Predição Não-supervisionada de Ataques DDoS por Sinais Precoces e *One-Class* SVM

Matheus H. Lima[1], **Anderson B. de Neira**[2], **Ligia F. Borges**[1], **Michele Nogueira**[1,2]

[1]Departamento de Ciência da Computação - Universidade Federal de Minas Gerais

[2]Departamento de Informática - Universidade Federal do Paraná

{mateus.lima,ligia.borges,michele}@dcc.ufmg.br, abneira@inf.ufpr.br

Resumo. *A predição de ataques de negação de serviço distribuído (DDoS) é essencial para aumentar o tempo no combate aos ataques. Grande parte das soluções de predição de ataques DDoS utiliza dados rotulados, que é um processo custoso e limita a aplicação em ambientes reais. Para diminuir a dependência de dados rotulados, este trabalho apresenta PREDICTOR, um sistema para a predição de ataques DDoS baseado na teoria dos sinais precoces de alerta e na detecção de outliers. O sistema usa a teoria dos sinais precoces de alerta para indicar sinais da preparação dos ataques. Ele prediz o ataque usando o algoritmo One-Class SVM, um detector de outlier. Os resultados indicam que a predição ocorreu 31 minutos antes do início do ataque com acurácia de 91%.*

Abstract. *Predicting Distributed Denial of Service (DDoS) attacks is essential to allow more time to combat attacks. Most DDoS attack prediction solutions utilize labeled data, which is costly and constraints their application in real environments. This work presents PREDICTOR, a system for predicting DDoS attacks based on the theory of early warning signals and outliers detection. PREDICTOR mainly aims at reducing the dependence on labeled data. It predicts the attack using the One-Class SVM algorithm, an outlier detector. The results indicate that the prediction occurred 31 minutes before the onset of the attack with an accuracy of 91%.*

1. Introdução

Os ataques cibernéticos estão cada vez mais sofisticados e mais presentes na sociedade. Entre os diversos tipos de ataques existentes, um dos que vem evoluindo e causando danos significativos é o ataque de negação de serviço distribuído (do inglês, *Distributed Denial of Service* — DDoS) [Jyoti and Behal 2021]. As equipes de segurança são surpreendidas pelo crescimento repentino no consumo de recursos durante um ataque de negação de serviço, prejudicando a experiência dos usuários legítimos. Em fevereiro de 2023, a empresa Cloudflare identificou um dos maiores ataques já reportados, alcançando uma taxa de 71 milhões de pacotes por segundo. Além disso, esse ataque apresentou uma característica perigosa ao atingir cerca de 65 milhões de requisições por segundo em menos de 30 segundos após o seu início [Yoachimik et al. 2023].

A literatura ainda continua em sua maioria se concentrando principalmente na tarefa de detecção dos ataques DDoS. Por exemplo, Said (2023) apresenta o uso do algoritmo máquina de vetores de suporte quântica (do inglês, *Quantum Support Vector Machine* — QSVM) para detectar ataques DDoS. Bouke *et al.* (2023) propuseram

o uso algoritmo *Decision Tree* aliado com uma técnica de seleção de atributos do tráfego de rede denominada *Gini index* para detectar ataques DDoS. Os resultados apontam que a solução proposta detectou ataque DDoS de maneira mais acurada que outros algoritmos como *eXtreme Gradient Boosting* e o *Random Forest*. Contudo, a literatura indica que o momento mais adequado para lidar com o ataque DDoS é antes do atacante efetivamente lançá-lo [Kivalov and Strelkovskaya 2022]. Isso ocorre pois, após o lançamento do ataque, os recursos das redes intermediárias e da vítima são consumidos. Essas consequências impactam diretamente os usuários e a detentora do serviço. Assim, a literatura recente começou a apresentar esforços para predizer os ataques DDoS e possibilitar que as equipes de segurança possam lidar com o ataque antes que ele seja lançado [Machaka et al. 2021, Kivalov and Strelkovskaya 2022].

Apesar da existência de trabalhos relacionados com a predição de ataques DDoS, ainda há muito espaço para explorar e avançar nessa área de pesquisa visando criar soluções mais simples e eficazes. Em geral, as soluções de predição tomam como entrada dados rotulados, o que é um grande limitador. Ao construir uma solução baseada em dados rotulados, a capacidade de prever ataques DDoS fica limitada a comportamentos semelhantes aos padrões de treinamento. Além disso, os atacantes estão se tornando cada vez mais habilidosos em mascarar suas ações no meio do tráfego normal [Antonakakis et al. 2017]. Isso torna ainda mais desafiador prever ataques DDoS, até porque as soluções também assumem o balanceamento entre os dados legítimos e os dados do ataque. Como os sinais da preparação dos ataques são raros em comparação ao tráfego normal da rede, existem mais dados provenientes dos usuários normais do que dados originados pelos atacantes, dificultando a classificação.

