptable system for DDoS attack prediction, which identifies attack evidence in the network before it reaches advanced stages. Based on the metastability theory, the STARK system provides an unsupervised statistical learning and identifies the DDoS attack imminence. That implies saying the system does not need prior knowledge of the network flow, of DDoS attack subscriptions or prior training of algorithms to predict the attack. The STARK system evaluation follows trace-driven approach in which three datasets are used. Hence, features are extracted from these datasets in order to submit to the statistical indicators and evaluate the data behavior trends. According to the trends in the dataset behavior is possible to identify the closeness of a critical transition, in this case the DDoS attack imminence. On the carried out evaluations, with different traces, the STARK system shows capability of predicting the respective DDoS attacks in minutes or hours in advance.

Keywords: Attacks, DDoS, Metastability, Prediction, Network Security.

Lista de Figuras

Lista de Tabelas

Lista de Acrônimos

ARIMA	*AutoRegressive Integrated Moving Average*
CAIDA	Centro para Análise de Dados Aplicada à Internet
CDC	*Centers for Disease Control and Prevention*
CERT.br	Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil
CPU	*Central Process Unit*
CTU	Universidade Técnica Checa
DARPA	Agência de Projetos de Pesquisa Avançada de Defesa
DDoS	*Distributed Denial of Service*
EWS	*Early Warning Signal*
EWSys	*Early Warning System*
HTTP	*HyperText Transfer Protocol*
ICMP	*Internet Control Message Protocol*
IDS	*Intrusion Detection System*
IP	*Internet Protocol*
IPS	*Intrusion Prevention System*
IRC	*Internet Relay Chat*
LSTM	*Long Short-Term Memory*
P2P	*Peer to Peer*
RNN	*Recurrent Neural Networks*
TCP	*Transmission Control Protocol*
UDP	*User Datagram Protocol*

Lista de Símbolos

${}_1\gamma$	assimetria
ρ_1	autocorrelação
σ^2	variância
Σ	somatório
μ	média dos elementos do conjunto
SD	desvio padrão
t	tempo
z	elemento do conjunto de dados
n	número de elementos do conjunto de dados
CV	coeficiente de variação

Sumário

1 Introdução

Os ataques de negação de serviço distribuídos (*Distributed Denial of Service – DDoS*) são uma ameaça de segurança que comprometem a rede e os serviços na Internet. Exemplos são aqueles contra o serviço *web* e contra o serviço de nomes ocorridos em 2016 Woolf (2016). Esses ataques têm avançado em quantidade, volume e técnicas. No Brasil, o CERT.br[1] ressalta um aumento de 138% na quantidade de ataques DDoS em 2016 NicBR (2017). No geral, eles geram volumes de dados inesperados, chegando a *Terabits*. A fim de sobrecarregar os servidores ou enlaces da rede, os atacantes empregam técnicas cada vez mais sofisticadas Woolf (2016). Além disso, eles exploram os recursos disponíveis nos sistemas computacionais, largura de banda e diversidade resultante da distribuição geográfica dos dispositivos. Outro aspecto é a abrangência das redes de dispositivos infectados (*botnets*), uma vez que podem ser compostas por dispositivos móveis com vulnerabilidades exploradas Zargar et al. (2013).

De acordo com Akamai (2017), no primeiro semestre de 2017, o Estados Unidos da América (EUA) ocupa a primeira posição no *ranking* dos países com maior número de endereços IPs classificados como fonte de ataques DDoS, contando com 44% e o Brasil figura na quinta posição neste mesmo período, com 3% de endereços IP envolvidos como fonte de DDoS. Além disso, o Brasil ocupa a mesma quinta posição na lista das fontes de ataque contra aplicações web, com aproximadamente 6% destes, no quadro mundial. Entretanto, ao considerar apenas os países das Américas, o Brasil ocupa a segunda posição deste ranking, perdendo apenas para os EUA. Mas, além de figurar como **fonte** de ataques DDoS, o Brasil também aparece na lista dos países **alvo** de ataques DDoS, tratando especialmente de aplicações web, ocupando a segunda posição em todo o mundo.

Ainda conforme Akamai (2017), os dois seguimentos da indústria mais afetados pelos ataques são: o de *games* e o de Internet & telecom com aproximadamente 90% dos ataques direcionados a elas. O setores Financeiro e Público também figuram em rankings como alvo de ataques DDoS Mansfield-Devine (2015). Em geral, as consequências de um ataque DDoS para esses setores são muito negativas, começando por manchar as marcas dos respetivos negócios até a indisponibilidade total dos serviços ofertados e ainda prejuízos financeiros diretos. Este último, em dois aspectos principais. O primeiro quando a infraestrutura de TI da organização é disponibilizada na nuvem e paga de acordo com a demanda, como exemplo o consumo de links e/ou de processamento. E o segundo aspecto, quando perdem em faturamento, impactando assim economicamente as organizações.

Este estudo apresenta o sistema STARK (do inglês, *prediction **S**ys**T**em against ddos **A**ttack on Netwo**RK***) um sistema autoadaptável para predição de ataques DDoS. Particularmente, o sistema identifica indícios do ataque na rede antes que o mesmo alcance estágios avançados. Com base na teoria de metaestabilidade, o sistema STARK realiza a predição de forma automatizada, sem supervisão e sem rótulos. Ele toma como base a análise de variações nos estados da rede, medidos através de indicadores estatísticos. Estes indicadores ajudam a identificar as transições

[1] Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil. https://www.cert.br/.

irreversíveis no estado da rede. Desta forma, o sistema STARK segue três etapas de operação: (*i*) medições e preparação dos dados, (*ii*) predição dos ataques, e (*iii*) emissão de alertas. O sistema recebe como entrada dados brutos do tráfego da rede, filtra esses dados e compõe séries temporais correspondentes aos períodos (janelas) de medições. Para cada janela, calcula-se os valores para os indicadores estatísticos, e com base nestes, o sistema STARK identifica sinais de um ataque DDoS e emite um alerta.

1.1 Motivação

As abordagens relacionadas aos ataques DDoS focam primeiramente na detecção e mitigação. Em geral, os ataques DDoS possuem duas etapas. A primeira de reconhecimento da vítima (ex. enlaces ou servidores), e a segunda que é a execução propriamente dita. No momento em que um ataque DDoS entra em execução a sobrecarga da vítima tende a ser alcançada em um curto espaço de tempo, dificultando ou inviabilizando o adequado tratamento para evitar ou minimizar suas consequências. Adicionalmente, a revisão da literatura tem mostrado estudos que começa a discutir sobre a emissão de alerta antecipados ou a predição de sinais que apontam os ataques DDoS na rede. Entretanto, tais estudos apontam caminhos que, em geral, abordam a predição por meio do uso de abordagens e técnicas relacionadas à inteligência artificial, redes neurais, ente outras. Contudo, estas abordagens dependem do conhecimento prévio do estado da rede ou das assinaturas do ataque DDoS, limitando-as a ataques conhecidos. Portanto, os desafios impostos para uma nova abordagem com relação aos ataques DDoS incluem predizer o ataque com a maior antecipação possível, ser independente de conhecimentos prévios do estado da rede e ser independente das assinaturas dos ataques.

1.2 Definição do Problema

Os ataques DDoS evoluem continuamente, expandindo em quantidade, volume de requisições (de linear para exponencial) e sofisticação técnica Cisco (2017) Networks (2017). Os estudos apontam técnicas e ferramentas capazes de detectar e mitigar os ataques DDoS, através do conhecimento do fluxo normal e anômalo da rede, dados históricos e assinaturas previamente conhecidas Bhuyan et al. (2015); Nezhad et al. (2016). As soluções atuais apresentam limitações ao tratar ataques DDoS desconhecidos (*zero-day or unknown attacks*). Por exemplo, ao utilizar um IDS (*Intrusion Detection System*) baseado em assinaturas, exige-se um conhecimento prévio do comportamento do fluxo de dados inerente que possa ser comparado ao fluxo corrente da rede e, então, apontar a ocorrência de um ataque DDoS. Outro exemplo consiste na aplicação de redes neurais, entretanto, estas necessitam de treinamento prévio através de conjuntos de dados que contém os fluxos da rede e dos ataques que se deseja detectar ou mitigar.

A janela de tempo entre o início do ataque e a saturação dos recursos da rede (ex. enlaces, servidores ou serviços) é muito pequena, portanto a ação de mitigação ou bloqueio efetivo do ataque deveria ser instantânea, a fim de evitar que usuários legítimos sejam penalizados com a indisponibilidade dos serviços. Pesquisas recentes propõem estratégias para a predição de ataques DDoS e emissão de alertas. No geral, as abordagens embasam-se em técnicas de mineração de dados, modelos estatísticos, redes neurais e modelos de Markov. Por exemplo, Kwon et al. (2017) expõem um método pró-ativo para predizer o volume de ataques DDoS em uma rede através da análise de regressão e correlação. Em Nijim et al. (2017), um sistema para predizer ataques DDoS na camada de aplicação segue a técnica de mineração de dados e classificação das requisições com base no histórico de uso de recursos. Wang et al. (2017) fazem correlações do

comportamento temporal, espacial e espaço-temporal dos ataques e suas dinâmicas. Os modelos estatísticos e comportamentos das *botnets* são usados para treinar o sistema e predizer ataques DDoS. Azzouni e Pujolle (2017) aplicam rede neural sob séries temporais para prever a matriz de tráfego para predizer anomalias. Entretanto, essas soluções ainda dependem de treinamentos prévios, muitos deles *off-line*, ou dependem de conhecimento do comportamento dos ataques.

As abordagens adotadas até então, limitam a atuação das soluções aos ataques previamente conhecidos. Assim, a predição não supervisionada torna-se crucial para evitar custos e perdas resultantes de ataques DDoS desconhecidos. O desafio está na antecipação de ataques DDoS com a finalidade de tratá-los de forma pró-ativa e não apenas de maneira reativa, quando o ataque está em progresso.

1.3 Objetivos

Este trabalho advoga pela predição de ataques DDoS conhecidos (*known*) e desconhecidos (*unknown*). O objetivo é identificar ataques DDoS antes da sobrecarga da rede ou do servidor alvo. O intervalo de tempo entre as ações coordenadas do atacante e a sobrecarga completa da vítima é muito curto, por isso, esta pesquisa busca predizer o ataque DDoS com a maior antecedência possível, para permitir ações preventivas e assim evitar as consequências resultantes Santos et al. (2017). Desta forma, esta pesquisa propõe a construção do sistema STARK (do inglês, *prediction **S**ys**T**em against ddos **A**ttack on Netwo**RK***), cujo objetivo é predizer a iminência de ataques DDoS volumétricos por meio da identificação de sinais de aproximação de transições críticas em dados de fluxo de rede, a fim de antecipar a iminência de ataques DDoS. Estes sinais são identificados com base em séries temporais, extraídas do fluxo de dados da rede. Estas séries temporais são submetidas ao conjunto de indicadores estatísticos que mostram os sinais com base no comportamento apresentado. A identificação destes sinais demonstra a possibilidade de uma ruptura no fluxo da rede, ou seja, um ataques DDoS.

Esta abordagem difere da detecção e mitigação, que em geral ocorrem de maneira reativa e quando a sobrecarga da vítima já está em estágios avançados Ramaki e Atani (2016). Assim sendo, este sistema visa complementar ferramentas e soluções existentes contra ataques DDoS volumétricos, ou seja, com volumes de dados significativos e com ampla capacidade de interrupção dos serviços de Internet, tanto como de enlaces e servidores. Desta forma, o sistema identifica sinais da aproximação de transições críticas no fluxo de dados da rede, e emite alertas quando observada a proximidade de ataques. Portanto, é possível realizar ajustes antecipados no ambiente de rede com a finalidade de minimizar, ou até mesmo evitar as consequências e impactos resultantes dos ataques DDoS.

1.4 Contribuições

O sistema STARK foi avaliado através de uma abordagem orientada a traços. As avaliações empregaram três conjuntos de dados que contém tráfego geral da rede, inclusive tráfego de ataques DDoS. Particularmente, os dados possuem ataques DDoS realizados sobre o ICMP e o UDP. Essas bases de dados são disponibilizadas pelo CAIDA (*Center for Applied Internet Data Analysis*) CAIDA (2007), pela CTU (*Czech Technical University*) García e Uhlir (2011) e pela DARPA (*Defense Advanced Research Projects Agency*) Laboratory (2000). Os traços são avaliados em janelas, das quais realiza-se a extração das informações que servem como entrada para a predição. Os resultados obtidos apontam a viabilidade e o potencial do sistema em predizer ataques DDoS sem assumir o conhecimento prévio do comportamento do ataque ou

treinamentos. O sistema é capaz de predizer um ataque com minutos ou horas de antecedência, visto que nos dados da CAIDA apontou o ataque 23 minutos antes da sobrecarga do alvo, nos dados da CTU identificou o ataque com 1 hora e 18 minutos de antecedência da sobrecarga, e nos dados da DARPA com 2 horas de antecedência. Portanto, entre as contribuições deste trabalho estão: (*i*) a revisão da literatura, (*ii*) a sistematização do processo para a predição de ataques, e (*iii*) o sistema STARK, que se propõe a predizer ataques DDoS volumétricos de forma dinâmica, *online* e com características autoajustáveis.

1.5 Estrutura da Proposta

Esta proposta está organizada como segue. O Capítulo 2 apresenta conceitos necessários para compreensão do arcabouço que compõe ataques DDoS, explana sobre a teoria da meta-estabilidade e sobre a aprendizagem estatística em associação com os indicadores aplicados à solução proposta. O Capítulo 3 expõe os trabalhos relacionados descrevendo aspectos referentes à predição de ataques DDoS. A proposta desta dissertação está detalhada no Capítulo 4 juntamente com a explanação da sistematização do método aplicado com a finalidade de predizer ataques DDoS. A validação do estudo é apresentada no Capítulo 5 com a aplicação dos indicadores nos conjuntos de dados do CAIDA, da CTU e da DARPA. As conclusões, com apontamentos referentes à continuidade da pesquisa e trabalhos futuros, são apresentados no Capítulo 6.

2 Fundamentos

Este capítulo descreve os principais conceitos relacionados a este trabalho. A Seção 2.1 aborda e mostra os fundamentos de ataques DDoS, suas caraterísticas e classificação. A Seção 2.2 apresenta os conceitos e características da aprendizagem estatística e aplicações. Ainda, essa seção descreve a teoria da metaestabilidade e como ocorre a predição de ataques DDoS, além de apresentar os indicadores estatísticos genéricos aplicados.

2.1 Ataques de Negação de Serviço

O principal objetivo do ataque DDoS é interromper recursos ou serviços específicos na Internet, tornando-os inoperantes e assim comprometendo o acesso de seus usuários legítimos, prejudicando-os diretamente Mirkovic e Reiher (2004). Uma das formas de tornar os serviços da Internet inoperantes é sobrecarregar o alvo (ex. enlaces ou servidores) com significativo volume de requisições. Em geral, o volume de requisições direcionadas a uma vítima é muito maior que a largura de banda dos enlaces ou a capacidade de processamento dos servidores. Desta forma, a consequência dos ataques DDoS é a interrupção dos serviços de Internet disponibilizados pela vítima. A seguir são apresentadas a taxonomia do ataque com suas principais características, as camadas de rede em que atuam, as motivações para os ataques e também os desafios relacionados.

Os ataques DDoS são classificados de diversas formas, que variam conforme a evolução do próprio ataque ou com a expansão dos recursos computacionais envolvidos. Desta forma, pesquisadores apresentam a classificação de acordo com o viés das respectivas pesquisas. Segundo Mirkovic e Reiher (2004) estes ataques são classificados conforme os mecanismos utilizados, sendo caracterizados com relação ao (*i*) grau de automação, (*ii*) vulnerabilidades exploradas para negar o serviço, (*iii*) validade do endereço de origem, (*iv*) dinâmica da taxa de ataque, (*v*) possibilidade de caracterização, (*vi*) persistência do conjunto de agentes, (*vii*) tipo de vítima e (*viii*) impactos na vítima. Ainda sobre a classificação do ataque, Zargar et al. (2013) afirma que os ataques DDoS de inundação são classificados com base na camada de rede em que o ataque é realizado. Entre os tipos de ataques estão os ataques DDoS de inundação na camada de rede ou transporte, que em geral ambos são alocados sob o mesmo grupo, e ataques DDoS de inundação na camada de aplicação.

