Um Sistema Autoadaptável para Predição de Ataques DDoS Fundado na Teoria da Metaestabilidade

Mateus Pelloso[1], **Andressa Vergütz**[1], **Aldri Santos**[1], **Michele Nogueira**[1]

[1]Centro de Ciência de Segurança Computacional (CCSC)
Universidade Federal Paraná (UFPR)

{mpelloso,avergutz,aldri,michele}@inf.ufpr.br

Abstract. *Distributed Denial of Service (DDoS) attacks grow in volume, sophistication, and impact. Examples are the recent DDoS attacks against the French company OVN and the name provider DYN, which have reached unprecedented volumes of malicious traffic. In general, these attacks have unexpected behaviors, being detected or mitigated only when they are in advanced stages. Thus, differently from other works, we advocate for the early prediction of DDoS attacks to assist in reducing or avoiding costs and losses due to DDoS attacks. This paper presents STARK, a self-adaptable DDoS attack prediction system. Unlike works from the literature, STARK identifies signs of attack on the network before reaching advanced stages. Based on the metastability theory, STARK provides unsupervised statistical learning and identifies the imminence of DDoS attacks. Its evaluation follows a trace-driven approach, in which three databases containing records of DDoS attacks are employed. Results show the prediction of DDoS attacks with minutes or hours in advance.*

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 possuem comportamentos inesperados, desta forma são detectados ou mitigados apenas quando se encontram em estágios avançados. Diferente de outros trabalhos, nós advogamos pelo prognóstico precoce de ataques DDoS a fim de evitar custos e perdas provenientes do ataque. Este trabalho apresenta 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, STARK provê um aprendizado estatístico não supervisionado e identifica a iminência de ataques DDoS. A avaliação do STARK segue uma abordagem orientada a traços, em que três bases de dados são utilizadas. Como resultado, STARK demonstra predizer os ataques DDoS com minutos ou horas de antecedência.*

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 os ataques 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, tais ataques geram volumes de dados inesperados, chegando a Terabytes. 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].

As soluções atuais possuem limitações ao tratar ataques DDoS desconhecidos (*zero-day or unknown attacks*). Por exemplo, ao utilizar um IDS baseado em assinaturas, exige-se um conhecimento prévio do comportamento do fluxo de dados inerente ao ataque para que possa ser comparado ao fluxo corrente da rede e, então, apontar a ocorrência de um ataque DDoS. Um outro exemplo consiste na aplicação de redes neurais para predição de ataque DDoS, entretanto estas necessitam de treinamento prévio através de conjuntos de dados contendo os fluxos de dados dos ataques que se deseja predizer. Isto limita a atuação dessas soluções aos ataques previamente conhecidos. Assim, a detecção, mitigação ou predição não supervisionada torna-se crucial para evitar custos e perdas resultantes de ataques DDoS desconhecidos.

Estudos 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, e [Nijim et al. 2017] um sistema para predizer ataques DDoS na camada de aplicação pela 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.

Nesse contexto, este trabalho advoga pela predição de ataques DDoS conhecidos (*known*) e desconhecidos (*unknown*). O objetivo é identificar esses tipos de ataques com antecedência à sobrecarga da rede ou do servidor alvo, motivado por resultados de pesquisa que indicam uma escala de tempo muito pequena entre o início das ações coordenadas do ataque e a sobrecarga total da vítima [Santos et al. 2017]. 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 and Atani 2016]. A predição indica sinais do ataque, ou seja, a possibilidade de ocorrer uma sobrecarga antes que o ataque atinja níveis irreversíveis no comprometimento do alvo.

Este artigo apresenta 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, 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 irreversíveis no estado da rede. Dessa forma, STARK segue três etapas de operação: (i) medições e preparação dos dados, (ii) predição dos ataques, e

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

(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, calculam-se os valores para os indicadores estatísticos, e com base nestes, STARK identifica sinais de um ataque DDoS e emite um alerta.

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 empíricas são disponibilizadas pelo CAIDA (*Center for Applied Internet Data Analysis*) [CAIDA 2007], pela CTU (*Czech Technical University*) [García and Uhlir 2011] e pela DARPA (*Defense Advanced Research Projects Agency*) [Laboratory 2000]. 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.