Este trabalho apresenta o PREDICTOR, um sistema que prediz ataques DDoS usando a teoria dos sinais precoces de alerta e a detecção de *outliers*. A teoria dos sinais precoces de alerta possibilita a identificação de sinais que representam eventos futuros, por exemplo, os ataques DDoS. Ao pré-processar o tráfego de rede usando esta teoria, a preparação do ataque é evidenciada. O sistema PREDICTOR automatiza o processo de predição usando um detector de *outliers* chamado de *One-Class Support Vector Machines* (SVM). O algoritmo *One-Class* SVM não necessita de dados rotulados para realizar a predição de ataques DDoS [Muller et al. 2001] e tem um histórico de funcionamento em cenários desbalanceados [Devi et al. 2019]. Ao detectar *outliers*, o sistema PREDICTOR notifica a equipe de segurança para que eles tomem as ações pertinentes.

Este trabalho avaliou o sistema PREDICTOR seguindo dois experimentos. O primeiro usa o tráfego da rede local disponibilizado pela Universidade Técnica Tcheca (do inglês, *Czech Technical University* - CTU) por meio do *dataset* CTU-13 [Garcia et al. 2014]. Sem usar rótulos, o sistema predisse um ataque DDoS com 31 minutos e 29 segundos antes do início do ataque com uma acurácia de 91,51%. O segundo experimento usa como entrada o *dataset* CIC-DDoS2019 [Sharafaldin et al. 2019]. Neste caso, o sistema proposto predisse o ataque DDoS com nove minutos e 41 segundos antes do início do ataque. Além de não usar dados rotulados, o sistema PREDICTOR lida com o desbalanceamento dos dados. Antes do início do ataque, a quantidade de dados originados pela preparação dos ataques (tráfego originado por *bots*, teste de ataques entre outros) é muito menor do que a quantidade de tráfego normal gerado pela rede. Essas características fazem com que o sistema PREDICTOR supere a literatura atual de predição

de ataque DDoS. No Experimento 1, o sistema PREDICTOR superou o tempo de predição apresentado em [Rahal et al. 2020]. Além disso, a acurácia obtida no Experimento 1 foi maior do que a obtida em [Silva et al. 2022]. A acurácia maior foi obtida usando o *One-Class* SVM que possui menos parâmetros do que o *Long short-term memory Auto-encoder* usado em [Silva et al. 2022]. Isso torna o sistema PREDICTOR mais fácil de ser implantado em ambientes reais.

Este trabalho prossegue como segue. A Seção 2 apresenta a literatura relacionada. A Seção 3 detalha o funcionamento do sistema PREDICTOR. A Seção 4 apresenta a avaliação do sistema proposto. Por fim, a Seção 5 traz as conclusões.

2. Trabalhos Relacionados

Esta seção apresenta trabalhos que investigaram a predição de ataques DDoS. Machaka *et al.* (2021) compararam o algoritmos *Logistic Regression* (LGR), o *Support Vector Regression* e o *Kernel Ridge Regression* para prever ataques DDoS. Os autores utilizaram o *dataset* DARPA 1999 e aumentaram manualmente o número de pacotes para introduzir tráfego de ataque DDoS. Os algoritmos foram treinados usando o número de pacotes por 10 segundos em 80% do conjunto de dados. Usando os 20% restantes, o algoritmo LGR alcançou uma acurácia de 98,60% em prever a ocorrência de ataques dentro de 15 minutos. O estudo de Muhammad *et al*. (2020) avaliou técnicas de aprendizado de máquina supervisionado para predizer os ataques DDoS. A predição do ataque foi baseada na identificação do tráfego de comando e controle usado para a preparação dos ataques. Os autores selecionaram 40 atributos que poderiam auxiliar na predição dos ataques. Durante os testes, o algoritmo *Random Forest* atingiu 99% de acurácia na predição dos ataques.

Liu *et al.* (2015) predisseram ataques usando atividades maliciosas originadas pela mesma rede baseada em Sistemas Autônomos (AS). Os autores coletaram dados de listas de reputação entre janeiro de 2013 e fevereiro de 2014. Usando os endereços do protocolo de Internet (do inglês, *Internet Protocol* - IP) de cada dia, os autores identificaram os AS e o prefixo do *Border Gateway Protocol* (BGP) relacionado aos endereços IP. Os autores coletaram os eventos de segurança relatados no site *hackmageddon.com* entre outubro de 2013 e fevereiro de 2014. Eles usaram os eventos de segurança para encontrar os prefixos BGP relacionados aos nomes de domínio dos eventos. Usando dados coletados, os autores treinaram um SVM para prever ataques. A solução predisse atividades maliciosas com verdadeiros positivos de 69%. Kivalov e Strelkovskaya (2022) utilizaram as extrapolações de *spline* linear e cúbico para prever ataques DDoS. A solução utiliza o tráfego de rede antes, durante e depois de um ataque DDoS para prever ataques semelhantes. Os autores realizaram uma simulação de um ataque DDoS para avaliar a solução. As *splines* cúbicas apresentaram os melhores resultados para prever o ataque nos próximos segundos.