Entre as motivações dos atacantes para a realização de ataques DDoS, encontram-se a financeira/econômica, vingança, crença ideológica, desafio intelectual e guerra cibernética Zargar et al. (2013); Networks (2017). Este aspecto merece ser observado do ponto de vista da fundamentação, pois refere-se aos estímulos para a realização do ataque. Isto se deve ao observar que tais ataques permanecerão em destaque observadas as motivações, que são em geral humanas e não técnicas. Portanto, mesmo com o avanço das pesquisas, da criação e implementação de soluções, as motivações citadas são indicativos de que continuamente haverão interessados em realizar tais ataques, salvo se houver técnicas, ferramentas e recursos que permitam a extinção completa de execução dos ataques DDoS.

Apesar da gravidade das consequências dos ataques DDoS e a disponibilidade de diversos mecanismos de defesa, ainda persistem os desafios para que seja possível evitá-los, tendo em vista sua contínua evolução em relação à crescente frequência, sofisticação e volume. Entre os desafios é possível verificar a importância de distribuir mecanismos de monitoramento ou defesa em pontos estratégicos, como ISPs responsáveis pelo núcleo da rede. Em geral, isso implica em custos e não necessariamente beneficiará diretamente a organização que os disponibiliza. Dessa forma, é preciso verificar os (*i*) pontos nos quais as soluções sejam mais efetivas e também os (*ii*) custos para essas organizações. Há ainda outros aspectos que são desafios para soluções mais efetivas, entre os quais estão a (*iii*) falta de informações detalhadas sobre os ataques, (*iv*) *benchmarks* entre sistemas de defesa e (*v*) a dificuldade de realizar testes em larga escala em ambientes reais Mirkovic e Reiher (2004).

De acordo com Zargar et al. (2013), os pontos de alocação de mecanismos contra DDoS podem ser: próximo à origem, próximo à vítima, baseados em rede ou híbrido. Os mecanismos de defesa alocados próximo à fonte, detectam e respondem aos ataques quando implantados próximos aos *hosts* de origem. Mecanismos designados próximo à vítima, concentram-se em detectar e responder a ataques, quando implantados próximos aos *hosts* de destino, ou seja, aos alvos do ataque. Também há a possibilidade de soluções baseadas em rede, que são implantadas em pontos intermediários das redes, como exemplo em roteadores. Por fim, há mecanismos classificados como híbridos, ou seja, estes compreendem soluções distribuídas que atuam tanto próximo às fontes, às vítimas e em pontos intermediários da rede, atuando de forma cooperativa.

Conforme Zargar et al. (2013), existem duas estratégias principais para disparar ataques DDoS. Na primeira, o atacante envia pacotes adulterados para o alvo com a finalidade de confundir o protocolo ou a aplicação em execução, interrompendo o serviço ou até mesmo o hospedeiro (ex. servidor). Esse ataque é conhecido como ataque de vulnerabilidade. A segunda estratégia pode ser implementada de duas formas: (*i*) o atacante interrompe as conexões de usuários legítimos através do esgotamento da largura de banda, da capacidade de processamento do roteador ou dos recursos da rede, através do envio de grande volume de requisições. Na outra forma, (*ii*) o atacante provoca a ruptura das conexões dos usuários legítimos exaurindo recursos dos servidores como *sockets*, CPU, memória e disco. A primeira forma refere-se a um ataque que é conhecido como ataque de inundação da camada de rede ou transporte (ex. ICMP *flooding*, TCP SYN *flooding*). A segunda forma, também é um ataque de inundação, contudo, é essencialmente da camada de aplicação (ex. HTTP *flooding*). Assim, é possível observar que os ataques DDoS de inundação também são classificados em relação ao tipo de ataque e respectiva camada de rede em que atuam.

De acordo com Zargar et al. (2013), os ataques DDoS são disparados remotamente de forma coordenada e por meio de dispositivos zumbis ou *botnets* amplamente distribuídas. Segundo Mahmoud et al. (2015), este ataque é comumente realizado através de *botnets*, ou seja, redes compostas por dispositivos infectados sob o controle dos atacantes. *Botnet* é um termo cunhado para descrever uma rede de hospedeiros (*bots*) infectados, os quais executam softwares robôs controlados por um humano (*botherder*), por meio de um ou mais controladores (*botmasters*). As *botnets* são implementadas utilizando diferentes arquiteturas, entre elas a centralizada (ex. IRC, HTTP) e a descentralizada (ex. P2P). No geral, o atacante utiliza um ou mais hospedeiros para ser o *bot* mestre (controlador dos *bots*) e os demais hospedeiros para atuarem como escravos ou zumbis. Com a *botnet* disponível, o atacante configura os hospedeiros escravos que apontarão para um determinado alvo e disparam o ataque DDoS.

Portanto, este ataque é elaborado em duas fases, sendo (*i*) a preparação e (*ii*) a execução propriamente dita. A fase da preparação compreende o estudo do alvo através de varreduras na rede para descobrir vulnerabilidades (ex. sistemas desatualizados, *bugs* conhecidos, etc) e a

construção de uma *botnet*. Na fase de execução os *bots*, sob o comando do *botmaster*, disparam o envio de requisições contra a vítima, alvo do ataque. Esta organização pode ser vista na Figura 2.1. Desta forma, o ataque DDoS é realizado com um volume de requisições que gera um fluxo de rede capaz de comprometer a rede ou seus serviços Mahmoud et al. (2015), conforme a estratégia e ataque adotados. Portanto, ataques DDoS em geral, utilizam *botnets* para gerar grandes volumes de requisições.

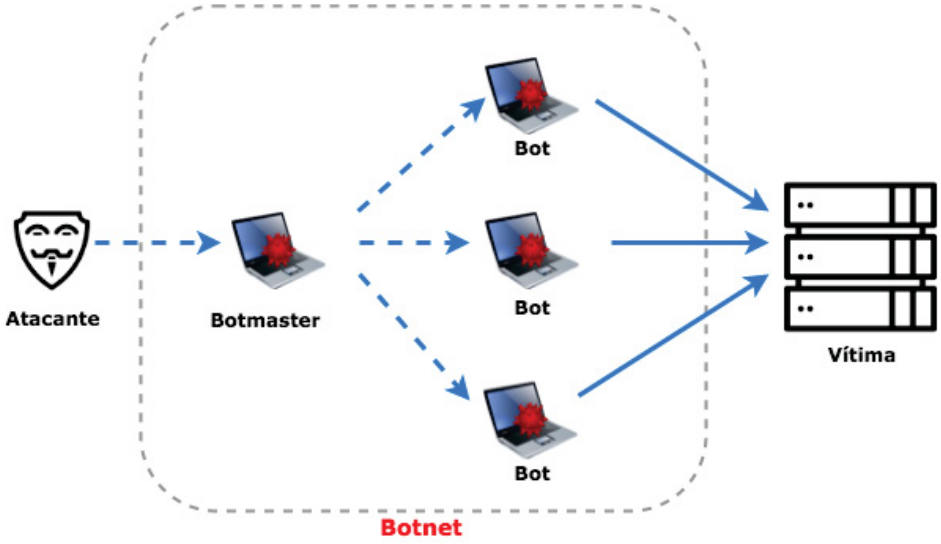

Figura 2.1: Diagrama de um ataque DDoS usando uma *botnet*

A literatura apresenta diversos tipos de ataques DDoS e suas respectivas características Mirkovic e Reiher (2004) Zargar et al. (2013) Bhuyan et al. (2015). Portanto, este trabalho trata o ataque DDoS *flooding* (inundação). Esta variação de ataque DDoS é lançada de forma direta ou refletora. Os ataques diretos disparam um volume de pacotes ou requisições maliciosas contra a vítima (*host* ou servidor), sendo classificados de acordo com a camada da rede utilizada para executar o ataque, sendo mais comum nas camadas de rede/transporte ou de aplicação. Entre os ataques da camada de rede/transporte estão TCP, UDP, ICMP e SYN *flooding*. Dentre os exemplos de ataque da camada de aplicação estão HTTP, HTTPS e FTP *flooding*. Os ataques do tipo refletor enviam requisições para um ou diversos computadores que atuam como um refletor com o IP da vítima (alvo) no remetente (IP *spoofing*). Entre os exemplos há o ICMP ECHO reply, SYN ACK RST e DNS Bhuyan et al. (2015).

Os ataques DDoS têm evoluído como todos os demais tipos de ameaças e também junto com a própria evolução das redes. A expansão e a popularização da Internet, a ampliação da capacidade da infraestrutura, como largura de banda, e a oferta contínua de novos serviços trazem o aumento dos riscos simultaneamente às melhorias. Com a popularização das câmeras de vigilância, *smartphones*, consoles de *video-games*, televisores e uma infinidade de dispositivos conectados à Internet, ampliaram-se exponencialmente as vulnerabilidades. Afinal, cada um desses dispositivos possui um sistema operacional ou um sistema embarcado, que precisam de manutenção e atualizações constantes para evitar vulnerabilidades de segurança. Cada um desses dispositivos é um potencial *bot* (zumbi) participante de uma *botnet*. Assim sendo, os atuais ataques DDoS são caracterizados pela inclusão contínua de novos dispositivos na rede, pelo grande volume de requisições e, na outra extremidade, por volume muito pequenos que se misturam ao fluxo normal da rede e demais variações de tipos de ataques considerado a diversidade de camadas de rede em que atuam. Esse conjunto de características dos ataques DDoS, impõem diversos desafios à comunidade científica, no que tange às abordagens e técnicas

aplicadas para a detecção e a mitigação dos mesmos. Em geral, a detecção e a mitigação dependem de dados históricos ou do ataque em progresso.

2.2 Aprendizagem Estatística

A Aprendizagem Estatística compreende os diversos artefatos da estatística descritiva e inferencial que quando aplicadas permitem a compreensão de sistemas complexos e dinâmicos James et al. (2014). Ou seja, a Aprendizagem Estatística permite aprender e compreender a dinâmica de funcionamento e o comportamento de diferentes sistemas a partir de dados extraídos dos mesmos. Conforme James et al. (2014), a Aprendizagem Estatística é classificada em supervisionada e não supervisionada. A Aprendizagem Estatística supervisionada envolve a construção de modelos estatísticos para predição ou estimativa que são baseadas em uma ou mais entradas (*inputs*) de dados. Neste caso, há a verificação dos resultados com a finalidade de validar o comportamento apresentado pelo modelo aplicado, ou seja, há supervisão no processo. Em contrapartida, na aprendizagem não supervisionada, o sistema tem que descobrir de forma autônoma relações, padrões ou categorias nos dados de entrada. A Aprendizagem Estatística vem crescendo junto com soluções (*hardware* e *software*) computacionais que possibilitam a implementação de modelos estatísticos robustos. Desta forma, tornou-se possível analisar volumes significativos de dados e inferir, ou até mesmo predizer, determinados comportamentos.

De acordo com James et al. (2014), a aprendizagem sem supervisão é muitas vezes muito mais desafiadora. O exercício tende a ser mais subjetivo, e não há um objetivo simples para a análise, como a expectativa de uma resposta. A aprendizagem não supervisionada é muitas vezes realizada como parte de uma análise exploratória de dados. Além disso, pode ser difícil avaliar os resultados obtidos a partir de métodos de aprendizagem não supervisionados, uma vez que não existe um mecanismo universalmente aceito para realizar a validação cruzada ou validar resultados em um conjunto de dados independente. A diferença entre a aprendizagem supervisionada e não supervisionada é percebida ao aplicar um modelo preditivo usando uma técnica de aprendizagem supervisionada. Nesta é possível verificar o trabalho, observando o quão próximo o modelo prediz a resposta Y, nas observações não utilizadas na elaboração do modelo. No entanto, quando aplicado um modelo de aprendizagem sem supervisão, não há como verificar o resultado porque a resposta não é conhecida, ou seja, aprendizagem não supervisionada. Contudo, as técnicas de aprendizagem não supervisionadas vêm ganhando espaço por possibilitar a identificação de determinados comportamentos baseados em características específicas dos dados, através de padrões similares, favorecendo assim a predição de determinados comportamentos. Entre as inúmeras vantagens, a aplicação de técnicas ou ferramentas de aprendizagem não supervisionada destaca-se por predizer comportamentos dos dados sem necessariamente requerer treinamento pregresso e por isso, conforme Scheffer et al. (2015); Dakos et al. (2012); Moon e Lu (2015), pode ser aplicada na predição de comportamento de diversos sistemas complexos e dinâmicos como padrões de circulação oceânica, lagos ou mercados financeiros. Outros estudos demonstram experimentos com a aplicação da aprendizagem não supervisionada nas áreas da saúde, em especial monitoramento de sinais vitais e predição de sinais de depressão Vergutz (2017); Wichers et al. (2016).

2.2.1 A Teoria da Metaestabilidade

A metaestabilidade é um fenômeno comum observado em uma grande variedade de situações, em geral, na natureza Bovier e Den Hollander (2016). Esse fenômeno está relacionado com a observação de equilíbrio e escalas de tempo em um sistema dinâmico e complexo. Uma das

características principais da metastabilidade é a observação de múltiplas escalas de tempo e bem separadas: (*i*) em uma escala de tempo curta, o sistema *parece* estar em um estado de equilíbrio (estado metaestável), explorando apenas uma seção confinada dos seus possíveis estados; (*ii*) enquanto em escalas de tempo muito maiores, *transições* entre os estados metaestáveis do sistema podem ser identificadas. Essa separação de tempos na dinâmica do sistema tem manifestações experimentais evidentes, modeladas matematicamente, por exemplo, em funções de correlação de decomposição de tempo, tais como as utilizadas neste trabalho.

Os primeiros trabalhos sobre metaestabilidade procuraram entendê-la no contexto de reações químicas, eletrônica e circuitos digitais. Hoje, o modelo de metaestabilidade é usado em diferentes aplicações, tais como o estudo das dinâmicas no mercado financeiro, análise de biomassa e até para o entendimento de doenças humanas Vergutz (2017); Dakos et al. (2012). Os estados metaestáveis são características inerentes dos sistemas digitais assíncronos, tais como a Internet. Existem várias formas de estudar o fenômeno da metaestabilidade. Este trabalho está embasado na abordagem que considera o fenômeno ocorrendo em processos estocásticos, e em particular, em um processo markoviano, ou seja, a análise da evolução de uma coleção de variáveis aleatórias – por exemplo, o tamanho dos pacotes de dados na rede – com estados (valores) discretos em que a predição dos estados seguintes depende apenas do conhecimento do estado atual, sendo então irrelevante conhecer os estados anteriores ao estado atual. Particularmente, a predição dos ataques DDoS desconhecidos neste trabalho toma como base a predição de *transições críticas*, ou seja, quando as mudanças de um estado metaestável para outro ocorrem de forma irreversível Dakos et al. (2012).

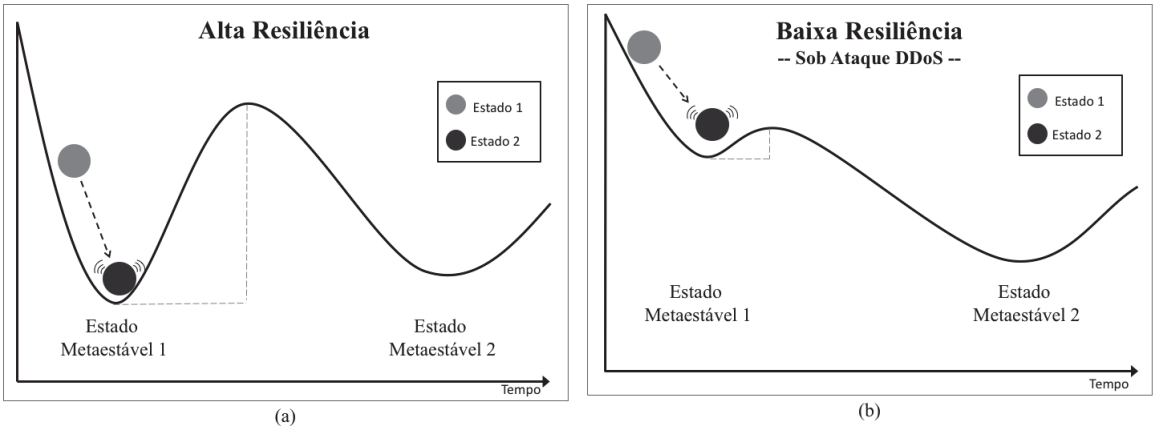

Figura 2.2: Perda de resiliência de um sistema metaestável devido ao ataque DDoS

A Figura 2.2 exibe duas situações para ilustrar o equilíbrio do sistema em estados metaestáveis e suas transições: (*i*) sistema com alta resiliência (transições críticas são improváveis), e (*ii*) sistema com baixa resiliência (iminência de transição crítica). A Figura 2.2(a) representa o sistema em uma condição considerada de equilíbrio com dois estados metaestáveis distintos. Existe uma barreira gerando uma alta resistência na passagem de um estado metaestável para outro. O estado do sistema, representado em dois momentos pelos círculos cinza (estado 1) e preto (estado 2), pode até mudar de estado (valor), mas permanece no estado metaestável (o vale). Na Figura 2.2(b), percebe-se também a existência de dois estados metaestáveis, porém a resistência que impede uma transição de um estado metaestável para outro é baixa. Dependendo da aplicação, existem várias causas para reduzir a barreira que controla a transição. Neste estudo, a causa na redução na barreira é instanciada para perturbações que ocorram devido ao ataque.