O restante do artigo está organizado como segue. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 detalha o sistema de predição de ataques DDoS. A Seção 4 descreve a metodologia de avaliação e os conjuntos de dados utilizados. A Seção 5 apresenta e discute os resultados obtidos. Por fim, a Seção 6 conclui o trabalho.

2. Trabalhos Relacionados

Existe um crescente interesse por soluções para predição de ataques DDoS. A pesquisa de [Kwon et al. 2017] propõem um método pró-ativo que estima o volume 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 de *bots*, foi realizada a estimativa do volume de ataques DDoS nesse ambiente através da análise de regressão e de correlação. Entretanto, esse estudo se restringiu a ataques DDoS volumétricos. 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 andamento. Além disso, o método necessita de dados históricos e assinaturas.

O estudo de [Wang et al. 2017] apresenta 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 carac-

terísticas são obtidas por análises de engenharia reversa. Em [Azzouni and Pujolle 2017], os autores aplicam uma rede neural sob 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. 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 de ataques. 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 dos passos dos ataques.

Dessa forma, considerando as características dos trabalhos apresentados e com base na pesquisa iniciada sobre predição de ataques DDoS [Nogueira et al. 2017], este trabalho apresenta STARK, um sistema de predição de ataques DDoS. O objetivo do sistema é prever um ataque a fim de possibilitar a tomada de medidas em tempo de prevenir a rede de graves consequências. Diferente de outros trabalhos na literatura, STARK prediz ataques DDoS *online* em seus estágios iniciais.

3. STARK: Sistema de Predição de Ataques DDoS

Esta seção detalha o sistema STARK (*Prediction **S**ys**T**em against DDoS **A**ttack on Netwo**RK***) que tem como objetivo predizer ataques DDoS conhecidos e desconhecidos em seus estágios iniciais e antes que estes sobrecarreguem a vítima. STARK complementa as soluções preventivas, reativas, e tolerantes. Ele alerta sobre possíveis sobrecargas para que ações na rede evitem maiores efeitos dos ataques. As subseções seguintes descrevem a teoria da metaestabilidade e o funcionamento do sistema proposto.

3.1. A Teoria da Metaestabilidade

Metaestabilidade é um fenômeno comum observado em uma grande variedade de situações, em geral, na natureza [Bovier and 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 escalas de tempo múltiplas e bem separadas: (i) em uma escala de tempo curta, o sistema *parece* estar em um estado de equilíbrio (estado metaestável), mas explora 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 e 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. 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 et al. 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].

A Figura 1 apresenta 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 1(a) representa o sistema em uma condição considerada de equilíbrio com dois estados metaestáveis bem 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 1(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 a redução na barreira que controla a transição. Neste trabalho, 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 STARK 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 as *features* (i.e. tamanho de pacotes, quantidade de pacotes, entre outros) como variáveis aleatórias e com base nas medições da rede consegue predizer a iminência de transições críticas. Ou seja, a iminência de ataques DDoS.

3.2. Detalhamento do Sistema STARK

A Figura 2 ilustra uma proposta de 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 equipamento de *hardware* dedicado para o sistema STARK, porém o mesmo é flexível para ser implementado junto ao *firewall* ou ao IDS.