Apesar da alta acurácia, os trabalhos relacionados ainda são limitados. O *dataset* usado por Machaka *et al.* (2021) pouco reflete as redes atuais pois é de 1999. Além disso, os autores aumentaram manualmente a quantidade de pacotes, isso pode facilitar a tarefa de predição de ataques DDoS. Muhammad *et al*. (2020) realizam a detecção do tráfego de comando e controle das *botnets* que usam arquiteturas centralizadas. Uma vez que a *botnet* use uma arquitetura diferente a solução pode não funcionar. Além disso, usar 40 atributos do tráfego de rede pode consumir muitos recursos computacionais e, eventualmente, atrasar a predição dos ataques. Uma limitação comum da literatura é a de-

pendência de dados rotulados para o treinamento das soluções. Isso dificulta a adoção da solução em ambientes reais pois limita a solução a funcionar apenas em cenários similares. Sendo esses rótulos raramente corretos e necessitam considerar incertezas e ruídos em seus projetos para se adequarem aos diferentes cenários de segurança [Arp et al. 2022]. O sistema PREDICTOR foi projetado para não usar dados rotulados. Isso facilita a adoção em ambientes reais e a solução não fica limitada a um único tipo de ataque.

3. Sistema PREDICTOR

Esta seção detalha o sistema PREDICTOR que usa a teoria dos sinais precoces de alerta para pré-processar o tráfego de rede coletado. Esse pré-processamento realça nuances nos dados que permite à técnica *One-Class* SVM realizar a predição dos ataques DDoS sem o uso de dados rotulados. A Figura 1 apresenta o sistema PREDICTOR com 5 etapas detalhadas ao longo desta seção. A Subseção 3.1 apresenta o funcionamento da captura do tráfego de rede (Etapa 1). A Subseção 3.2 detalha o pré-processamento do tráfego de rede usado pelo sistema (Etapa 2). A Subseção 3.3 descreve o uso do algoritmo *One-Class* SVM, o detector de *outliers* usado no sistema proposto (Etapa 3). A Subseção 3.4 apresenta a Etapa 4 que compreende a predição dos ataques DDoS. Por fim, a Subseção 3.5 apresenta a ação de notificação do ataque relacionada a predição dele (Etapa 5).

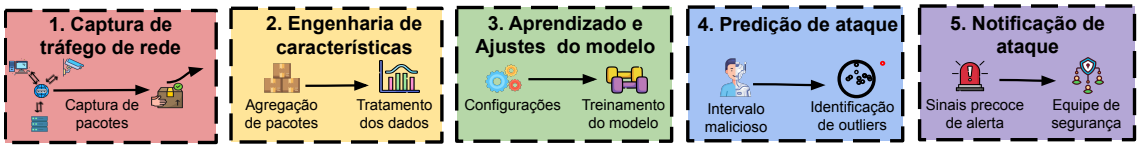

Figura 1. Arquitetura do Sistema PREDICTOR

3.1. Captura do Tráfego de Rede

Na Etapa 1, o sistema captura os pacotes da rede. Para isso, as equipes de segurança podem utilizar ferramentas do tipo *sniffer*; como *WireShark*, *tcpDump*, *Windump*, *OmniPeek*; de acordo com suas preferências. Essas ferramentas capturam o tráfego tanto em redes cabeadas quanto em redes sem fio. A forma como a coleta de dados é realizada impacta diretamente todas as etapas subsequentes, pois quanto mais célere for a coleta, mais rápido o sistema poderá predizer os ataques. De posse dos dados coletados e salvos, é possível extrair os atributos do tráfego de rede que serão usados na próxima etapa.

As equipes de segurança podem escolher os atributos do tráfego de rede que serão extraídos da captura. A escolha permite às equipes de segurança adaptarem o sistema para predizer os ataques DDoS segundo seus objetivos. Por esse motivo, este trabalho não define previamente quais atributos do tráfego de rede que serão coletados. A literatura apresenta alguns atributos que podem ser utilizados para a predição dos ataques DDoS [Muhammad et al. 2020]. Assim, o sistema PREDICTOR está apto a usá-los caso as equipes de segurança assim o decidam. A única restrição que o sistema PREDICTOR impõe é que os atributos do tráfego de rede devem ser obtidos apenas ao processar os cabeçalhos dos pacotes, nunca a carga útil (do inglês, *payload*) dos pacotes. Essa restrição visa proteger a privacidade dos usuários ao não manipular a carga útil dos pacotes. O número total de pacotes, a soma de *bytes* dos pacotes e o número de pacotes por protocolo são exemplos de atributos do tráfego de rede baseados apenas nos cabeçalhos.

3.2. Pré-processamento do tráfego de rede

Na Etapa 2, o sistema PREDICTOR usa a teoria dos sinais precoces de alerta para pré-processar o tráfego de rede coletado na forma dos atributos. A teoria dos sinais precoces de alerta possibilita a identificação de sinais que podem representar a ocorrência de eventos futuros, inclusive ataques DDoS. Esta teoria baseia-se na transição de estados de eventos observáveis. Uma das formas de transição entre estados é a transição crítica. Na transição crítica, um dado sistema evolui de um estado para o outro apresentando instabilidades. Existem alguns tipos de transição crítica, mas em todos os tipos os efeitos podem ser devastadores. Em decorrência dessas instabilidades, um sistema observável pode evoluir para um estado que apresente perturbações e inconsistências. Por exemplo, sistemas podem parar de responder devido a consumo excessivo de recursos ou redes podem ficar congestionadas devido ao número elevado de acessos. A teoria dos sinais precoces de alerta produz os sinais precoces quando o comportamento de indicadores estatísticos mudam devido a aproximação de uma transição crítica. O *skewness*, o *kurtosis*, a autocorrelação de *lag* 1, o coeficiente de variação (CV) e a variância são exemplos de indicadores estatísticos que podem produzir sinais precoces.