Embasado nessas dinâmicas e nos modelos matemáticos que as representam, o sistema proposto neste trabalho prediz os ataques DDoS através do cálculo de indicadores estatísticos

que permitem observar tendências no comportamento do fluxo da rede e identificar transições críticas. O sistema trata *features* (i.e., tamanho de pacotes, quantidade de pacotes, entre outros) como variáveis aleatórias e com base nas medições da rede consegue predizer a aproximação de transições críticas. Ou seja, a iminência de ataques DDoS.

2.2.2 Indicadores Estatísticos Genéricos

O conjunto de indicadores estatísticos genéricos avaliados tem como finalidade expor o comportamento dos dados. Com o mapeamento deste comportamento é possível identificar a aproximação de transições críticas que apresentam a tendência de ruptura em consequência de perturbações impostas a um sistema. A ruptura por sua vez mostra que o estado metaestável deverá sofrer mudança abrupta e assim atingir um novo estado. Essa mudança de estado tem significados distintos, variando conforme os dados avaliados. Como exemplo é possível citar mudanças abruptas na biomassa, nos mercados financeiros ou no estado de saúde de uma pessoa por meio dos seus sinais vitais. E, neste caso em especial, um ataque DDoS.

A **taxa de retorno** corresponde em termos estatísticos à probabilidade de um sistema se recuperar de uma perturbação (ex. queda repentina no mercado de ações) e retornar à estabilidade. O sistema se revela resiliente quando a taxa de retorno é crescente, ou seja, indica que tende a não sofrer uma transição crítica. Contudo, quando o sistema é submetido a perturbações severas, estas forçam o seu estado a atingir um estado crítico, mudando para um novo estado, em geral inesperado. Este novo estado crítico provoca grande instabilidade no sistema, resistindo ou até mesmo impossibilitando o retorno ao estado antigo (metaestável). A taxa de retorno quantifica o tempo de retorno a um estado estável. Portanto, um aumento na taxa de retorno indica que o sistema possui resiliência e está se recuperando facilmente das perturbações sofridas. Dessa forma, o sistema perde a resiliência quando há queda na taxa de retorno tornando-o vulnerável às perturbações. Assim, quando a taxa de retorno é decrescente há indícios de que se aproxima uma transição crítica, apontando que há a iminência de que um fenômeno possa criar perturbações suficientes para causar uma ruptura no estado do sistema levando-o a um novo estado inesperado.

A **autocorrelação** estima a correlação entre observações sucessivas de uma série temporal. Ela calcula o quanto o estado do sistema se tornou similar entre observações consecutivas Dakos et al. (2012). Assim, um aumento na autocorrelação é esperado, fornecendo indícios sobre a aproximação de tais eventos Scheffer et al. (2009), nesse caso, como exemplo um ataque DDoS. A autocorrelação é calculada neste trabalho seguindo a Equação 2.1, onde as variáveis z_t e z_{t+1} representam duas observações consecutivas, μ a média das observações da série temporal, e σ a variância da variável z no tempo t. Com base nisto, o aumento da autocorrelação é esperado para indicar a aproximação da transição crítica, ou seja, perturbações que provocam a ruptura do estado do sistema.

$$\rho_1 = \frac{E[(z_t - \mu)(z_{t+1} - \mu)]}{\sigma_z^2} \tag{2.1}$$

O **coeficiente de variação** analisa a dispersão das observações de uma série temporal em relação a sua média μ. Desse modo, uma grande dispersão nos valores das observações indica a presença de valores instáveis nos dados Scheffer et al. (2009). Na iminência de eventos críticos ou após perturbações, o estado de um sistema tende a se alterar amplamente em torno de um estado estável, atingindo valores considerados críticos (ex. tamanho do pacote em torno de 1500 bytes), e assim aumenta a variabilidade das observações. Estatisticamente, o coeficiente de variação representa a variância das observações, sendo calculado através da seguinte equação: $CV = \frac{SD}{\mu}$, onde SD é o desvio padrão das observações da série temporal. O desvio padrão representa a

variabilidade dos dados e pode ser estimado a partir da Equação 2.2, onde n representa o número de observações da série temporal, z a variável analisada no tempo t e μ a média. Conforme a literatura, um aumento no coeficiente de variação aponta um aumento na instabilidade e a possibilidade de ocorrer um evento crítico em um sistema, neste caso, como exemplo um ataque DDoS Dakos et al. (2012) Nogueira et al. (2017).

$$SD = \frac{1}{n-1} \sum_{n}^{t=1} (z_t - \mu)^2 \tag{2.2}$$

Por fim, a **assimetria** é um indicador que compara a distribuição de uma série temporal em relação a uma distribuição simétrica (distribuição normal). Na Estatística, uma distribuição simétrica representa uma distribuição estável, sem condições críticas. Por outro lado, a apresentação de uma assimetria na distribuição da série temporal revela a presença de valores críticos nas observações. Mais especificamente, a Equação 2.3 apresenta a função para estimar a medida da assimetria. Nesse sentido, para apontar a iminência de um evento crítico, neste cenário um ataque DDoS, é esperado um aumento na curva da assimetria da distribuição. Além disto, a assimetria pode aumentar ou diminuir dependendo se os valores críticos do estado da rede tendem para um estado maior ou menor em relação ao estado atual Scheffer et al. (2009).

$$1\gamma = \frac{\frac{1}{n}\sum_{t=1}^{n}(z_t - \mu)^3}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(z_t - \mu)^2}} \tag{2.3}$$

O conjunto destes quatro indicadores estatísticos apresenta um comportamento genérico quando o estado de um sistema se encontra próximo de uma mudança de estado. Esse comportamento engloba um aumento nos valores da autocorrelação, coeficiente de variação, e assimetria, enquanto mostra uma queda na taxa de retorno Scheffer et al. (2009). Este comportamento para os indicadores acima aponta a possibilidade do estado de um sistema atravessar uma transição crítica. Dessa maneira, permite prever a tendência de ocorrer uma transição crítica (ex. ataque DDoS) Dakos et al. (2012).

2.3 Aplicação dos Indicadores Estatísticos

Com a finalidade de ilustrar a tendência que pode ser exposta pelos indicadores, a Figura 2.3 apresenta o comportamento esperado do grupo de indicadores avaliados para mostrar a tendência de aproximação da transição crítica. A taxa de retorno apresenta um comportamento decrescente na Figura 2.3. Este comportamento indica que o retorno ao estado de equilíbrio está cada vez mais difícil e portanto mais lento. Conforme aumentam as perturbações no sistema avaliado, este tende a perder sua resiliência, e assim torna-se mais difícil retomar a estabilidade. Neste cenário, a perturbação pode ser o processo de preparação de um ataque DDoS, quando ocorrem varreduras na rede, a busca por vulnerabilidades, teste da capacidade de banda e/ou da capacidade de resposta do alvo (ex. servidores, serviços ou enlaces). Neste caso, o tamanho dos pacotes da rede com relação ao tamanho dos pacotes em relação ao tempo. Dessa forma, quanto mais lento o tempo de retorno ao estado normal, maior a proximidade de uma transição crítica se encontra o sistema. A autocorrelação, com defasagem no tempo, se apresenta com tendência de crescimento. Este comportamento é um indicativo da iminência de uma transição crítica. Ou seja, conforme o estado da rede se aproxima dos limites críticos a autocorrelação tende a aumentar. A variância nas observações, representadas neste pelo coeficiente de variação, mostra instabilidade na rede, através de sua tendência de crescimento e a assimetria apresenta a

existência de valores distantes do estado estável, o que provoca aumento na assimetria da curva de distribuição dos valores observados. Apontando que o estado da rede permanece próximo a um evento crítico. Cabe observar que se o estado tende para maior ou menor em relação ao estado atual, a assimetria os refletirá indicando a tendência de aumento ou diminuição respectivamente.

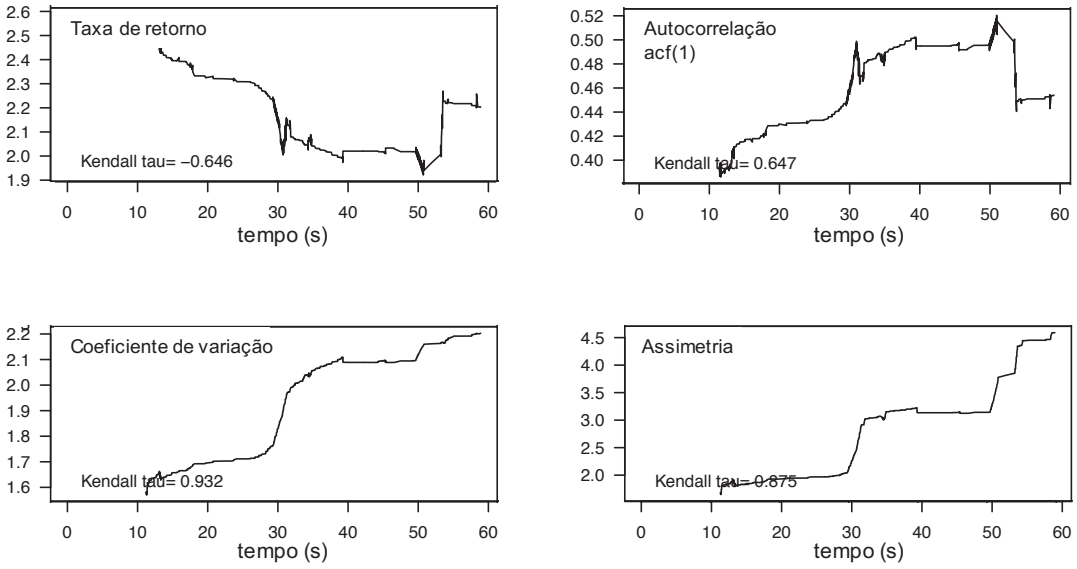

Figura 2.3: Exemplo de indicadores com comportamento de aproximação da transição crítica

2.4 Resumo

Este capítulo apresentou os conceitos necessários para o entendimento do restante deste manuscrito. Iniciou pela fundamentação dos ataques ataques DDoS de inundação, conceituando-os com sua dinâmica e apresentando suas características e respectivos exemplos. Conceituou também a aprendizagem estatística como um conjunto de conhecimentos e artefatos que tem como finalidade disponibilizar as ferramentas necessárias para a compreensão da análise preditiva com base na teoria da metaestabilidade juntamente com a compreensão de resiliência aplicada a sistemas. Adicionalmente, auxilia no entendimento dos próprios indicadores estatísticos genéricos, também descritos neste capítulo a fim de construir o arcabouço de conceitos utilizados nesta pesquisa que propõe o sistema de predição de ataques DDoS.

3 Revisão Bibliográfica

Diversos estudos abordam o desenvolvimento de técnicas para predição do comportamento de sistemas complexos e dinâmicos. Em geral, estas pesquisas utilizam estatística ou inteligência artificial para antecipar o comportamento de sistemas em diversas áreas do conhecimento. Existem diversos estudos sobre predição e suas aplicações. Neste caso, mais especificamente, a predição de ataques de negação de serviços distribuído (DDoS).

3.1 Predição de Ataques DDoS

Na literatura existem trabalhos que abordam o desenvolvimento de técnicas para detecção e predição de ataques DDoS em seus estágios iniciais. Em geral, as técnicas utilizadas envolvem artefatos estatísticos, de inteligência artificial (ex. redes neurais) e aprendizagem de máquina. Em Nezhad et al. (2016), os autores abordam o problema da detecção de ataques DDoS e propõem um algoritmo para classificar o tráfego como normal ou malicioso. Para tal, constituem uma série temporal baseada no número de pacotes e a normalizaram por meio da transformação de Box-Cox. Também implementam o modelo estatístico de predição ARIMA e calculam o expoente de Lyapunov que indica a estabilidade na série temporal. De acordo com os autores, o algoritmo proposto é capaz de classificar o tráfego normal e o tráfego malicioso com precisão de 99,5%. Contudo, este estudo precisa de conhecimento pregresso da dinâmica do fluxo da rede com a finalidade de treinar o algoritmo para classificar o tráfego normal do anômalo, o que restringe a predição somente a ataques previamente conhecidos e com a dinâmica do tráfego da rede devidamente mapeada.

Em Tsai et al. (2010), os autores propõem um sistema de alertas antecipados multicamadas para geração de alertas precoces baseado em redes neurais e implementado sobre um IDS. Neste sistema, os computadores atuam de forma colaborativa em que cada um monitora seus nós vizinhos. Os dispositivos enviam os dados coletados para um módulo que analisa os dados e verifica combinações com os padrões de DDoS. Assim, tal sistema é capaz de detectar o ataque DDoS com sucesso a uma taxa de 82,7%. Desta forma, o sistema proposto precisa do conhecimento prévio das assinaturas de ataques DDoS, inviabilizando sua aplicação na predição de ataques DDoS desconhecidos ou ainda com a assinatura não identificada, pois a rede neural precisa ser treinada previamente e ter seus pesos devidamente ajustados. Outra proposta observada em Xiao et al. (2006) apresenta um sistema colaborativo que visa alertar quando ocorre um ataque DDoS ainda em estágios iniciais. O sistema é composto de um módulo de detecção no cliente e um módulo no servidor. Neste estudo, o sistema é baseado num filtro de Bloom modificado que monitora *handshakes* anormais e utiliza uma estrutura de dados baseada em *hash*. O filtro de Bloom identifica se determinado elemento pertence a um conjunto de dados e armazena a informação de forma probabilística. Conforme os autores, o sistema emitiu alerta de ataque DDoS logo em seus estágios iniciais. Contudo, se o volume de entradas (*handshakes* anormais) na tabela *hash* é significativo, o algoritmo sofrerá com as

colisões de tabelas *hash*, que potencialmente implicarão em maior número de falsos positivos e falsos negativos. Adicionalmente, a função *hash* (k) precisa ser ajustada para garantir maior efetividade. Esses aspectos exigem o conhecimento prévio das características do tráfego da rede.

O estudo de Kwon et al. (2017) propõe um método pró-ativo que estima o volume (quantidade) de ataques DDoS. A intenção dos pesquisadores é superar limitações impostas pelos sistemas de segurança reativos baseados na detecção de intrusão e avaliar a necessidade de implantação de sistemas IPSs na rede. A medição do fluxo da rede foi realizada através de sistemas *honeynet*, *logs* IDS e traços de atividade de intrusão que identifica tentativas de ataques. A partir disso, os autores estimaram o número de *bots* com base no número de usuários da rede em associação com os dados disponibilizados por uma pesquisa que informa a porcentagem de *bots* estimada para o país. Com base nas características do tráfego da rede, no número de usuários e na porcentagem de *bots*, foi realizada a estimativa do volume de ataques DDoS nesse ambiente através da análise de regressão e de correlação. Portanto, foi necessário o levantamento de dados históricos do tráfego de rede e das tentativas de ataques DDoS. Além disso, a predição utiliza como parâmetro a estimativa do número de *bots* por país. Este parâmetro é um dado externo, ou seja, existe a possibilidade da indisponibilidade desses dados para determinado país, ou ainda a rápida desatualização dessa informação, comprometendo a qualidade da solução proposta.