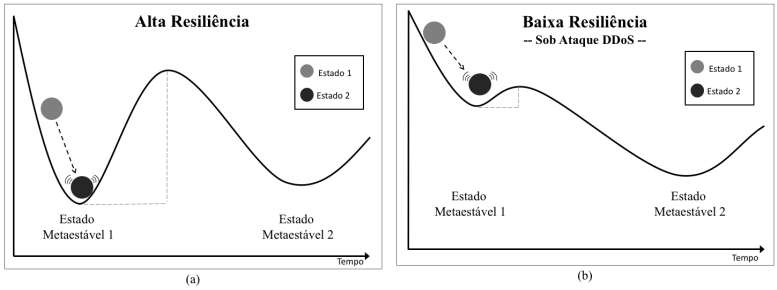

Figura 1. Perda de resiliência de um sistema metaestável devido ao ataque DDoS

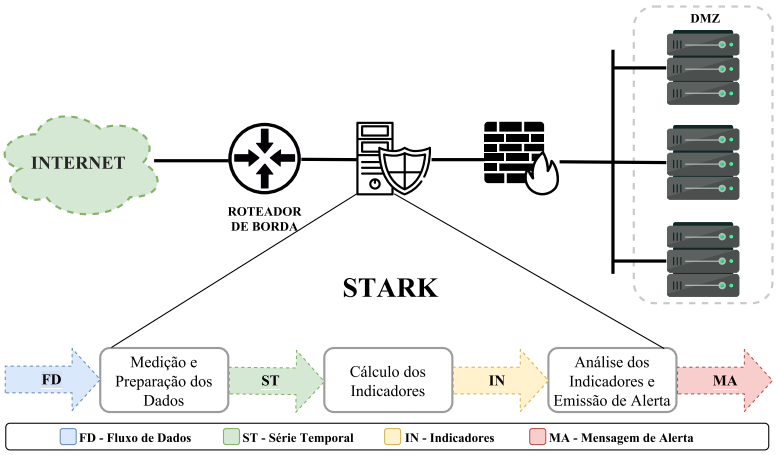

Figura 2. Posicionamento e etapas do sistema STARK

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 STARK consiste em dados de medições da rede contendo características (*features*), como tamanho dos pacotes, requisições por serviços e outras. Com base nas medições, as *features* relevantes para o cálculo dos indicadores são extraídas para criar séries temporais. A composição das séries se dá através de dados globais 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. Então, os valores dos indicadores estatísticos são analisados.

3.2.1. 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 e 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 com base nas características dos dados avaliados. Desta forma, o tamanho da janela pode ser definido por tempo, por exemplo, X segundos/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 poderá ser autoajustável. A autoadaptação é uma característica fundamental, visto que permite o autoajuste na janela de análise. Ele oferece uma dinâmica para a predição ajustando-se ao fluxo de rede sem comprometer o desempenho. Sua importância é ilustrada por viabilizar a análise de grande volume de dados, pois a mesma se adapta sem comprometer o desempenho da predição [Lima et al. 2009]. É com base nos dados contidos na janela de tamanho N que o STARK efetua o processo de cálculo dos indicadores estatísticos. Dessa 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 (P é o valor percentual sobre N), ampliando ou reduzindo a janela conforme o caso, autoajustando 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 exemplo, 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 etapa de *cálculo dos indicadores estatísticos*, tal como ilustra a Figura 2.

3.2.2. 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 neste trabalho: *taxa de retorno*, *autocorrelação*, *coeficiente de variação* e *assimetria*. Esta subseção detalha esses indicadores e como são realizados os cálculos.

A taxa de retorno representa em termos estatísticos a probabilidade da rede se recuperar de uma perturbação (ex. varredura para preparação do ataque) e se manter no estado metaestável em que se encontra (ou seja, retornar à estabilidade). Quanto maior a taxa de retorno, maior é a probabilidade da rede se manter no estado metaestável atual, ou seja, maior é a resiliência da rede (Figura 1(a)). No entanto, a ocorrência de perturbações consecutivas ou perturbações mais severas (tais como um ataque DDoS) forçam o estado da rede a sofrer uma transição crítica, migrando assim para um novo estado metaestável, muitas vezes inesperado. A transição crítica é resultado de uma grande instabilidade na rede, dificultando ou até mesmo impossibilitando o retorno ao estado metaestável em que a rede se encontrava. Desta forma, uma queda na taxa de retorno revela uma baixa resiliência (Figura 1(b)) no estado da rede, indicando a iminência de transições de estado metaestável, neste caso decorrentes de um ataque DDoS.

A autocorrelação estima a correlação entre as observações sucessivas de uma série temporal. Ela calcula o quanto o estado da rede se tornou similar entre observações consecutivas [Dakos et al. 2012]. Logo, um aumento na autocorrelação é esperado para fornecer indícios sobre a aproximação de uma transição crítica [Scheffer et al. 2009], nesse caso, de um ataque DDoS. Portanto, diante da iminência do ataque é esperado um aumento na autocorrelação, que é calculada neste trabalho seguindo a Eq. 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 num dado tempo t.