A Figura 2 exemplifica a produção dos sinais precoces e a predição dos ataques DDoS. O quadro A apresenta uma transição crítica do tipo *fold*. A transição crítica do tipo *fold* ocorre quando o sistema ou a rede, representados pela linha, ultrapassa o final do ponto de equilíbrio denotado por $F2$ indo em direção ao início do ponto de equilíbrio definido em $F1$. O quadro B representa os indicadores estatísticos em um momento longe do ataque DDoS (transição crítica). Os resultados obtidos no quadro B determinam o perfil dos dados no dado momento. Por exemplo, os resultados poderiam indicar que os dados observados apresentam características de uma distribuição mais plana do que pontiagudas. No momento retratado no quadro C, os indicadores estatísticos determinam novamente o perfil dos dados. A diferença entre os quadros A e B é o momento em que a medição foi realizada, pois o quadro C está em um momento mais próximo ao ataque DDoS (transição crítica). Os novos valores dos indicadores estatísticos indicam que a distribuição dos dados mudaram. O novo resultado indica uma distribuição dos dados mais pontiagudas do que planas. Essa variação pode ser um indicativo de que, no futuro, uma transição crítica vai acontecer. Este trabalho relaciona a ocorrência de transições críticas com a ocorrência de ataques DDoS. Assim, ao antecipar uma transição crítica o sistema PREDICTOR é capaz de predizer ataques DDoS.

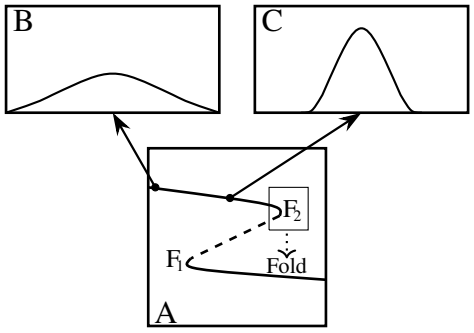

Figura 2. Aproximação de uma transição crítica [Guttal and Jayaprakash 2008]

Para processar o tráfego de rede sob a perspectiva dos indicadores estatísticos, o sistema PREDICTOR organiza cada atributo do tráfego de rede na forma de séries

temporais. Uma série temporal é constituída da coleta do tráfego de rede sob a forma dos atributos definidos na Subseção 3.1 e organizadas sequencialmente no tempo. Por exemplo, as observações sequenciais x_1 e x_2, ..., x_N representam o número de pacotes por minuto (atributo do tráfego de rede) coletados na rede do usuário. Assim, x_1 x_2 e x_N representam respectivamente o valor da primeira coleta do número de pacotes por minuto, o valor da segunda coleta do número de pacotes e o valor da enésima coleta.

O sistema PREDICTOR está fundamentado em três indicadores estatísticos: *skewness*, CV e o *kurtosis*. Tais indicadores foram selecionados pois alterações nesses valores acarretam em alterações nas características dessas medidas e essas alterações permitem predizer os ataques DDoS (Figura 2). O primeiro indicador estatístico implementado é o ***Skewness*** (Eq. 1). Na Eq. 1 a expressão x_t refere-se a cada elemento presente na série temporal. Assim, x_t refere-se ao valor de cada coleta de um atributo do tráfego de rede. O N representa a quantidade total de valores um atributo do tráfego de rede coletado. O $\bar{x}$ refere-se a média aritmética dos valores do atributo do tráfego de rede coletado. Por fim, σ representa o desvio padrão dos dados coletados.(Eq. 2). O *Skewness* mede a assimetria da distribuição dos dados em relação a sua média. Guttal e Jayaprakash (2008) observaram que variações na assimetria dos dados coletados indicam um alerta precoce confiável.

$$Skewness = \frac{N\sum_{t=1}^{N}(x_t - \bar{x})^3}{(N-1)(N-2)\sigma^3} \quad (1) \qquad \sigma = \sqrt{\frac{\sum x - \bar{x}^2}{(N-1)}} \quad (2)$$

O segundo indicador estatístico utilizado é o **CV**. Para obter o CV (Eq. 3) é preciso dividir o desvio padrão (σ) dos dados referentes ao tráfego de rede coletados pela média dos dados coletados ($\bar{x}$), onde $\bar{x} \neq 0$ [Bedeian and Mossholder 2000]. Esse indicador mostra o comportamento da variação dos dados quando comparados com a média dos conjuntos dos dados analisados. Assim, o limite inferior de CV indica uniformidade do conjunto de dados [Bedeian and Mossholder 2000]. É possível utilizar o CV como um sinal precoce pois variações podem indicar uma futura transição crítica [Dakos et al. 2012].