O trabalho de Nijim et al. (2017) argumenta a favor de um sistema que utiliza a mineração de dados para predizer e prevenir ataques DDoS na camada de aplicação. A proposta é priorizar requisições legítimas em detrimento do tráfego de ataque por meio de um mecanismo automático de priorização da comunicação. A priorização ocorre através da classificação das requisições com base no histórico do uso de recursos como tempo de CPU, memória, espaço em disco e tráfego da rede. No entanto, esse trabalho não apresentou resultados de predição, pois se encontra em fase de desenvolvimento. Além disso, o método necessita de dados históricos e assinaturas. Em Wang et al. (2017) são apresentados três modelos orientados a dados que capturam o comportamento temporal, espacial e espaço-temporal dos ataques DDoS, caracterizados pelas suas dinâmicas e comportamentos. O objetivo desses modelos estatísticos é predizer a ocorrência de ataques DDoS bem como as características comportamentais das *botnets*. Para isso, empregam traços de ataques DDoS verificados a partir de operações de mitigação e calculam as correlações temporais, espaciais e espaço-temporais das características de ataques (ex. número de *bots* e duração do ataque). Essas características são obtidas por análises de engenharia reversa. Os modelos propostos neste trabalho tem como requisito o conhecimento de dados históricos de ataques DDoS por meio de *botnets*.

Em Azzouni e Pujolle (2017), os autores aplicam uma rede neural sobre um modelo de séries temporais (*Long Short-Term Memory*) para classificar, processar e prever a matriz de tráfego em grandes redes. A matriz de tráfego da rede tem entre suas aplicações contribuir com o gerenciamento da rede e consequentemente coopera com a detecção de anomalias. Dessa forma, ao predizer a matriz de tráfego é possível aplicá-la na predição de ataques. Contudo, a abordagem adotada demanda o treinamento prévio da solução para que seja efetiva na predição de ataques DDoS. Portanto, necessita de conhecimento prévio do tráfego da rede. Outros estudos Zan et al. (2009); Holgado et al. (2017) propõem métodos baseados em Modelos Ocultos de Markov (*HMM - Hidden Markov Model*) para predizer ataques de múltiplos passos, como ataques DDoS. Os múltiplos passos compreendem as fases realizadas durante um ataque, como busca de vulnerabilidade e fase de preparação. Dessa forma, os métodos objetivaram prever os próximos passos dos ataques com base nos passos anteriores e dados históricos. Para esse fim, o processo estocástico oculto do modelo de Markov foi representado pela sequência de diferentes passos de ataques observados nos alertas emitidos por IDSs. Esses alertas são transformados em

observações e agrupados em *clusters* conforme a gravidade. Uma vez treinado o modelo, ele é capaz de predizer a probabilidade das etapas dos ataques.

No caso dos estudos de Nezhad et al. (2016), Tsai et al. (2010) e Xiao et al. (2006) é possível observar, que apesar de discorrerem sobre a predição ou a emissão de alertas antecipados aos ataques DDoS por meio do uso de classificadores, é possível afirmar que somente apontam o ataque DDoS no início da sobrecarga do alvo, caracterizando na prática a detecção e não predição de ataques. Somando-se a isto, os estudos analisados demandam conhecimento histórico da rede, das assinaturas dos ataques DDoS ou ainda o treinamento prévio de algoritmos. Os textos afirmam que as soluções emitem alerta logo nos estágios iniciais do ataque DDoS. Contudo, não há clareza quanto à definição de estágio inicial, visto que tanto pode estar relacionado às etapas de preparação do ataque como varreduras para o reconhecimento da rede, varreduras em busca de vulnerabilidades ou ainda os instantes (*ms*) prévios à sobrecarga de enlaces ou servidores. As características levantadas a partir das soluções propostas estão resumidas na Tabela 3.1.

Tabela 3.1: Levantamento de características das soluções propostas

		Dependente de		
Autor	Abordagem	Assinatura DDoS	Tráfego da Rede	Treinamento
Xiao et al. (2006)	Estatística			✓
Zan et al. (2009)	Estatística	✓	✓	✓
Nezhad et al. (2016)	Estatística		✓	
Kwon et al. (2017)	Estatística		✓	
Nijim et al. (2017)	Estatística		✓	
Wang et al. (2017)	Estatística		✓	
Holgado et al. (2017)	Estatística	✓	✓	✓
Tsai et al. (2010)	Redes neurais	✓		✓
Azzouni e Pujolle (2017)	Redes Neurais	✓	✓	✓

Consideradas as limitações dos trabalhos anteriores, neste estudo procuramos aprofundar a investigação sobre a possibilidade de antecipar ataques DDoS por meio dos indicadores estatísticos genéricos. Através desta pesquisa, argumentamos sobre a possibilidade de identificar sinais precoces que apontam a aproximação de mudanças disruptivas na rede a fim de indicar a tendência de ataques DDoS, mesmo sem o uso de dados históricos ou assinaturas previamente conhecidas. De acordo com uma avaliação preliminar de Nogueira et al. (2017) é possível aplicar indicadores estatísticos genéricos como forma de predizer ataques DDoS. O estudo propõe a utilização de um conjunto de indicadores genéricos como *taxa de retorno*, *autocorrelação*, *coeficiente de variação* e *assimetria* com a finalidade de apontar a aproximação de transições críticas que causam mudanças disruptivas na rede. Estas mudanças, quando associadas aos ataques DDoS podem então indicar a iminência destes ataques. O estudo demonstra empiricamente a aplicação do conjunto de indicadores para identificar a tendência de aproximação do ataque.

Além das técnicas aplicadas para a antecipação de alertas, Ramaki e Atani (2016) faz uma revisão abrangente sobre EWSys (*Early Warning Systems*) demonstrando propostas para a construção de sistemas de alertas precoces a fim de antecipar ataques DDoS. Tais sistemas são uma abordagem pró-ativa contra ameaças como o DDoS e são complementares aos IDSs (*Intrusion Detection Systems*) e IPSs (*Intrusion Prevention Systems*). Os autores apontam ainda uma série de desafios relacionados aos EWSys, como a precisão na predição de ataques e a priorização da geração de alertas antecipados. As diversas técnicas para antecipação de ataques DDoS, atualmente em construção, farão parte de ferramentas mais abrangentes que os tradicionais IDSs e IPSs. Dessa forma, considerando as características dos trabalhos apresentados e com

base na pesquisa iniciada sobre predição de ataques DDoS, este trabalho apresenta o STARK, um sistema de predição de ataques DDoS. O sistema faz uso da teoria da metaestabilidade e da aprendizagem estatística em associação com os indicadores estatísticos genéricos. O objetivo do sistema é prever um ataque com a maior antecedência possível, a fim de possibilitar a tomada de medidas preventivas em tempo de evitar graves consequências.

3.2 Resumo

Os trabalhos relacionados foram listados e discutidos com base nas características e técnicas utilizadas. Entre as abordagens discutidas estão redes neurais e estatística. O capítulo apresentou referências que preconizam uma nova geração de ferramentas capazes de tratar os ataques DDoS de forma pró-ativa, antecipando os sinais e assim compondo sistemas de alertas antecipados (do inglês *Early Warning Systems*) para complementar e avançar as tecnologias já existentes para detecção e reação aos ataques DDoS. E por fim, este abordou a aplicação da teoria da metaestabilidade e a aprendizagem estatística utilizadas na predição de ataques DDoS.

4 Sistema de Predição de Ataques DDoS

Este capítulo apresenta o sistema para a identificação da iminência de ataques DDoS volumétricos em redes, denominado STARK (*Prediction **S**ys**T**em against DDoS **A**ttack on Netwo**RK***). O sistema STARK complementa as estratégias e soluções preventivas, reativas, e tolerantes com o objetivo de predizer ataques conhecidos e desconhecidos em seus estágios iniciais e antes que sobrecarreguem a vítima. O sistema proposto aponta a aproximação de ataques DDoS volumétricos por meio de séries temporais submetidas à indicadores estatísticos genéricos como taxa de retorno, autocorrelação, assimetria e coeficiente de variação. Os indicadores se baseiam em mudanças ocorridas nos dados que compõem as séries temporais. As séries temporais por sua vez são extraídas do fluxo de dados da rede e submetidas aos indicadores estatísticos. Com base nas características mencionadas o sistema STARK emite alertas sobre a iminência de sobrecargas na rede (ex. enlaces ou servidores) para que sejam tomadas ações preventivas na rede e assim evitem maiores efeitos dos ataques. Desta forma, a Seção 4.1 apresenta uma visão geral do sistema e seus principais aspectos. A Seção 4.2 descreve detalhadamente o sistema STARK, suas características, a dinâmica do comportamento dos indicadores em relação aos ataques DDoS e o funcionamento do sistema como um todo. A Seção 4.3 apresenta o resumo do capítulo.

4.1 Visão Geral

O sistema STARK destaca-se por identificar tendências de aproximação de ataques DDoS com base no comportamento de indicadores estatísticos genéricos, ou seja, sem o conhecimento prévio do estado da rede. Por meio da antecipação de um ataque DDoS, os artefatos de monitoramento da rede, como exemplos IDS e IPS, podem disparar gatilhos para ajustes das configurações em ferramentas específicas como *firewall* a fim de parametrizar a rede para contingenciar, evitar ou conter o ataque. Desta forma, o sistema STARK realiza a predição dos ataques através de etapas que resumidamente consistem da medição e preparação dos dados, o cálculo dos indicadores estatísticos, a análise dos indicadores e a emissão de alertas. As etapas foram sistematizadas de modo que o sistema STARK receba como entrada os dados do fluxo da rede de maneira que os mesmos são manipulados e tratados para a compor às séries temporais. Estas são submetidas aos indicadores estatísticos. De acordo com o comportamento resultante dos dados é possível identificar a tendência de aproximação de ataques DDoS.

4.2 Detalhamento do Sistema STARK

A Figura 4.1 ilustra o sistema e o posicionamento do sistema STARK na rede e as etapas de seu funcionamento. A decisão sobre o posicionamento de soluções de segurança em uma rede depende, em geral, dos requisitos e características da rede. Entretanto, o seu posicionamento pode implicar em maior ou menor eficiência na defesa do ambiente. Desta forma, a proposta

de posicionamento considera os objetivos do sistema STARK, sendo este posicionado entre o roteador de borda e o sistema de *firewall*. Assume-se a existência de um *hardware* dedicado para o sistema STARK, porém o mesmo é flexível para ser implementado junto ao *firewall* ou ao sistema de detecção de instrução (da sigla em inglês IDS).

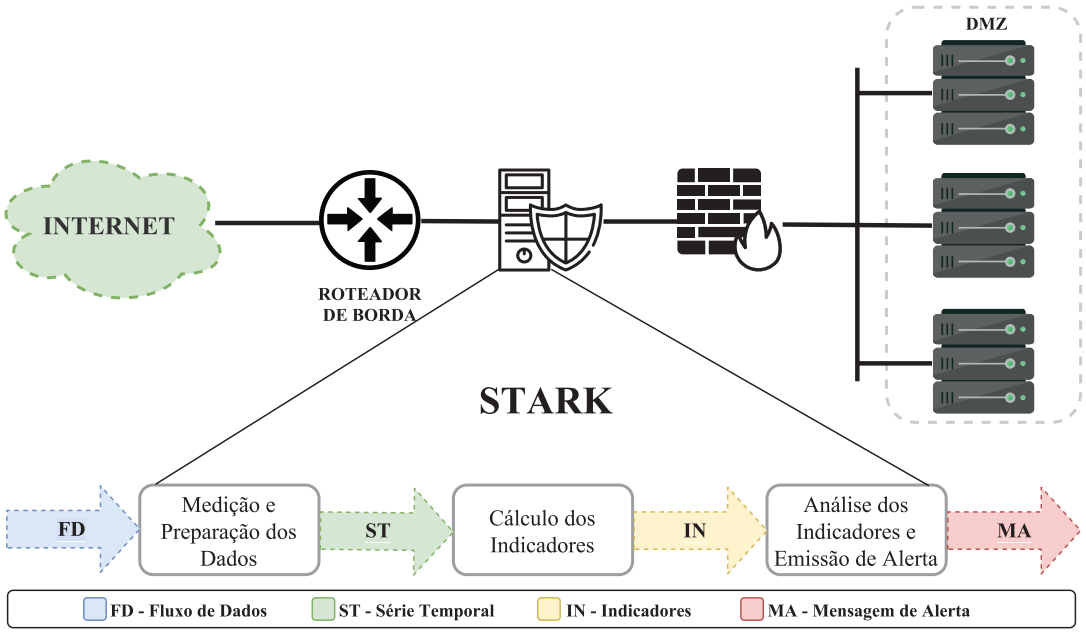

Figura 4.1: Posicionamento e etapas do sistema STARK

O sistema segue três etapas: (*i*) medições e preparação dos dados, (*ii*) cálculo dos indicadores estatísticos, e (*iii*) análise dos indicadores e emissão de alertas. A entrada do sistema consiste em dados de medições da rede contendo características (*features*), como por exemplo tamanho dos pacotes, número de requisições. Com base nas medições, as *features* relevantes para o cálculo dos indicadores são extraídas a fim de compor as séries temporais. A composição das séries temporais se dá através dos dados globais do fluxo da rede, ou seja, não se limitam de maneira específica ao fluxo proveniente da vítima, para a vítima, ou por um endereço IP e porta específicos. Após a composição das séries temporais, os valores dos indicadores são analisados.

4.2.1 Medições e Preparação dos Dados

Esta etapa engloba *(i) a coleta do fluxo de dados da rede*, *(ii) a definição do tamanho da janela de tempo* e *(iii) a filtragem dos dados* (extração da característica). Assume-se o funcionamento em modo promíscuo para a interface de rede do *hardware* em que o sistema estiver sendo executado. A coleta do fluxo de dados da rede ocorre continuamente através de ferramentas de monitoramento de rede, tais como tcpdump, tshark, wireshark.

Os dados coletados podem ser armazenados em uma base de dados sem ser formatados ou analisados, ou ainda podem ser analisados imediatamente. Este último é preferível, uma vez que a principal ideia do sistema STARK é prover resultados das predições online. As análises ocorrerão sobre janelas de coleta de tamanho N. O tamanho da janela de dados deve ser definido por tempo, por exemplo, X segundos (ou minutos), ou por quantidade de amostras, por exemplo, quantidade de pacotes coletados. Uma vez definido se a janela será por tempo ou quantidade de elementos, o seu tamanho é auto-ajustável. Com base nos dados contidos na janela de tamanho N , o sistema STARK efetua o processo de cálculo dos indicadores estatísticos. Desta forma, se

o sistema emitir a mensagem que indica ausência de recurso computacional para calcular sobre os dados que compõem a janela ou ainda que os dados contidos são insuficientes para alimentar os indicadores, o STARK descarta o conjunto de dados dessa janela e submete uma nova janela de tamanho $N+P$ ou $N-P$ (onde P é o valor percentual sobre N), ampliando ou reduzindo o tamanho da janela conforme o caso, auto-ajustando o valor de N.

Por fim, uma filtragem nos dados ocorrerá sobre cada janela. A filtragem dos dados percorre as amostras dos conjuntos de dados e extrai as características desejadas (ex. tamanho do pacote). A composição das séries temporais é realizada sobre a janela de tamanho N (ex. em segundos). As amostras da janela são utilizadas como entrada para a função F que aplica parâmetros específicos para extrair somente a característica desejada, neste experimento, o tamanho dos pacotes. A função F percorre esse conjunto de dados, de tamanho N (segundos) e evidencia o tamanho do pacote no respectivo tempo (T). Como saída, a função F gera uma estrutura (série temporal) em que indica a marcação do tempo T e o tamanho do pacote (ex. em *bytes*) correspondente para o dado instante. Essas séries temporais são as saídas da etapa de medição e preparação dos dados e servem de entrada para a etapa de cálculo dos indicadores estatísticos, tal como ilustra a Figura 4.1.

4.2.2 Cálculo dos Indicadores Estatísticos

Esta etapa do sistema STARK calcula os valores dos indicadores estatísticos. Cada série temporal serve de base para o cálculo dos valores dos quatro indicadores estatísticos utilizados nesta pesquisa: *taxa de retorno, autocorrelação, coeficiente de variação e assimetria*. Esta subseção detalha esses indicadores e como são realizados os cálculos, além de demonstrar o comportamento esperado que prediz os ataques DDoS.

Para o cálculo dos indicadores estatísticos, o sistema STARK tem como base as equações expostas na Seção 2.2.2 para os respectivos indicadores. O processo de cálculo tem início com a conversão dos valores observados na série temporal em um vetor numérico (V). Adicionalmente, o sistema utiliza um parâmetro W para determinar uma janela de rolamento. Esta janela W é expressa em porcentagem. A janela de rolamento indica uma determinada fração do conjunto de dados contido no vetor V que é utilizado como carga para o início do cálculo dos respectivos indicadores. Em geral, é utilizado como valor padrão para a janela de rolamento (W) cinquenta por cento das amostras contidas no vetor. Dessa forma, o sistema submete as amostras contidas no vetor V e o valor determinado pela janela W a cada um dos indicadores estatísticos. Como resultado desta etapa, o sistema apresenta a tendência de comportamento de cada um dos indicadores. Esse comportamento resultante indica se a tendência do indicador é crescente ou decrescente, bem como a intensidade desta tendência por meio do respectivo $Kendall\ tau$ do indicador estatístico. Esses parâmetros são representados por kTR ($Kendall\ tau$ da taxa de retorno), kAC ($Kendall\ tau$ da autocorrelação), kCV ($Kendall\ tau$ do coeficiente de variação), kAS ($Kendall\ tau$ da assimetria).