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

O coeficiente de variação analisa a dispersão das observações de uma série temporal [Scheffer et al. 2009]. Na iminência de transições críticas ou após perturbações, o estado da rede tende a alterar amplamente em torno de um estado metaestável, e assim aumenta a variabilidade das observações. Estatisticamente, o coeficiente de variação pode ser calculado através da seguinte equação: $CV = \frac{SD}{\mu}$, onde SD é o desvio padrão das observações da série temporal. Dessa forma, um aumento na curva que demonstra o comportamento deste coeficiente aponta a tendência de ocorrer uma transição crítica.

Por fim, a assimetria compara a distribuição de uma série temporal em relação a distribuição simétrica (normal), que na estatística, representa uma distribuição estável. Por outro lado, a apresentação de uma assimetria na distribuição da série temporal revela a presença de valores dispersos nas observações. Nesse sentido, para apontar a iminência de uma transição crítica (um ataque DDoS), é esperada uma assimetria na curva da distribuição das observações. Além disto, a assimetria (calculada através da Eq. 2) 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].

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

3.2.3. 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 (Figura 2). Com base nos valores dos indicadores, esta etapa analisa seus comportamentos 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 correlação conhecido como *Kendall tau* é calculado para quantificar a força da tendência do seu comportamento [Dakos et al. 2012, Mattos et al. 2017]. Os resultados desse coeficiente variam entre -1 e +1. Assim, os valores próximos ou maiores que -0.7 e 0.7 indicam uma forte tendência na intensidade do indicador [Mattos et al. 2017].

Conforme indicado na literatura [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. Assim, o comportamento dos indicadores precisa ser analisado em conjunto, pois mesmo que sejam adotadas estratégias para distorcer um deles, dificilmente o ataque conseguiria manipular o comportamento de todos os indicadores de forma coordenada.

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 emitir 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 é 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 Eq. 3. Quando a saída da FL for maior ou igual a um determinado valor, aqui representado por L (Limiar), o alerta de predição de ataque é gerado e enviado.

$$FL = \frac{-(kTR) + (kAC) + (kCV) + \sqrt{(kAS^2)}}{nTI} \quad (3)$$

4. Avaliação de Desempenho

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, CTU e DARPA. 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 literatura que permitem a fácil 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, associada 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 rede 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.

Características das Bases de Dados

Os três conjuntos de dados empregados nessa análise são os disponibilizados pelo CAIDA (**conjunto de dados 1**), CTU (**conjunto de dados 2**) e DARPA (**conjunto de dados 3**). Esta subseção descreve as principais características desses conjuntos. 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 and 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 treze 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. O 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.

Detalhes na Preparação dos Dados em 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 menores correspondiam a 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. Resultados

Esta seção apresenta os resultados de predição dos ataques DDoS sob os três conjuntos de dados descritos na Seção 4. As próximas subseções demonstram os resultados obtidos juntamente com uma discussão e análise crítica para cada conjunto de dados.

5.1. Análise do Conjunto de Dados 1

As Figuras 3 e 4 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 quatro indicadores são apresentados. A Figura 3 demonstra os resultados para os indicadores calculados sobre um série temporal de 60 segundos extraída do intervalo entre 20:50:36 às 20:51:36. Observa-se que os indicadores apresentam o comportamento esperado para indicar a iminência de um ataque DDoS, ou seja, 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 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. Ao considerar o comportamento dos indicadores juntos 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 identificar o ataque DDoS com 23 minutos de antecedência do início da sobrecarga gerada pelo ataque (que ocorreu às 21:13).