$$Coeficiente\ de\ variação = \frac{\sigma}{\bar{x}} \quad (3)$$

O terceiro e último indicador estatístico utilizado nesse trabalho é o ***Kurtosis*** (Eq. 4). Os termos N, x_t e $\bar{x}$ representa, respectivamente, a quantidade total de itens observados em uma série temporal, ou seja, a quantidade total dados coletados no tráfego de rede; cada valor referente aos dados coletados no tráfego de rede; e a média simples de todos os dados coletados no tráfego de rede [Joanes and Gill 1998]. Para calcular o *Kurtosis* é necessário o $\hat{y}$ (Eq. 5) [Joanes and Gill 1998]. Apesar das diferentes interpretações para o *Kurtosis*, uma das mais difundidas e controversas relaciona o valor do *Kurtosis* com grau de achatamento da curva da série temporal. Quando o valor do *Kurtosis* for positivo, a tendência da distribuição é possuir um pico. Para *Kurtosis* negativo, a tendência da série temporal é possuir uma distribuição plana, e quando o valor do *Kurtosis* for igual a zero, a distribuição assemelha-se a distribuição normal [Zhong et al. 2017]. O *Kurtosis* pode ser utilizado como sinal precoce de alerta pois Biggs *et al.* (2009) e Dakos *et al.* (2012) verificaram que o valor do *Kurtosis* pode variar ou apresentar picos próximos de transições críticas. Por exemplo, a distribuição de dados do sistema pode apresentar um padrão plano longe da transição crítica e um padrão com pico próximo à transição crítica. Essas variações indicam mudanças na distribuição de dados. Essas mudanças podem antecipar ataques DDoS (transições críticas) [Dakos et al. 2012].

$$Kurtosis = \frac{(N-1)}{(N-2)(N-3)}(N-1)\hat{y}+6 \quad (4) \qquad \hat{y} = \frac{N\sum(x_t-\bar{x})^4}{[\sum(x_t-\bar{x})^2]^2} \quad (5)$$

Após a coleta do tráfego de rede na forma dos atributos e da organização na forma de séries temporais, o sistema PREDICTOR agrega o tráfego de rede em intervalos de tempo. A escolha do intervalo de tempo é de responsabilidade da equipe de segurança e deve ser feita com cautela. Definir o intervalo de tempo ideal maximiza o aproveitamento dos recursos disponíveis e faz com que a predição dos ataques seja acurada e antecipada. Assim, o pré-processamento do tráfego de rede acontece quando a teoria dos sinais precoces de alerta, por intermédio dos indicadores estatísticos, processa os atributos do tráfego de rede organizado nas séries temporais e agrupados a cada intervalo de tempo para identificar os sinais da ocorrência de ataques DDoS futuros.

3.3. Treinamento

Na Etapa 3, o sistema usa o algoritmo *One-Class* SVM para detectar *outliers* no tráfego de rede pré-processado com a teoria dos sinais precoces de alerta. O *One-Class* SVM é um algoritmo de aprendizado de máquina não supervisionado que realiza a detecção de uma única classe. Devido ao desbalanceamento dos dados, onde existe mais tráfego normal do que o tráfego de preparação do ataque, esse trabalho usou o tráfego normal como a única classe. Isso torna o *One-Class* SVM adequado para detecção de *outliers* [Devi et al. 2019] sendo é ideal para a predição dos ataques DDoS.

O *One-Class* SVM funciona como uma função de decisão que envolve a maioria do tráfego normal formando uma barreira limite. Essa função de decisão é definida por meio do treinamento do modelo. O treinamento é realizado apenas com o tráfego de rede pré-processado com a teoria dos sinais precoces de alerta (Etapa 2) sem o uso de rótulos. Todos os pontos dentro dessa barreira fazem parte da classe dos dados normais, e todos os pontos fora dessa barreira serão classificados como *outliers* [Amer et al. 2013] e representam a preparação dos ataques DDoS. O *One-Class* SVM possui parâmetros que desempenham papéis fundamentais na capacidade do algoritmo de trabalhar com dados de alta dimensionalidade. O *Kernel* determina a forma pela qual os dados serão separados no espaço. Possíveis valores para o parâmetro *Kernel* são: $poly$, $linear$, rbf, $sigmoid$ e $precomputed$. O parâmetro nu atua como um limite para a taxa de erros durante o treinamento e também como um limite inferior para a proporção de vetores de suporte. Ao ajustar o valor do parâmetro nu, é possível controlar a quantidade de pontos de dados que serão considerados como anomalias durante a fase de treinamento. Esse valor deve estar no intervalo $(0, 1]$, representando a porcentagem de pontos que serão classificados como *outliers*. A configuração e o treinamento do *One-Class* SVM impactam diretamente nos resultados obtidos. Assim, ao usar o *One-Class* SVM, o sistema PREDICTOR proporciona às equipes de segurança uma maior autonomia para adaptar o algoritmo, desenvolvendo assim um modelo de predição adequado e ajustado para atender suas necessidades.