Conforme a Seção 2.2.2, quando a taxa de retorno é crescente e a autocorrelação, o coeficiente de variação e a assimetria são decrescentes, indica-se a probabilidade da rede se recuperar de uma perturbação e se manter no estado metaestável em que se encontra (ou seja, retornar à estabilidade). Já, se a taxa de retorno apresenta tendência decrescente e a autocorrelação, o coeficiente de variação e assimetria exibem tendências de crescimento, e quando avaliados em conjunto, indicam a aproximação da transição crítica, ou seja, a mudança do estado metaestável para outro estado. Isso demonstra a iminência de um ataque DDoS.

Entre os resultados objetivos desta etapa estão o comportamento de cada um dos indicadores avaliados, bem como seus respectivos coeficientes de correlação conhecido como

Kendall tau. A partir dessas informações a etapa de análise dos indicadores e emissão de alerta, descrita na Subseção 4.2.3, tem os subsídios necessários para realizar a análise e avaliar sobre a emissão ou não de um sinal de alerta.

4.2.3 Análise dos Indicadores e Emissão de Alerta

Esta etapa do sistema STARK toma como entrada os valores dos indicadores estatísticos calculados para cada série temporal e provê como saída alertas sobre a predição dos ataques DDoS, em caso positivo. Com base nos valores dos indicadores, esta etapa analisa o comportamento ao longo do tempo. Os valores calculados para cada indicador podem ser associados à sua representação gráfica. Os quatro indicadores e seus comportamentos devem ser analisados em conjunto. Para cada indicador e série temporal, o coeficiente de intensidade conhecido como *Kendall tau* é calculado para quantificar a força da tendência do seu comportamento Dakos et al. (2012). Os resultados desse coeficiente variam entre -1 e +1. Dessa forma, os valores próximos ou maiores que -0.7 e 0.7 indicam uma forte tendência na intensidade do indicador Dakos et al. (2012).

Conforme demonstrado em Scheffer et al. (2009), uma transição crítica pode ser prevista através da análise conjunta destes quatro indicadores estatísticos. Um comportamento específico caracteriza esses quatro indicadores na iminência de uma transição crítica, (*i*) em que a taxa de retorno tende a reduzir, (*ii*) a autocorrelação e o coeficiente de variação tendem a aumentar, e (*iii*) a assimetria pode aumentar ou diminuir. Essas três condições precisam ser verificadas para se considerar a iminência de uma transição crítica, neste caso proveniente de um ataque DDoS conhecido ou desconhecido.

Sabendo desta caracterização para os comportamentos dos valores dos indicadores estatísticos na iminência de uma transição crítica, esta etapa do sistema STARK analisa as tendências nos valores dos indicadores calculados para cada série temporal e emite um alerta em caso positivo da predição. Para identificar as tendências, a etapa toma como referência o valor do *Kendall tau* de cada indicador. Para isso, foi definido um limiar para a análise das tendências. Se as três condições de comportamento descritas acima forem observadas, um alerta de predição de ataque será gerado. Como forma de definir um limiar para disparar o alerta, foi utilizada a função FL (Função Limiar) que tem como entrada o *Kendall tau kTR* (*Kendall tau* da taxa de retorno), *kAC* (*Kendall tau* da autocorrelação), *kCV* (*Kendall tau* do coeficiente de variação) e *kAS* (*Kendall tau* da assimetria) resultante dos quatro indicadores avaliados e o número total de indicadores (*nTI*). A FL realiza o cálculo descrito na Equação 4.1. Quando a saída da FL for maior ou igual a um determinado valor, aqui representado por L (Limiar), o alerta de predição de ataque será gerado e enviado.

$$FL = \frac{-(kTR) + (kAC) + (kCV) + \sqrt{(kAS^2)}}{nTI} \tag{4.1}$$

4.3 Resumo

Este capítulo apresentou o sistema de predição de ataques DDoS denominado STARK, detalhando suas características e comportamento. Adicionalmente, também demonstrou o funcionamento das três etapas do sistema que são: (*i*) medições e preparação dos dados, (*ii*) cálculos dos indicadores estatísticos e (*iii*) análise dos indicadores e emissão de alerta. Sendo que a primeira corresponde ao processo de manipulação dos dados, a segunda está relacionada

a dinâmica do comportamento dos indicadores estatísticos, e a terceira refere-se ao processo de análise do comportamento resultante dos indicadores e emissão do alerta antecipadamente, indicando a predição do ataque DDoS.

5 Avaliação

Este capítulo expõe a metodologia adotada durante o desenvolvimento desta pesquisa, além de apresentar e detalhar as características dos conjuntos de dados utilizados. Foram listadas as ferramentas aplicadas no tratamento e manipulação dos dados juntamente com a dinâmica do processo realizado. Além disso, o capítulo apresenta as análises realizadas sobre os conjuntos de dados avaliados, juntamente com os resultados obtidos, demonstrando a predição de ataques DDoS. E, adicionalmente, é exibida uma discussão sobre os resultados alcançados com a finalidade de facilitar a compreensão da aplicação dos indicadores estatísticos e a predição de transições críticas, que neste caso, aponta a iminência do ataque DDoS.

5.1 Metodologia

A avaliação do sistema STARK passou pela seleção do conjunto de dados e análise da predição dos ataques em estágios iniciais. O sistema é projetado para funcionar online mas, para fins de avaliação a abordagem seguida é orientada a traços, devido ao melhor controle no cenário de avaliação, além de poder usar dados contendo registros de ataques reais. Assim, em um primeiro momento, comparou-se o uso de diferentes conjuntos de dados. Em geral, as medições desses dados foram realizadas pela ferramenta TCPdump em conjunto com tshark e os conjuntos de dados foram gerados contendo dados brutos sobre os fluxos da rede. Esses conjuntos de dados são disponibilizados pelo CAIDA (*Center for Applied Internet Data Analysis*), CTU (*Czech Technical University*) e DARPA (*Defense Advanced Research Projects Agency*). A motivação para o uso de tais conjuntos de dados se deve aos seguintes fatores: (*i*) conterem ataques DDoS rotulados, o que possibilita melhor compreensão dos dados para análise e conclusões; (*ii*) utilizarem o padrão pcap, empregado por muitas ferramentas de redes; (*iii*) amplo uso desses conjuntos de dados em outras pesquisas da literatura, permitindo verificações; e (*iv*) serem conjuntos de dados disponíveis na Internet que permitem a reprodução dos resultados.

Esses conjuntos de dados são a base para a extração de séries temporais, tendo por referência o tamanho dos pacotes trafegados na rede. Assim, as séries possuem a marcação do tempo e, associado a isso o valor do tamanho do pacote. A marcação do tempo foi normalizada para o intervalo aberto de 0 a 60 segundos ou de 0 a 10 segundos (no caso dos dados disponibilizados pela CTU). Esse intervalo também é chamado de janela de tempo. Neste estudo, refere-se a pacotes na camada de enlace e a informação do tamanho foi extraída dos cabeçalhos dos pacotes através de comandos oferecidos para tratar arquivos do tipo pcap. Nas séries temporais não se faz distinção entre origem e destino dos pacotes, o comportamento do fluxo da rede foi analisado apenas sob a perspectiva da *feature* tamanho do pacote.

Para o cálculo dos indicadores estatísticos (taxa de retorno, autocorrelação, coeficiente de variação e assimetria), que é a base para a predição dos ataques, emprega-se a biblioteca *Early Warning Signals (EWS)* implementada em R. As séries temporais compostas pelos tamanhos dos pacotes servem de entrada para o cálculo dos valores dos indicadores, os quais são calculados

para cada janela de tempo. Seguindo o padrão recomendado da EWS, a curva resultante dos indicadores engloba 50% da janela Dakos et al. (2012). Após o cálculo dos valores dos indicadores, uma análise de tendência é realizada tomando como referência o *Kendall tau*. Caso um ataque DDoS seja previsto, um alerta é emitido, sendo, neste trabalho uma mensagem enviada para os sistemas de detecção ou mitigação e/ou ao administrador da rede. Está fora do escopo deste trabalho analisar o envio, a transmissão e a garantia de entrega dos alertas.

5.2 Características das Bases de Dados

Nesta seção são apresentadas as características dos conjuntos de dados utilizados neste trabalho, como: volume de dados, duração e tempo em que ocorre o ataque DDoS. O **conjunto de dados 1** CAIDA (2007) possui aproximadamente uma hora de registros (20:50:08 UTC a 21:56:16 UTC) de fluxo de dados coletados da rede. Neste, os dados estão distribuídos em três subconjuntos, *all-victim*, *to-victim* e *from-victim*. Utiliza-se o subconjunto *all-victim*, pois compreende todo o fluxo de entrada e saída da vítima. De acordo com a documentação da CAIDA, o ataque teve início por volta das 21:13, quando a carga da rede aumenta em poucos minutos de uma taxa perto de 200 kbits/s para cerca de 80 Mbits/s. Desta forma, entende-se aqui por início do ataque o momento em que uma sobrecarga é percebida no servidor. Além disso, o tamanho dos pacotes de dados oscila de 48 bytes até 1500 bytes. O sistema STARK procura predizer o ataque antes do início dessa sobrecarga.

O **conjunto de dados 2** possui tráfego de *botnet* capturado na CTU, em 2011 García e Uhlir (2011). Este conjunto de dados oferece uma grande quantidade de tráfego real de *botnet* misturado com tráfego normal e tráfego de *background*. Ele consiste em dados coletados considerando 13 cenários diferentes. Em cada cenário foi executado um *malware* específico que usou ao mesmo tempo vários protocolos de rede. Entre os cenários disponibilizados, o cenário 4 é utilizado neste trabalho, pois contém traços de ataques ICMP e UDP *Flooding*. O conjunto de dados é registrado em um arquivo de tamanho de 55 GB com aproximadamente quatro horas de gravação (11:00 às 15:11 horas). A sobrecarga do ataque DDoS teve início em torno das 12:21 e término às 13:06 da marcação de tempo nos traços, havendo uma alteração significativa no tamanho dos pacotes que oscilaram entre 60 e 1514 bytes.

O **conjunto de dados 3** disponibilizado pela DARPA Laboratory (2000) apresenta três horas (09:21 às 12:35 - EST) de registro de fluxo de dados. No arquivo de tamanho total de 111 MB há registro de um ataque de negação de serviço distribuído executado por um invasor. O ataque segue varreduras de vulnerabilidades, invasão, instalação do *malware* e execução do ataque DDoS. O início da sobrecarga do ataque ocorreu em torno das 11:29 horas, apresentando grande variação no tamanho dos pacotes (oscilação entre 60 e 1514 bytes) e aumento no fluxo de pacotes na rede. As características apresentadas dos conjuntos de dados utilizados neste estudo está disponível de forma resumida na Tabela 5.1.

Tabela 5.1: Características dos Conjuntos de Dados

Dataset	Duração	Tamanho	Ataque	Sobrecarga	Tamanho Pacotes
CAIDA	20:50 às 21:56 UTC	21 GB	ICMP *Flooding*	21:13	48 a 1500 bytes
CTU	11:00 às 15:11	55 GB	ICMP e UDP *Flooding*	12:21	60 a 1514 bytes
DARPA	09:21 às 12:35 EST	111 MB	ICMP	11:29	60 a 1514 bytes

Após a pesquisa e seleção dos conjuntos de dados listados o trabalho desenvolveu-se sobre o os traços de fluxo de rede, considerando ferramentas e formas de realizar a extração e manipulação dos dados com a finalidade de compor as séries temporais, descrito na Seção 5.3.

5.3 Extração das Séries Temporais

Devido ao tamanho dos arquivos referentes aos conjuntos de dados, estes foram fracionados em arquivos menores, a fim de facilitar a extração da *feature* tamanho dos pacotes e compor as séries temporais. Os arquivos resultantes correspondem à janelas de tempo de 10 ou 60 segundos. As janelas foram assim dimensionadas pois em situação real espera-se poder analisar online esses dados, e um dos aspectos avaliados é a influencia do tamanho da janela nos resultados de predição. Para criação dessas séries temporais, foi extraída a característica do tamanho do pacote *versus* o tempo (em segundos), sendo essa uma característica afetada em grande proporção nos ataques DDoS. Dessa forma, cada conjunto de dados gera várias séries temporais que são avaliadas individualmente. Cada série é submetida ao cálculo dos indicadores estatísticos utilizando a biblioteca EWS do R. A força da tendência do comportamento dos indicadores também é calculada. O coeficiente de correlação *Kendall tau* quantifica essa tendência. Dessa maneira, os quatro indicadores possuem um resultado relacionado ao seu respectivo *Kendall tau*.

5.4 Manipulação dos Dados

A análise estatística exploratória e inferencial depende da manipulação massiva dos dados juntamente com a extração das séries temporais e do cálculo dos indicadores estatísticos. Em geral, estes processos são custosos em termos computacionais e por consequência também em tempo. Inicialmente, os manipuladores e extratores (*scripts*) implementados utilizavam mono instruções que normalmente atribuem a apenas um núcleo do processador a tarefa solicitada, mantendo assim os demais núcleos do *hardware* subutilizados. Desta forma, o custo computacional e de tempo tornou o processo de experimentação bastante caro. Após identificar a demanda por desempenho, ainda durante os experimentos, os extratores e manipuladores sofreram refatoração no código com o objetivo principal de diminuir o custo computacional, principalmente em termos de tempo consumido, e otimizar o uso do processador. A refatoração concentrou-se no aspecto de maximizar o uso dos núcleos do *hardware*. A princípio foram reavaliadas todas as etapas do processo de manipulação e extração realizada. Resumidamente o processo de experimentação segue as seguintes etapas (i) fracionamento do conjunto de dados de acordo com a definição do tamanho de janela (N), (ii) varredura os pacotes resultantes do fracionamento e extração do tamanho do pacote (*feature* avaliada) compondo as séries temporais e (iii) submissão das séries temporais contendo o tamanho dos pacotes aos indicadores estatísticos. O tempo consumido para realizar a segunda etapa deste grupo de tarefas, em geral, chega a consumir horas, variando conforme o volume de dados. Dessa forma, a fim de viabilizar o avanço da pesquisa em tempo hábil, a refatoração permitiu diminuir o tempo consumido na segunda etapa em torno de 60 a 70% utilizando a paralelização através da atribuição de uma instrução de extração da característica para cada núcleo do *hardware* utilizado simultaneamente. Este é um dado empírico e sem a pretensão de que seja validado ou confirmado neste estudo pois está fora de escopo neste momento. Contudo, o relato é considerado por que demanda o desenvolvimento de conhecimentos para otimizar o uso do *hardware* em prol da pesquisa desenvolvida, e, assim, viabilizar o andamento desta bem como facilitar e otimizar a realização de futuros e novos

experimentos. Como resultado, a refatoração permitiu paralelizar a extração da característica avaliada pelos indicadores, bem como identificar outros pontos possíveis de serem otimizados. A paralelização da etapa de extração se deu através do uso do GNU Parallel Tange (2011).

Outro desafio que se fez presente foi a análise exploratória dos dados. Esta por sua vez demanda visualizar os dados sob diferentes aspectos, como exemplo, médias, medianas, modas e sumarizações e como resultado plotar gráficos para fins de identificação, análise e validação do comportamento dos dados. Esse tratamento possibilita a verificação dos resultados identificados junto aos indicadores estatísticos aos quais as séries temporais foram submetidas. A demanda é identificada em momento posterior às três etapas anteriores, pois a necessidade de visualizar os dados sob outros aspectos também apresenta questões de desempenho em que foi necessário um tratamento cuidadoso na manipulação dos dados por meio de fracionamento prévio dos dados e posteriormente sua manipulação através do R.