Em contrapartida, a Figura 4 ilustra o comportamento em que **não** é identificada a predição de uma transição crítica. Esta figura ilustra o comportamento identificado nas demais séries temporais antes das 21:13 e não apresentadas neste artigo devido à limitação de espaço. Ela foi incluída aqui para ilustrar uma situação em que os indicadores não seguem o comportamento que indica a iminência de uma transição crítica. 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) e o coeficiente de variação (*Kendall tau* -0.570) em descréscimo indicam pouca similaridade entre as observações dos dados e a existência de pouca variação, portanto demonstra estabilidade na rede. Já a assimetria se mostra crescente (*Kendall tau* 0.592), indicando o deslocamento das médias de tamanho

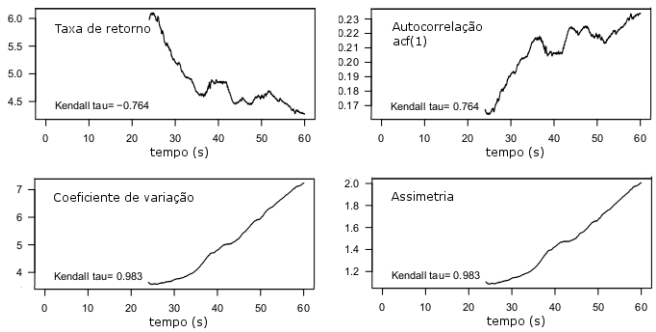

Figura 3. CAIDA: Comportamento dos indicadores no momento prévio ao ataque

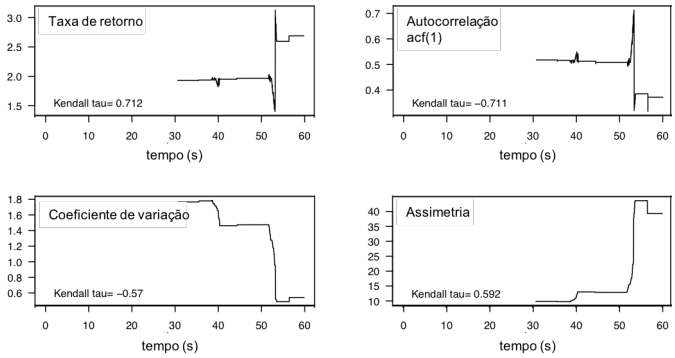

Figura 4. CAIDA: Comportamento inverso dos indicadores

dos pacotes. Desse modo, nessa série temporal é possível observar a estabilidade da rede, diante das poucas perturbações ou variações.

5.2. Análise do Conjunto de Dados 2

Os resultados apresentados na Figura 5 mostram o comportamento dos indicadores estatísticos calculados a partir de uma série temporal extraída do conjunto de dados 2. É importante ressaltar que apresentam-se apenas os resultados para a série temporal cujos indicadores predizem o ataque e por limitações de espaço os resultados para as demais séries são excluídos. Nesta série temporal há uma alteração significativa no tamanho dos pacotes variando entre 60 e 1514 bytes. Observa-se que a autocorrelação aumentou significativamente, indicando forte tendência do tamanho dos pacotes de dados permanecerem com valores em torno de 1500 bytes. A queda na taxa de retorno confirma que o fluxo da rede sofreu oscilações severas. O elevado aumento no coeficiente de variação indica instabilidade no estado da rede, ocasionado pela variação extrema no tamanho dos pacotes de 60 a 1514 bytes. Além disso, essa grande variação justifica o aumento da assimetria da distribuição dos dados. Isso revela a alta concentração em torno de pacotes com 1500 bytes, o que indica a tendência do estado da rede permanecer em torno de um estado crítico nas observações seguintes. A concentração de pacotes com o tamanho de 1500 bytes consiste de um comportamento originado por fases de preparação de ataques DDoS. Os indicadores apontam a predição do ataque com antecedência de aproximadamente 1 hora e 18 minutos de antecedência do início da sobrecarga do ataque.

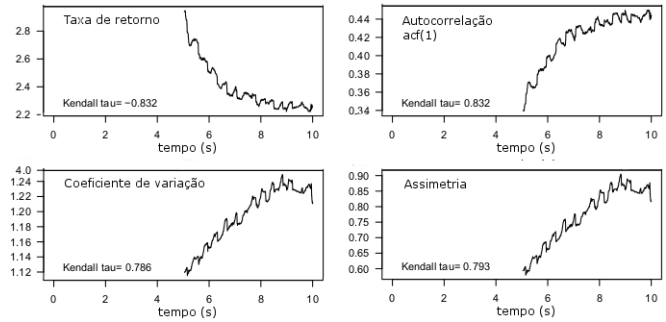

Figura 5. CTU: Comportamento dos indicadores estatísticos sob ataque

5.3. Análise do Conjunto de Dados 3

A Figura 6 ilustra o resultado dos quatro indicadores estatísticos obtidos para predição do ataque DDoS. Observa-se uma queda na taxa de retorno, apontando a dificuldade do estado da rede em retornar à estabilidade após sofrer perturbações (aumento no tamanho dos pacotes). Em conjunto, observa-se um crescimento significativo da autocorrelação,