3.4. Predição do ataque

Na Etapa 4, o sistema utiliza o detector de *outliers* para predizer os ataques DDoS considerando os dados coletados e pré-processados. A predição dos ataques acontece quando o *One-Class* SVM identifica um *outlier* nos atributos do tráfego de rede que foram processados usando a teoria dos sinais precoces de alerta. Os *outliers* representam mudanças

anormais no tráfego de rede, que podem não ser percebidos mesmo por especialistas. Essas mudanças no tráfego de rede são realçadas graças ao pré-processamento do sistema PREDICTOR (Etapa 2). Assim, ao aplicar a teoria dos sinais precoces de alerta sobre os atributos do tráfego de rede, a mudança do comportamento da rede, refletida nos resultados dos indicadores estatísticos, possibilita a predição dos ataques DDoS. Com base nessa premissa, o sistema PREDICTOR utiliza o detector de *outliers* para processar os dados coletados. Isso permitirá ao sistema identificar a presença ou ausência de *outliers*. No caso da detecção de *outliers*, o sistema avança para Etapa 5.

3.5. Notificação do ataque

Na Etapa 5, ocorre a notificação dos ataques, uma fase crucial que permite à equipe de segurança dispor de um tempo de reação mais amplo para mitigar eventuais danos. Após o sistema identificar um determinado momento de captura de tráfego de rede como *outlier* (Etapa 4), o sistema envia uma notificação à equipe de segurança. Essa notificação pode ser estabelecida por meio de interfaces de programação de aplicações (do inglês, *Application Programming Interface* - API), as quais comunicam o sistema PREDICTOR aos sistemas de monitoramento utilizados pela equipe de segurança. É necessário realizar uma configuração cuidadosa nesse sentido, de forma que o sistema assegure que os alertas emitidos sejam encaminhados ao sistema de monitoramento. Uma vez que a equipe de segurança receba uma notificação, ela poderá adotar as medidas de proteção adequadas. Por exemplo, aplicar técnicas de *blackholing* [Dietzel et al. 2016, Wichtlhuber et al. 2022].

4. Avaliação

Esta seção apresenta a metodologia de avaliação do sistema PREDICTOR e os resultados. Este trabalho avaliou o sistema proposto por meio de dois experimentos. O primeiro foca em predizer ataques DDoS na rede interna de uma universidade. O segundo prediz o ataque DDoS em uma rede onde os *bots* estão ligados à vítima por meio da Internet. Esta seção está dividida em três subseções. A Subseção 4.1 apresenta os parâmetros da experimentação. A Subseção 4.2 apresenta os resultados. Por fim, a Subseção 4.3 discute os resultados obtidos e os compara com a atual literatura.

4.1. Cenários de avaliação

O Experimento 1 usa a captura 51 da CTU-13 [Garcia et al. 2014]. A captura foi realizada sobre o tráfego da rede da universidade CTU ao longo de 8803 segundos. A captura apresenta apenas os cabeçalhos dos 46 milhões de pacotes, totalizando 41 GB de dados. Dez *bots* lançam os ataques de inundação de pacotes dos protocolos *User Datagram Protocol* e *Internet Control Message Protocol*. O primeiro ataque aconteceu no segundo 5632. Apesar de não necessitar de dados rotulados para realizar as predições, o *One-Class* SVM necessita de treinamento. Este trabalho treinou o identificador de *outlier* usando um terço de todo o *dataset* (até o segundo 2934). Os dados utilizados para o teste variam dos segundo 2934 até 5631. Isso ocorre pois o propósito do sistema PREDICTOR é predizer os ataques DDoS. Assim, a fase de teste varia até o segundo anterior ao início do ataque. O *One-Class* SVM usado nesse experimento usa todas as configurações padrão da biblioteca[1], exceto pelos parâmetros nu e $kernel$. A configuração que maximizou a acurácia

[1]https://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html

e a predição dos ataques DDoS foi $kernel = poly$ e $nu = 0.325$. É importante ressaltar que esses hiperparâmetros foram obtidos empiricamente e o código fonte está online[2].

O *dataset* CIC-DDoS2019 [Sharafaldin et al. 2019] foi avaliado no Experimento 2. Este *dataset* apresenta mais de 61 milhões de pacotes completos (cabeçalho e carga útil). O *dataset* possui 19 ataques DDoS disparados ao longo de dois dias. A avaliação do sistema usou apenas o primeiro ataque DDoS realizado no primeiro dia da captura. O arquivo correspondente a este ataque tem mais de 27 GB de dados referente ao tráfego malicioso e tráfego normal. O primeiro ataque iniciou no segundo 1484 da captura e durou 540 segundos. Este ataque é do tipo *Portmap* que explora as portas UDP ou TCP utilizadas pelo serviço *Portmap*. Seguindo o mesmo padrão do primeiro experimento, este trabalho treinou o *One-Class* SVM usando um terço de todo o *dataset* (até o segundo 674). Como o objetivo é realizar a predição do ataque DDoS, os dados usados no teste variam entre os segundos 674 e 1483. *One-Class* SVM foi configurado seguindo o padrão da biblioteca Scikit-learn, exceto pelo $kernel$ e nu. A configuração que maximizou os resultados foi $kernel = poly$ e $nu = 0.13$, obtida empiricamente. A automatização da obtenção dos hiperparâmetros será o foco de pesquisas futuras.