5.5 Resultados e Análises

Este capítulo demonstra os resultados obtidos na avaliação de desempenho do sistema STARK. São apresentados o comportamento dos indicadores estatísticos relacionando-os à teoria da metaestabilidade e o comportamento esperado que identifica a aproximação da transição crítica, ou seja, a iminência de ataques DDoS. Também, é exposto o comportamento contrário dos respectivos indicadores, quando não indica a tendência de transição crítica. Adicionalmente, é apresentada a análise exploratória que demonstra as características estatísticas dos dados contidos nos conjuntos avaliados. O processo de análise tem início pelo conjunto de dados 1 (CAIDA), seguido pelos conjuntos 2 (CTU) e 3 (DARPA). Os gráficos gerados foram obtidos a partir da aplicação dos indicadores estatísticos sobre as séries temporais, ilustradas nas próximas seções.

5.5.1 Resultados do Conjunto de Dados CAIDA

As Figuras 5.1 e 5.2 ilustram duas situações nos resultados obtidos sobre as séries temporais de janelas de tempo de 60 segundos criadas a partir do conjunto de dados da CAIDA. Sobre as séries temporais, o comportamento dos indicadores taxa de retorno, autocorrelação, coeficiente de variação e assimetria são apresentados. A Figura 5.1 demonstra os resultados dos indicadores calculados sobre um série temporal extraída do intervalo entre 20:50:36 às 20:51:36. Nesta série, observa-se que os indicadores apresentam o comportamento esperado para indicar a iminência de um ataque DDoS. Esse comportamento engloba uma queda na curva da taxa de retorno, enquanto mostra um aumento na curva da autocorrelação, coeficiente de variação e assimetria. A queda na taxa de retorno revela a presença de oscilações significativas no fluxo de dados da rede, apresentando assim uma grande dificuldade de se manter em um estado metaestável. A propensão de forte incremento na autocorrelação (*Kendall tau* positivo de 0.764) indica similaridade nos valores do tamanho dos pacotes ao longo do tempo. Isso significa que há uma forte tendência dos pacotes de dados permanecerem com tamanho em torno de 1500 bytes. O aumento na curva do coeficiente de variação revela forte instabilidade na rede devido à presença de valores extremos e à alta variação no tamanho dos pacotes. Por fim, a assimetria positiva, com forte tendência de crescimento, aponta uma significativa concentração do tamanho de pacotes com valores em torno de 1500 bytes.

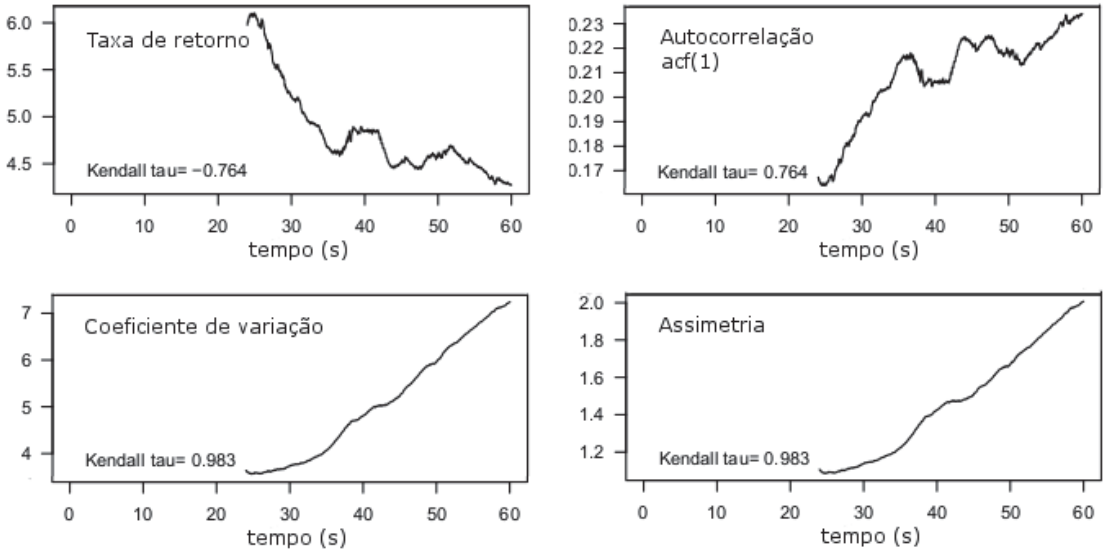

Figura 5.1: Comportamento dos indicadores estatísticos no momento prévio ao ataque DDoS sob o conjunto de dados do CAIDA

Ao considerar o conjunto de comportamento dos indicadores e as características do conjunto de dados analisados, é possível afirmar a indicação da aproximação de uma transição crítica, neste caso representando a aproximação de um ataque DDoS. Dessa forma, o sistema STARK, nesta base de dados, foi capaz de predizer o ataque DDoS com 23 minutos de antecedência do início da sobrecarga gerada pelo ataque (que ocorreu às 21:13). Além disso, o sistema STARK também apontou a iminência de ataque em outra amostra das séries temporais. A janela de tempo que compôs a série temporal pertencente ao intervalo entre 21:00:36 às 21:01:36 apontou a iminência do ataque, ou seja, emitiu o alerta com 13 minutos de antecedência com relação ao tempo de início da sobrecarga da rede (21:13).

Resultado Inverso dos Indicadores

Para auxiliar na compreensão do comportamento dos indicadores estatísticos, a Figura 5.2 ilustra o comportamento em que **NÃO** é identificada a predição de uma transição crítica, ou seja, não aponta a aproximação de um ataque DDoS na série temporal. Esta figura ilustra o comportamento identificado quando não há indícios de ataque na rede. Ela foi incluída aqui para ilustrar uma situação em que os indicadores não seguem o comportamento que indica a iminência de uma transição crítica, ou seja, quando os indicadores apresentam comportamento inverso ao esperado. Esses resultados foram calculados da série temporal do período das 20:53:36 às 20:54:36. A taxa de retorno apresentou uma tendência de crescimento, com seu coeficiente de intensidade (*Kendall tau*) em 0.712 positivo. Isso mostra pouca variação nos dados, mantendo o seu estado metaestável. A autocorrelação (*Kendall tau* -0.711) em decréscimo indica que não há valores consecutivos semelhantes, não havendo comportamento genérico. O coeficiente de variação (*Kendall tau* -0.570) em decréscimo indica a existência de pouca variação nos dados, portanto demonstra estabilidade na rede. Já a assimetria se mostra crescente (*Kendall tau* 0.592), indicando o deslocamento das médias de tamanho dos pacotes, o que apresenta alguma variação nos dados, porém não o suficiente para apontar a tendência de aproximação crítica visto que os demais indicadores tem comportamento adverso ao esperado. Desse modo, nessa série temporal é possível observar a estabilidade da rede, diante das poucas perturbações ou variações, o que demonstra o suporte do metaestado atual.

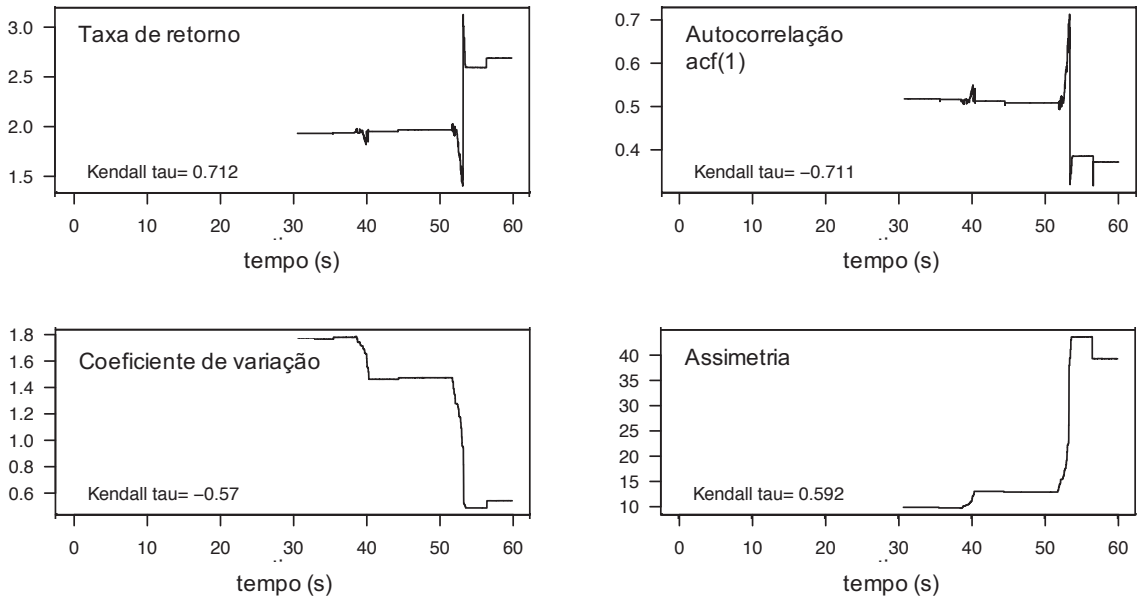

Figura 5.2: Comportamento inverso dos indicadores do CAIDA

Análise dos Dados

Esta análise tem como premissa observar as amostras em toda a extensão do conjunto de dados e assim possibilitar a visualização do comportamento geral na rede. O conjunto contém 3955 segundos de fluxo de dados registrados. Além do fluxo de rede, neste conjunto de dados está documentado um ataque DDoS às 21:13. Conforme pode ser visto na Figura 5.3, indicado no tempo $t3$, o volume de pacotes sofre ampliação significativa passando de quarenta e quatro mil (44k) para mais de um milhão e quinhentos mil (1.567k) pacotes. Os dados mostram um avanço de 12 MBytes para 107 MBytes. Essa evolução ocorreu ao longo de 60 segundos sobrecarregando a rede neste curto espaço de tempo. Nos períodos de tempo de 20:50:36 à 20:51:36 e de 21:00:36 à 21:01:36, ambos séries temporais onde o STARK emitiu o alerta, houve significativa diversidade no tamanho de pacotes variando de 60 à 1500 bytes. Isso é possível verificar na Figura 5.3 que ilustra o tempo (em segundos) *versus* as médias dos tamanhos dos pacotes de rede. O comportamento apresentado ilustra de forma clara toda a extensão do fluxo da rede. Os apontamentos indicados em $t1$ e $t2$ apontam as séries temporais em que a predição foi realizada. Nestas séries temporais são apresentadas as tendências da aproximação de transições críticas. Dessa forma, o sistema STARK emite alerta indicando a iminência de um ataque DDoS na rede. A variação no tamanho dos pacotes, exibidas instantes antes dos tempos $t1$ e $t2$ expostos na Figura 5.3, justifica o comportamento dos indicadores estatísticos genéricos que demonstram a aproximação de uma mudança de estado na rede resultante das etapas de preparação do ataque.

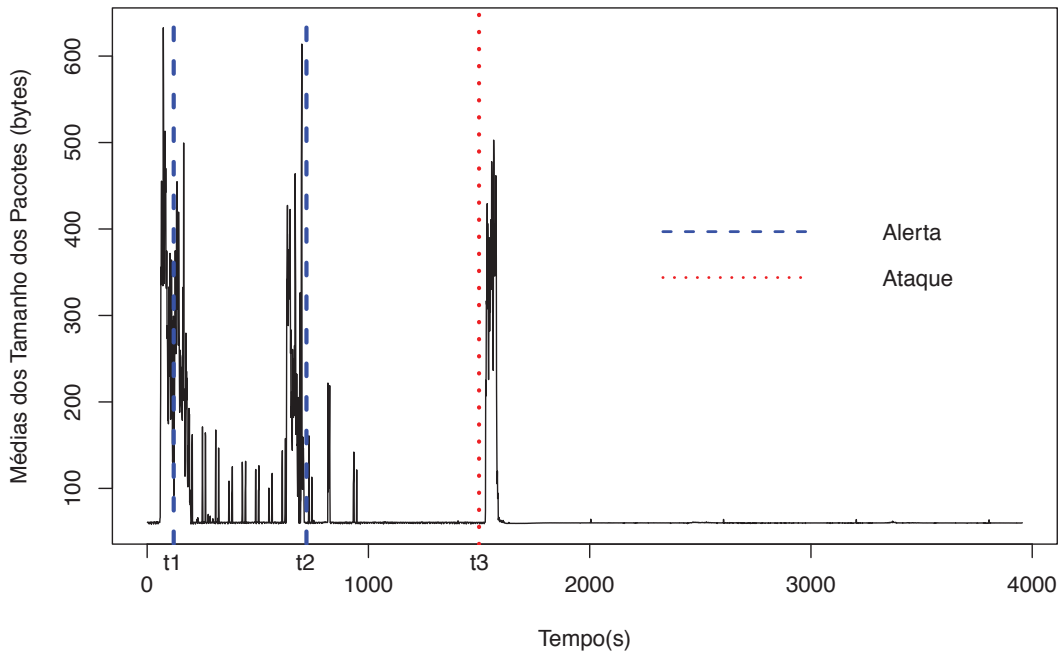

Figura 5.3: Médias dos tamanhos dos pacotes *versus* tempo

Com base no comportamento das médias dos tamanhos dos pacotes e nos apontamentos de tempo $t1$, $t2$ e $t3$ ilustrados na Figura 5.3, é possível observar a variação do tamanho dos pacotes em toda a extensão do traço da rede. Essa instabilidade repentina expõe as oscilações inesperadas no fluxo da rede. Dessa forma, o sistema STARK se mostrou capaz de predizer a iminência do ataques DDoS contido no conjunto de dados, com antecedência de 23 minutos em $t1$ e de 13 minutos em $t2$, em relação ao instante ($t3$) da sobrecarga da rede, que é uma consequência resultante do ataque DDoS.

5.5.2 Resultados do Conjunto de Dados CTU

Os resultados apresentados na Figura 5.4 mostram o comportamento dos indicadores estatísticos aplicados sobre a série temporal extraída do conjunto de dados disponibilizado pela CTU. Esta série temporal possui o registro de traços da rede no período de 11:00:05 à 15:11:19. Nesta série temporal, há alteração significativa no tamanho dos pacotes variando entre 60 e 1514 bytes. Devido à alta variação dos dados, os indicadores estatísticos apresentaram o comportamento esperado para a predição. A autocorrelação aumentou significativamente, indicando forte tendência na manutenção do tamanho dos pacotes de dados permanecer com valores em torno de 1500 bytes e assim ocorrer o evento crítico. A diminuição na taxa de retorno confirma que o fluxo de rede sofreu oscilações severas. O coeficiente de variação apresenta significativa tendência de crescimento. Este comportamento indica instabilidade no estado da rede, ocasionado pela variação no tamanho dos pacotes de 60 a 1514 bytes. Além disso, a variação no tamanho dos pacotes justifica o aumento na curva da assimetria da distribuição dos dados, apontando que há uma alta concentração de pacotes grandes (em torno de 1500 bytes), o que indica a tendência do estado da rede permanecer em torno de um estado crítico nas observações seguintes. Dessa forma, é possível identificar a iminente aproximação do ataque

DDoS, com base no comportamento dos indicadores da Figura 5.4. Logo, cabe observar que a série temporal apresentada compreende o período de 11:03:25 às 11:03:35 e o ataque tem início por volta de 12:21. Portanto, neste conjunto de dados é possível verificar que o comportamento dos indicadores estatísticos apontam a tendência de um ataque DDoS com antecedência de aproximadamente uma hora e dezoito minutos.