apontando a tendência do tamanho dos pacotes permanecerem em torno de 1500 bytes. Quando analisado em conjunto com os demais indicadores, esta tendência pode significar a aproximação de uma transição crítica, neste caso resultante do ataque. 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. A taxa de retorno apresenta comportamento decrescente e 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). O comportamento conjunto dos indicadores aponta uma transição crítica, ou seja, a tendência de um ataque DDoS. A predição ocorreu com duas horas de antecedência do início da sobrecarga da rede pelo ataque.

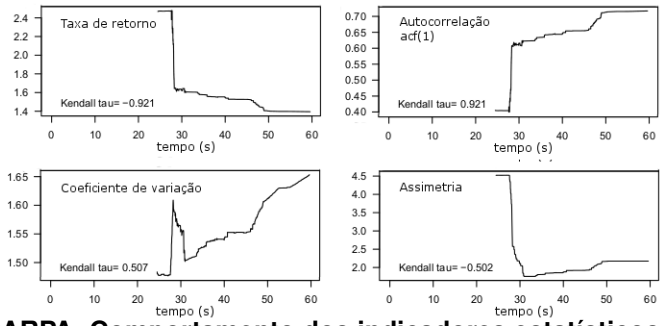

Figura 6. DARPA: Comportamento dos indicadores estatísticos sob ataque

6. Conclusão

Este artigo apresentou o sistema STARK, que aplica conceitos de metaestabilidade e sistemas dinâmicos para predizer ataques DDoS conhecidos e desconhecidos. O sistema traça o comportamento do fluxo de dados da rede de forma adaptativa e sem conhecimento prévio e utiliza um conjunto de indicadores estatísticos para predizer a iminência dos ataques antes da sobrecarga da vítima (rede ou servidor). Sua avaliação seguiu uma abordagem orientada a traços tomando como entrada os conjuntos de dados disponibilizados pela CAIDA, CTU e DARPA em que ataques DDoS reais foram registrados. Os resultados com o sistema STARK demonstraram a iminência do ataque com antecedência de aproximadamente vinte e três minutos, uma hora e duas horas à sobrecarga da rede nos traços da CAIDA, CTU e DARPA, respectivamente. Como trabalhos futuros, expandiremos as análises e as *features* avaliadas quando submetidas aos indicadores de metaestabilidade.

Referências

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

Bovier, A. and 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`. Acesso em Jun/2017.

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., and Scheffer, M. (2012). Methods for detecting early

warnings of critical transitions in time series illustrated using simulated ecological data. *PloS one*, 7(7).

García, S. and 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. Acesso em Jun/2017.

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

Kwon, D., Kim, H., An, D., and Ju, H. (2017). Ddos attack volume forecasting using a statistical approach. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pages 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. Acesso em Jun/2017.

Lima, M. N., Dos Santos, A. L., and Pujolle, G. (2009). A survey of survivability in mobile ad hoc networks. *IEEE Communications Surveys & Tutorials*, 11(1):66–77.

Mattos, V., Konrath, A., and Azambuja, A. (2017). *Introdução á Estatística: Aplicações em Ciências Exatas*. Livros Técnicos e Científicos Editora-LTC.

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., and Rao, D. (2017). Fastdetict: A data mining engine for predecting and preventing ddos attacks. In *IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–5. IEEE.

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

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

Santos, A. A., Nogueira, M., and 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., and Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260):53–59.

Vergutz, A., da Silva, R., Vieira, A. B., and Nogueira, M. (2017). Um sistema de identificação antecipada e transmissão prioritária de alertas médicos sobre WBANs e WLANs. In *Anais SBRC, Trilha Principal*. (SBRC).

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

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. Acesso em Jun/2017.

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

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