A avaliação do sistema utilizou o intervalo de tempo de apenas um segundo para agrupar o tráfego de rede disponibilizado nos *datasets*. Usando um segundo para agrupar o tráfego de rede, este trabalho objetiva obter resultados mais precisos e aumentar o tempo de predição dos ataques DDoS. Apesar da solução não usar rótulos para realizar a predição dos ataques, este trabalho usa os rótulos apenas para medir os acertos e os erros do sistema PREDICTOR. Assim, cada intervalo de tempo foi rotulado usando a presença de tráfego originado ou destinado aos dispositivos usados para o ataque (*bots*). Caso o intervalo apresente ao menos um pacote onde o endereço IP de origem ou destino for de um *bot*, este intervalo foi considerado malicioso. Caso contrário, o intervalo foi considerado normal.

Os experimentos usaram como atributos o total de endereços IP de origem e destino e o total de pacotes trafegados em cada intervalo de tempo. O total de pacotes é um importante atributo na identificação do tráfego de comando e controle que geralmente ocorre antes do início do ataque [Feng et al. 2018] e a falsificação de IPs é um ação comum em ataques DDoS [Jyoti and Behal 2021]. Por esses motivos, este trabalho escolheu esses três atributos para avaliar o sistema PREDICTOR. Além disso, esses atributos podem ser extraídos apenas usando o cabeçalho dos pacotes. Isso facilita a captura dos e o armazenamento dos dados, pois a carga útil dos pacotes não precisa ser coletada.

Como definido na Seção 3, o sistema aplica a teoria dos sinais precoces de alerta sobre os atributos do tráfego de rede. Essa ação gera três características para cada atributo. Portanto, o sistema calcula o valor do *kurtosis*, *skewness* e CV (indicadores estatísticos) para o total de endereço IP de origem e destino e o total de pacotes para cada intervalo de tempo (atributos do tráfego de rede). Esse processo gera nove séries temporais, uma para cada combinação entre indicadores estatísticos e atributos do tráfego de rede. Cada item da série temporal corresponde ao valor dos indicadores estatísticos calculados sobre as coletas do tráfego de rede nos intervalos de tempo um segundo. Por exemplo, cada item da série temporal representa o valor dos atributos do tráfego de rede (quantidade de pacotes e total de endereços IP de origem e destino) por segundo. Assim, são consideradas

[2]https://github.com/1995-Matheus-Lima/sbseg-2023-sistema-predictor

nove características distintas onde as três primeiras correspondem ao valor de *Kurtosis*, *Skewness* e coeficiente de variação relacionadas ao total de endereços IPs de origem. As três características seguintes registram esses valores em relação ao total de IPs de destino e as três últimas capturam esses valores em relação ao número total de pacotes.

O pré-processamento dos atributos do tráfego de rede sobre a perspectiva das séries temporais usou o conceito das janelas deslizantes de tamanho fixo. O objetivo é avaliar o sistema ao longo do *dataset* e prevenir tendências equivocadas [Bury et al. 2020]. A Figura 3 ilustra o conceito de janelas deslizantes de tamanho fixo. Ela apresenta uma série temporal com a medição da quantidade de pacotes recebidos em uma rede por segundo. Para uma janela móvel de tamanho dois, o sistema PREDICTOR calcula os sinais precoces de alerta utilizando os valores presentes nas duas primeiras posições. Quando uma nova observação estiver disponível, nesse caso representado pela terceira posição da série temporal, o sistema desliza e calcula os sinais precoces de alerta utilizando os valores das posições dois e três. O sistema segue até o fim da série temporal ou enquanto recebe novos dados. O tamanho da janela deslizante de tamanho fixo nesta avaliação compreende 10% do tamanho do *dataset*. Como a literatura ainda não é unânime sobre o valor ideal, este trabalho definiu o valor de 10% para otimizar o tempo de predição dos ataques.

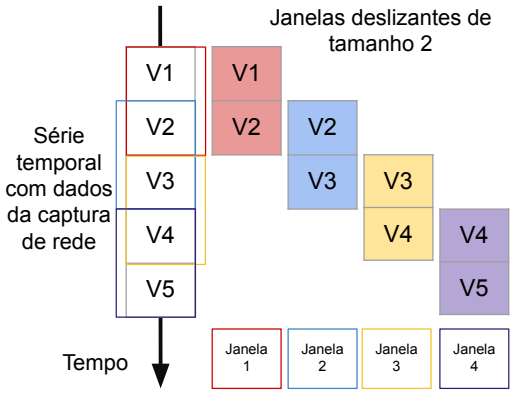

Figura 3. Ilustração do conceito de janelas deslizantes de tamanho fixo

Para demonstrar a eficácia do sistema, este trabalho utiliza as métricas de acurácia (Eq. 6), o *recall* (Eq. 7) e a precisão (Eq. 8). O *recall* e a precisão podem ser obtidos para os intervalos normais e para os intervalos maliciosos. Além disso, a predição dos ataques DDoS convive com o grande desbalanceamento das classes, onde existem mais intervalos normais do que intervalos maliciosos. Assim, para simplificar a apresentação dos resultados, contemplar o desbalanceamento e reportar informação sobre os dois tipos de intervalos, este trabalho apresenta o *recall* e a precisão e médias ponderadas pelo total de intervalos maliciosos e normais. Nas Eq. 6, 7 e 8, N indica o total de intervalos; VP indica o total de verdadeiros positivos; o FP representa o total de total de falsos positivos; FN é o total de falsos negativos; e VN indica o total de verdadeiros negativos.