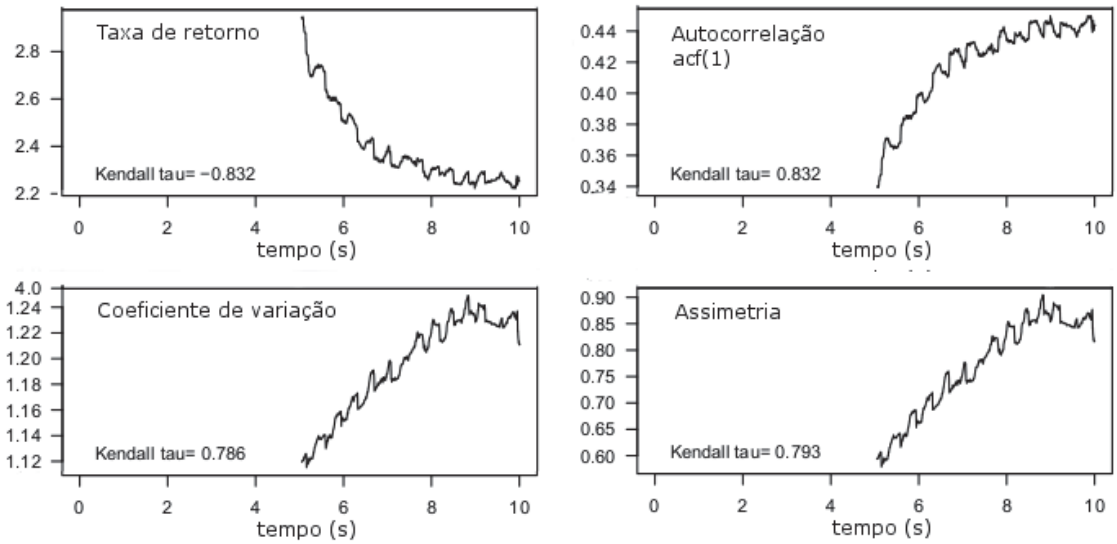

Figura 5.4: Comportamento dos indicadores estatísticos no momento prévio ao ataque DDoS sob o conjunto de dados da CTU

Resultado Inverso dos Indicadores

As demais séries temporais avaliadas neste conjunto de dados (CTU), exibem um comportamento adverso ao esperado para a predição de ataques DDoS. A Figura 5.5 ilustra o comportamento inverso ao esperado, ou seja, que não emite o alerta de predição. Com a finalidade de demonstrar esta situação, foi selecionada arbitrariamente a primeira série temporal deste conjunto. Assim, estes resultados são projetados com base na série temporal que compreende o período de 11:00:05 às 11:00:15. A taxa de retorno apresentou uma predisposição de crescimento, com o seu coeficiente de intensidade (*Kendall tau*) em 0.590 positivo. Desta forma, mostra pouca variação nos dados, mantendo o seu estado metaestável. A autocorrelação (*Kendall tau* -0.590) e o coeficiente de variação (*Kendall tau* -0.503) em queda indicam pouca similaridade entre as amostras dos dados e a existência de pouca variação, portanto indica estabilidade na rede. Já a assimetria se mostra crescente (*Kendall tau* 0.475), o que indica o deslocamento das médias de tamanho dos pacotes. Dessa forma, a série temporal mantém seu metaestado atual por meio da estabilidade da rede, devido a poucas perturbações ou variações.

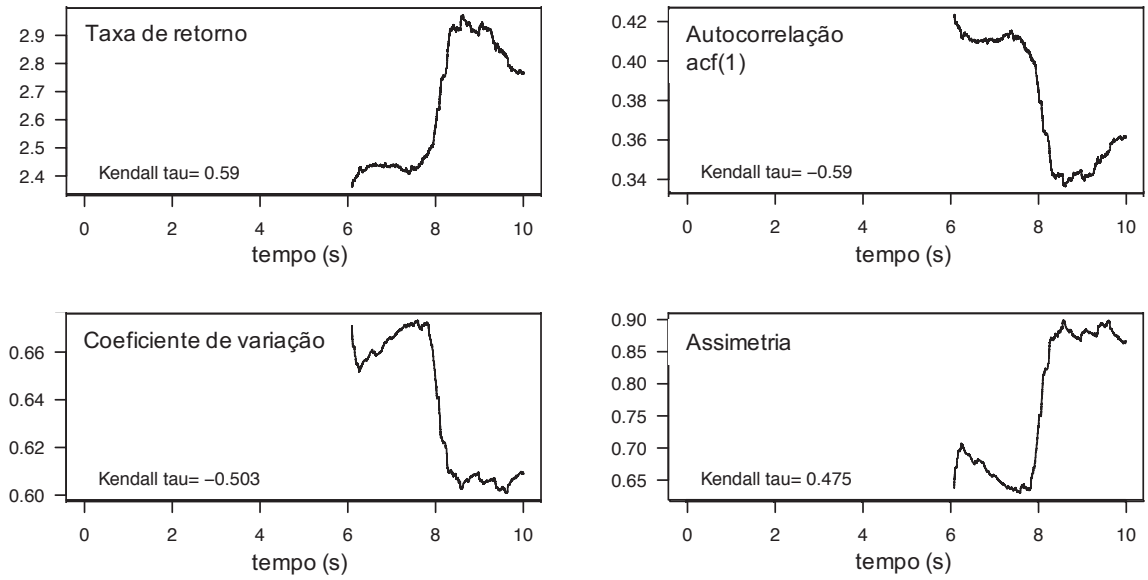

Figura 5.5: Comportamento inverso dos indicadores da CTU

Análise dos Dados

A Figura 5.6 corresponde à série temporal de todo o fluxo da rede disponibilizada pela CTU. Este traço é composto por 15.074 segundos, totalizando aproximadamente 4 horas de fluxo de rede registrado. A série temporal exposta na Figura 5.6 apresenta o tempo (em segundos) *versus* as médias dos pacotes de rede. Neste conjunto de dados é possível observar nos tempos $t1$ e $t2$ os indicativos da predição do ataque e o instante em que o ataque é iniciado respectivamente. O instante de tempo $t1$ refere-se ao intervalo de tempo das 11:03:25 às 11:03:35. Neste intervalo de tempo, os indicadores estatísticos exibiram o comportamento esperado na predição do ataque DDoS, conforme demonstrado na Figura 5.4. No tempo $t2$, a Figura 5.6 expõe o início do ataque DDoS no tempo 12:21. A oscilação do tamanho dos pacotes, visualizado por meio das médias dos tamanho de pacotes justifica o comportamento nos indicadores estatísticos que demonstram a iminência da transição crítica, i.e., o ataque DDoS. As variações apresentadas corroboram o movimento exibido pelos indicadores estatístico genéricos avaliados, conforme Figura 5.4.

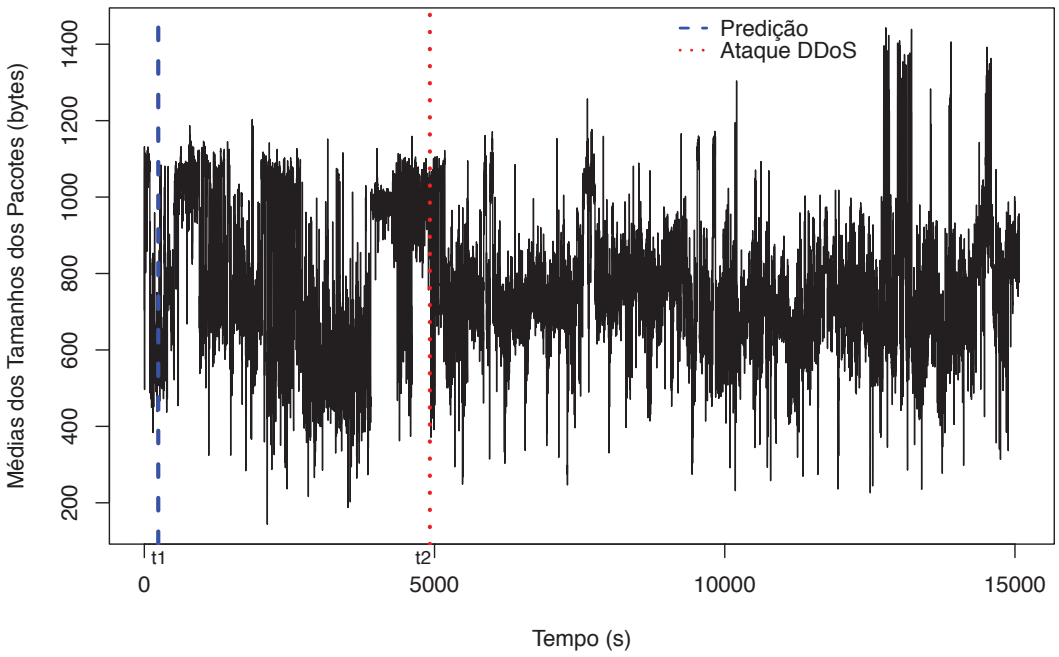

Figura 5.6: Médias dos tamanhos dos pacotes *versus* tempo

5.5.3 Resultados do Conjunto de Dados DARPA

No período de tempo das 09:29:36 às 09:30:36 foi observado o comportamento esperado dos indicadores estatísticos para a predição do ataque DDoS no conjunto de dados da DARPA. A Figura 5.7 ilustra o resultado dos quatro indicadores estatísticos obtidos para predição do ataque DDoS neste conjunto de dados. Observa-se uma queda na taxa de retorno, apontando a dificuldade do estado da rede em retornar à estabilidade após sofrer perturbações, como exemplo, um aumento no tamanho dos pacotes ou varreduras na rede. Ao analisar os indicadores em conjunto, observa-se um crescimento significativo da autocorrelação, revelando a tendência do tamanho dos pacotes permanecerem em torno de 1500 bytes. Quando analisado com os demais indicadores, esta tendência pode significar a aproximação de uma transição crítica, em direção ao ataque DDoS. O aumento do coeficiente de variação indica instabilidade no estado da rede, provocada pela grande oscilação no tamanho dos pacotes. Por fim, a curva da assimetria negativa aponta a presença de valores extremos e distantes dos valores considerados estáveis na distribuição das observações. No geral, a taxa de retorno apresenta comportamento decrescente enquanto os demais indicadores mostram comportamento crescente. Contudo, a assimetria é calculada em relação à média do tamanho dos pacotes, logo, o comportamento negativo e o positivo mostram que os valores da série temporal não se encontram em torno da média (valor estável). Dessa forma, o comportamento conjunto dos indicadores indica a aproximação de uma transição crítica, apontando a tendência de um ataque DDoS. A série temporal, cujos indicadores predizem o ataque DDoS está com duas horas de antecedência do início da sobrecarga da rede.

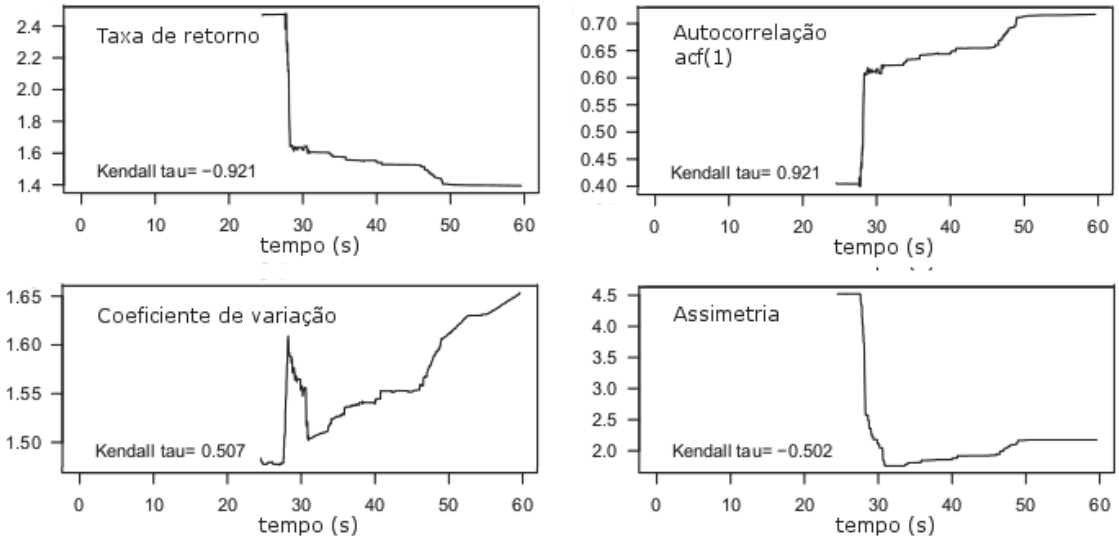

Figura 5.7: Comportamento dos indicadores estatísticos no momento prévio ao ataque DDoS sob o conjunto de dados da DARPA

Resultado Inverso dos Indicadores

Neste conjunto de dados observamos que as demais janelas de tempo avaliadas exibem um comportamento inverso ao comportamento esperado para a predição de ataques DDoS. Dessa forma a Figura 5.8 ilustra este comportamento, ou seja, esta janela de tempo não emite o alerta de predição. Este resultado é observado com base na série temporal que compreende o período de 09:21:36 às 09:22:36. Esta série temporal é a primeira de todo o conjunto e foi selecionada arbitrariamente. A taxa de retorno exibe uma tendência de crescimento, com o seu coeficiente de intensidade (*Kendall tau*) em -0.407 positivo. Desta forma, mostra resistência, mantendo o seu estado metaestável. A dinâmica do comportamento dos indicadores nesta série temporal mostra a autocorrelação (*Kendall tau* 0.356) e o coeficiente de variação (*Kendall tau* 0.220) decrescentes. Com base nisso observa-se pouca similaridade entre as amostras e a existência de pouca variação, portanto, indica estabilidade na rede. Já a assimetria se mostra decrescente e com valor de intensidade em *Kendall tau* -0.294, ou seja, uma assimetria pequena, que confirma junto com os demais indicadores a tendência da rede se manter estável.Dessa maneira, nessa série temporal é possível observar a manutenção do metaestado atual por meio da estabilidade da rede e mediante as poucas perturbações ou variações.

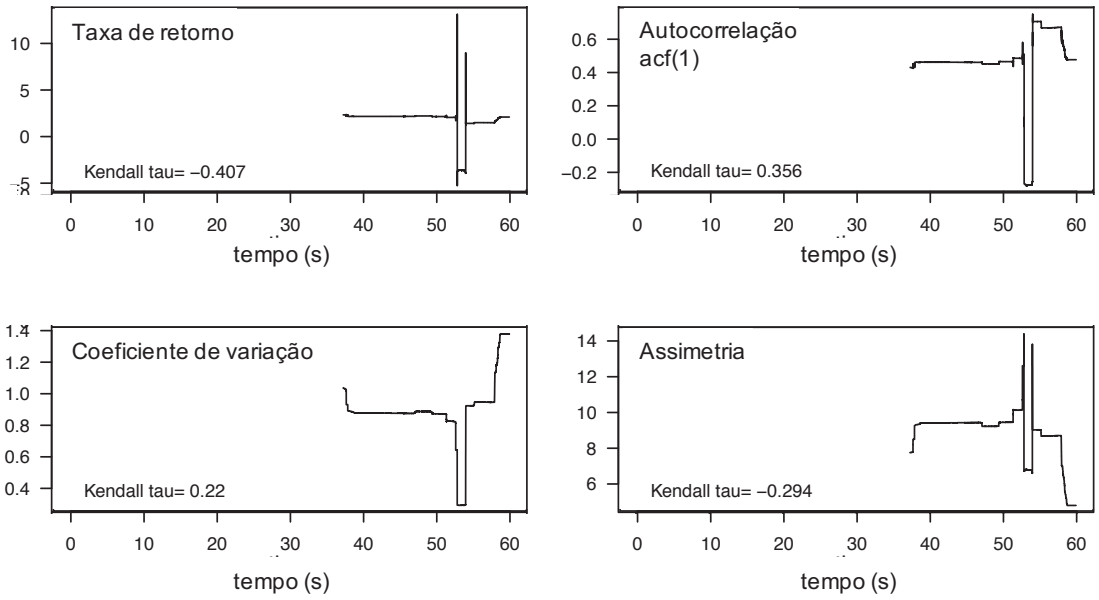

Figura 5.8: Comportamento inverso dos indicadores da DARPA

Análise dos Dados

O conjunto de dados compreende 11.652 segundos de fluxo, totalizando aproximadamente 3 horas de pacotes registrados. Neste, está documentado um ataque DDoS no tempo 11:29 representado pelo $t7$ onde são evidenciados os pacotes do ataque. A Figura 5.9 exibe as médias dos tamanhos dos pacotes em toda extensão do fluxo de dados da rede. O tamanho dos pacotes varia entre 60 e 1514 bytes. Os tempos identificados como $t1$ a $t6$ na figura 5.9 expõem as séries temporais em que os indicadores estatísticos mostraram a tendência de aproximação de uma transição crítica, ou seja, a iminência de um ataque DDoS na rede. A Figura 5.9 ilustra ainda oscilações nos dados instantes antes dos tempos $t1$ a $t6$. A documentação deste conjunto de dados aponta que existem processos de varreduras, instalação de ferramentas para explorar vulnerabilidades como *malwares* e *exploits* e a coordenação do ataque. Essa movimentação resulta em variações no fluxo da rede, e, neste caso, justifica as oscilações apresentadas pelos indicadores estatísticos apresentados na Figura 5.9 nos tempos $t1$ a $t6$ e também as tendências expostas por meio dos indicadores estatísticos na Figura 5.7.

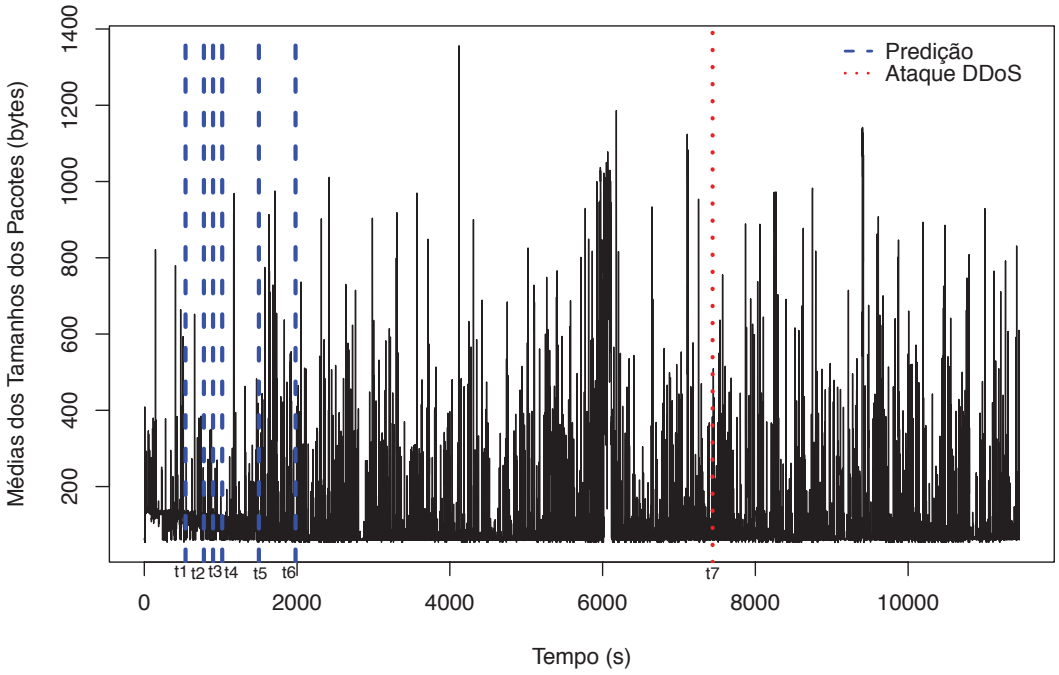

Figura 5.9: Médias dos tamanhos dos pacotes

5.6 Discussão

A análise dos conjuntos de dados em associação com os resultados apontados mostram que a proposta do sistema STARK é promissora e tem potencial de atuar de forma complementar às demais soluções que tem como objetivo a predição de ataques DDoS. Nos resultados, é possível observar que em diversos momentos prévios aos ataques DDoS o sistema STARK indica a aproximação da transição crítica que caracteriza uma futura ruptura na rede, ou seja, a sobrecarga resultante de um ataque DDoS. No conjunto de dados 1 (CAIDA), observou-se dois instantes em que o sistema STARK emite o alerta. O primeiro alerta foi emitido com 23 minutos de antecedência do ataque e o segundo com 13 minutos. Na série temporal extraída do conjunto de dados 2 (CTU) o sistema STARK indica a predição do ataque DDoS com antecedência de 78 minutos do momento do ataque. Já nas séries temporais extraídas do conjunto 3 (DARPA) são observadas a emissão de seis alertas sendo o primeiro com 120 minutos de antecipação e os demais com 116, 114, 112, 104 e 96 minutos respectivamente. Nessas séries temporais há variação no tamanho dos pacotes que justificam a presença de valores extremos (ex. 1500 bytes) e assim o comportamento exibido pelos indicadores estatísticos. As oscilações observadas ao longo das séries temporais estão relacionadas às etapas de preparação do ataque. Entre elas estão a comunicação entre atacantes, varreduras por vulnerabilidades, varreduras para reconhecimento da arquitetura da rede, instalação de *exploits*, entre outros.

A exposição de toda a série temporal de cada um dos conjuntos permite visualizar tais oscilações, bem como o instante dos respectivos ataques DDoS ao longo do fluxo registrado. A realização dessas análises trouxe à tona alguns desafios relacionados ao sistema proposto, entre eles a necessidade de tornar o tamanho da janela (N) dinâmico, conforme detalhado no Capítulo 4. O tamanho da janela precisa ser autoajustável devido (i) a variação do volume do

fluxo de dados da redes e (*ii*) o *warm-up* dos indicadores. A variação do volume de dados da rede impacta diretamente no desempenho do sistema e, portanto, no tempo de resposta. Como a proposta é que o sistema seja *online*, o autoajuste do tamanho da janela permite que o sistema STARK se adapte ao fluxo da rede. Ao submeter as séries temporais aos indicadores estatísticos, estes precisam dispor de um determinado número de ocorrências (tamanho de pacotes) para que seja possível efetuar os cálculos, e, assim, expor a tendência de comportamento de tais indicadores. Além disso, a autoadaptação considera aspectos como processamento e memória disponíveis.

A predição de ataques DDoS associado aos indicadores estatísticos genéricos está relacionada à oscilação do fluxo da rede, em especial na variação do tamanho dos pacotes. Por isso, alguns questionamentos são levantados. Entre as questões levantadas estão: (*i*) se um usuário legítimo fizer o *download* de arquivos significativamente grandes para o dimensionamento da rede, o sistema STARK responderá com um alerta, resultando em um falso positivo?, ou ainda (*ii*) um ataque DDoS *flash crowd* pode passar despercebido pelo sistema STARK, e assim, não gerar o alerta necessário criando um falso negativo? No caso em que muitos pacotes apresentaram o tamanho máximo em poucos instantes de tempo (ex. no caso de um *download*), os indicadores que medem a similaridade entre as amostras, como a autocorrelação, em poucos instantes exibirão esse comportamento, demonstrando assim que a variação abrupta no tamanho dos pacotes não indica a presença de um ataque DDoS. Portanto, como os indicadores estatísticos são avaliados em conjunto para a emissão do alerta, em geral, não será possível um atacante manipular todos os indicadores simultaneamente e assim burlar o sistema STARK. Ainda assim, cabe uma investigação mais profunda com a finalidade de identificar em qual proporção um *download* no ambiente de rede pode gerar oscilações no tamanho dos pacotes que faça parecer um ataque, ou ainda, que um ataque de fato não seja identificado pelo o STARK.

Com a aplicação dos indicadores estatísticos genéricos utilizados pelo sistema STARK para a predição de ataques DDoS não é preciso treinamento prévio, como ocorre com outras técnicas (ex. redes neurais). Essa característica é vantajosa porque evita o custo (ex. em tempo, capacidade de processamento) que em geral são consumidos para o treinamento de algoritmos usados em demais soluções. Contudo, ao mesmo tempo em que este aspecto é uma vantagem, impõe também novos desafios. Entre os desafios para o uso dos indicadores estatísticos está a dificuldade em avaliar a acurácia dos resultados apresentados. Assim, é necessário avaliar as abordagens disponíveis para verificação de falsos positivos ou falsos negativos quando dos resultados expostos.

5.7 Resumo

Este capítulo apresentou a metodologia aplicada para a realização da avaliação e verificação dos resultados apresentados com o uso dos indicadores estatísticos. A avaliação foi realizada com base nos conjuntos de dados disponiblizados pelo CAIDA, CTU e DARPA. Os dados foram descritos com a finalidade de facilitar a compreensão e visualização do uso dos indicadores, expondo assim suas principais características, permitindo assim identificar os pontos em que houve a predição do ataque, bem como o instante em que o ataque é apontado em cada um dos conjuntos. Para cada um dos conjuntos de dados foram demonstrados o comportamento esperado dos indicadores através dos resultados, a análise e a visualização das respectivas séries temporais. Adicionalmente, o capítulo expôs a discussão sobre os resultados apontando os avanços e os desafios ainda em estudo. Contudo, o sistema STARK se mostra capaz de indicar a aproximação da transição crítica, neste caso a iminência de um ataque DDoS.

6 Conclusões

Esta pesquisa propôs uma abordagem para predizer a iminência de ataques DDoS volumétricos na Internet por meio do sistema STARK. O sistema utiliza a identificação de sinais de aproximação de transições críticas em dados de fluxo de rede, a fim de antecipar a iminência de ataques DDoS. A predição ocorre com base na aplicação da teoria da metaestabilidade por meio dos indicadores estatísticos genéricos. As séries temporais são submetidas aos conjunto de indicadores estatísticos que exibem os sinais com base no comportamento exposto. O reconhecimento deste comportamento demonstra possibilidade de uma ruptura no fluxo da rede, ou seja, uma ataque DDoS. Foi realizado um levantamento da literatura relacionado ao tema com a finalidade de fundamentar e apresentar os conceitos relacionados a esta pesquisa envolvendo ataques DDoS, metaestabilidade, aprendizagem estatística em associação com os respectivos indicadores estatísticos avaliados, sistemas dinâmicos e também técnicas e soluções aplicadas à predição de ataques DDoS. As principais vantagens desses indicadores em relação à literatura consiste no fato de não precisar treinar o sistema a fim de predizer o ataque. Esses indicadores estão fundamentados em estudos estatísticos e em sistemas dinâmicos. Para demonstrar a sua viabilidade aplicamos o conjunto de indicadores sobre traços contendo registros de ataques DDoS disponibilizados pelo CAIDA, CTU e DARPA. Com a base da CAIDA (conjunto de dados 1) o sistema foi capaz de predizer o ataque com 23 e 13 minutos de antecedência comparado ao instante da sobrecarga da rede. No caso da CTU (conjunto de dados 2) o sistema antecipou o ataque com 78 minutos prévios ao seu início. Já com o conjunto de dados da DARPA (conjunto de dados 3) o sistema STARK foi capaz de predizer o ataque com 120, 116, 114, 112, 104 e 96 minutos de antecedência. A análise dos resultados mostra, nos indicadores utilizados, o conjunto de comportamentos característicos da iminência de uma transição crítica antes do lançamento dos ataques. A partir dos resultados, é possível predizer a iminência do ataque em até duas horas, demonstrando o seu potencial de uso. O sistema traça o comportamento do fluxo de dados da rede de forma adaptativa e sem conhecimento prévio. Além disso, o sistema utiliza um conjunto de indicadores estatísticos para predizer a iminência dos ataques DDoS antes da sobrecarga da vítima (ex. enlaces ou servidor).

6.1 Trabalhos Futuros

Como trabalhos futuros esperamos expandir as análises em ambientes monitorados *online* e as *features* avaliadas quando submetidas aos indicadores de metaestabilidade. Outro aspecto que deve ser avaliado é a acurácia do sistema, ou seja, a verificação dos falsos positivos ou falsos negativos resultantes da predição do ataque. Além disso, há também a necessidade de avaliar abordagens que permitam a comparação do sistema STARK com demais soluções existentes, apesar de que, em geral, o propósito das ferramentas existentes está relacionado à detecção de ataques DDoS e não com a predição destes. Desta forma, como trabalhos futuros

pretende-se contribuir para evolução deste debate e espera-se expandir as análises sob ambientes monitorados *online* através de ambientes experimentais.

Referências

Akamai (2017). Akamai's state of the internet | security report. https://www.akamai.com/de/de/multimedia/documents/state-of-the-internet/q3-2017-state-of-the-internet-security-report.pdf. Último acesso em Mar/2018.

Azzouni, A. e Pujolle, G. (2017). A long short-term memory recurrent neural network framework for network traffic matrix prediction.

Bhuyan, M. H., Bhattacharyya, D. e Kalita, J. K. (2015). An empirical evaluation of information metrics for low-rate and high-rate ddos attack detection. *Pattern Recognition Letters*, 51:1–7.

Bovier, A. e Den Hollander, F. (2016). *Metastability: a potential-theoretic approach*, volume 351. Springer.

CAIDA, U. (2007). The CAIDA UCSD "DDoS attack 2007"dataset. Disponível em `https://www.caida.org/data/passive/ddos-20070804_dataset.xml`. Último acesso em Jun/2017.

Cisco, V. N. I. (2017). The zettabyte era: Trends and analysis. *http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html*.

Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Kéfi, S., Livina, V., Seekell, D. A., van Nes, E. H. e Scheffer, M. (2012). Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PloS one*, 7(7):e41010.

García, S. e Uhlir, V. (2011). Malware capture facility project. Disponível em `http://mcfp.weebly.com/the-ctu-13-dataset-a-labeled-dataset-with-botnet-normal-and-background-traffic.html`. Último acesso em Jun/2017.

Holgado, P., VILLAGRA, V. A. e Vazquez, L. (2017). Real-time multistep attack prediction based on hidden markov models. *IEEE Transactions on Dependable and Secure Computing*.

James, G., Witten, D., Hastie, T. e Tibshirani, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

Kwon, D., Kim, H., An, D. e Ju, H. (2017). DDoS attack volume forecasting using a statistical approach. Em *Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium on*, páginas 1083–1086. IEEE.

Laboratory, L. (2000). DARPA intrusion detection evaluation. Disponível em `https://www.ll.mit.edu/ideval/data/2000/LLS_DDOS_1.0.html`. Último acesso em Jun/2017.

Mahmoud, M., Nir, M. e Matrawy, A. (2015). A survey on botnet architectures, detection and defences. *IJ Network Security*, 17(3):264–281.

Mansfield-Devine, S. (2015). The growth and evolution of ddos. *Network Security*, 2015(10):13–20.

Mirkovic, J. e Reiher, P. (2004). A taxonomy of ddos attack and ddos defense mechanisms. *ACM SIGCOMM Computer Communication Review*, 34(2):39–53.

Moon, H. e Lu, T.-C. (2015). Network catastrophe: self-organized patterns reveal both the instability and the structure of complex networks. *Scientific reports*, 5:9450.

Networks, A. (2017). Worldwide infrastructure security report. https://pages.arbornetworks.com/rs/082-KNA-087/images/12th_Worldwide_Infrastructure_Security_Report.pdf. Último acesso em Ago/2017.

Nezhad, S. M. T., Nazari, M. e Gharavol, E. A. (2016). A novel DoS and DDoS attacks detection algorithm using ARIMA time series model and chaotic system in computer networks. *IEEE Communications Letters*, 20(4).

NicBR (2017). CERT.br registra aumento de ataques de negação de serviço em 2016. `http://www.nic.br/noticia/releases/cert-br-registra-aumento-de-ataques-de-negacao-de-servico-em-2016/`. [Último acesso em Jul/2017].

Nijim, M., Albataineh, H., Khan, M. e Rao, D. (2017). Fastdetict: A data mining engine for predecting and preventing ddos attacks. Em *Technologies for Homeland Security (HST), 2017 IEEE International Symposium on*, páginas 1–5. IEEE.

Nogueira, M., Santos, A. A. e Moura, J. M. F. (2017). Early signals from volumetric ddos attacks: An empirical study. 2.

Ramaki, A. A. e Atani, R. E. (2016). A survey of it early warning systems: architectures, challenges, and solutions. *Security and Communication Networks*.

Santos, A. A., Nogueira, M. e Moura, J. M. (2017). A stochastic adaptive model to explore mobile botnet dynamics. *IEEE Communications Letters*, 21(4):753–756.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M. e Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260):53–59.

Scheffer, M., Carpenter, S. R., Dakos, V. e van Nes, E. H. (2015). Generic indicators of ecological resilience: inferring the chance of a critical transition. *Annual Review of Ecology, Evolution, and Systematics*, 46:145–167.

Tange, O. (2011). Gnu parallel - the command-line power tool. *;login: The USENIX Magazine*, 36(1):42–47.

Tsai, C.-L., Chang, A. Y. e Ming-Szu, H. (2010). Early warning system for ddos attacking based on multilayer deployment of time delay neural network. *Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*.

Vergutz, A. (2017). Um sistema de identificação antecipada e transmissão prioritária de alertas médicos sobre wbans e wlans.

Wang, A., Mohaisen, A. e Chen, S. (2017). An adversary-centric behavior modeling of ddos attacks. Em *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*, páginas 1126–1136. IEEE.

Wichers, M., Groot, P. C., Psychosystems, E., Group, E. et al. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and psychosomatics*, 85(2):114–116.

Woolf, N. (2016). DDoS attack that disrupted internet was largest of its kind in history, experts say. Disponível em `https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet`. Último acesso em Jun/2017.

Xiao, B., Chen, W. e He, Y. (2006). A novel approach to detecting DDoS attacks at an early stage. *J Supercomput*.

Zan, X., Gao, F., Han, J. e Sun, Y. (2009). A hidden markov model based framework for tracking and predicting of attack intention. Em *Multimedia Information Networking and Security, 2009. MINES'09. International Conference on*, volume 2, páginas 498–501. IEEE.

Zargar, S. T., Joshi, J. e Tipper, D. (2013). A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE communications surveys & tutorials*, 15(4):2046–2069.