$$Acurácia = \frac{VP + VN}{N} \quad (6) \qquad Recall = \frac{VP}{VP + FN} \quad (7) \qquad Precisão = \frac{VP}{VP + FP} \quad (8)$$

4.2. Resultados da avaliação

A Tabela 1 apresenta os resultados obtidos no Experimento 1. É oportuno destacar os desbalanceamento dos dados. Enquanto existem 2482 intervalos normais na base de teste,

existem apenas 216 intervalos maliciosos. O sistema detectou corretamente 2467 intervalos normais, estes são VPs. O sistema detectou erroneamente 214 intervalos maliciosos como normais, produzindo assim FPs. O sistema gerou 15 FNs ao relacionar intervalos maliciosos com intervalos normais. A predição do ataque ocorreu quando o sistema detectou dois *outliers* que realmente eram intervalos maliciosos, produzindo então VNs. O sistema atingiu uma acurácia de 91,51%, um valor muito expressivo para a predição de ataques DDoS não usando dados rotulados. A precisão e o *recall* médios ponderados são respectivamente de 85,59% e 91,51%. Por fim, a equipe de segurança recebeu o primeiro alerta 31 minutos e 29 segundos antes do início do ataque.

Tabela 1. Resultados do sistema PREDICTOR no CTU-13 (Experimento 1)

Matriz de confusão		Classe Real			
		Intervalo normal	Intervalo malicioso	Acurácia	91,51%
Classe Predita	Intervalo normal	2467	214	Precisão	85,59%
	Intervalo malicioso	15	2	*Recall*	91,51%

Os resultados obtidos no Experimento 2 usando o *dataset* CIC-DDoS2019 estão dispostos na Tabela 2. Assim como no experimento anterior, o desbalanceamento dos dados está presente. Enquanto existem 560 intervalos normais na base de teste, existem apenas 250 intervalos maliciosos. O sistema detectou corretamente 536 intervalos normais (VPs). O sistema detectou erroneamente 243 intervalos maliciosos como normais (FPs). O sistema gerou 24 FN ao relacionar intervalos maliciosos com intervalos normais. A predição do ataque ocorreu quando o sistema detectou sete *outliers* que realmente eram intervalos maliciosos (VN). O sistema atingiu uma acurácia de 67,04%. A precisão e o *recall* médios são respectivamente de 54,54% e 67,04%. Por fim, uma equipe de segurança receberia o primeiro alerta nove minutos e 41 segundos antes do início do ataque.

Tabela 2. Resultados do sistema PREDICTOR no CIC-DDoS2019 (Experimento 2)

Matriz de confusão		Classe Real			
		Intervalo normal	Intervalo malicioso	Acurácia	67,04%
Classe Predita	Intervalo normal	536	243	Precisão	54,54%
	Intervalo malicioso	24	7	*Recall*	67,04%

4.3. Discussão dos resultados

Os resultados apresentados são relevantes, pois mesmo em ambientes intensamente desbalanceados, o sistema proposto foi capaz de predizer o ataque DDoS sem o uso de rótulos. Mesmo com a presença de pouco tráfego originado por *bots* antes do início do ataque a predição ocorreria 31 minutos e 29 segundos antes do início do ataque no Experimento 1. Isso significa que a equipe de segurança teria mais de 30 minutos para conduzir ações para evitar que o ataque DDoS causasse prejuízos. No Experimento 2, o primeiro alerta ocorreu nove minutos e 31 segundos antes do ataque. O resultado do Experimento 2 fica ainda mais expressivo pois todo o processo de ataque (antes, durante e depois do ataque) possui apenas 33 minutos e 44 segundos de duração e a solução demorou apenas 3 minutos e 49 para realizar esse alerta (após o treinamento). Além disso, a predição ocorreu em um ambiente onde a grande maioria dos pacotes é originada por dispositivos normais, não infectados. Contudo, devido a acurácia obtida no Experimento 2 (67,04%) é necessário

aprimorar o modelo para obter um resultado mais seguro gerando menos erros mesmo que a predição do ataque tenha sido obtida sem o uso de rótulos.

Além da predição dos ataques, é oportuno discutir sobre os VPs, FPs, e FNs. Como os *datasets* são desbalanceados, a maioria dos intervalos são normais. Identificá-los é um desafio que o sistema PREDICTOR conseguiu fazer com exatidão. No Experimento 1, dentre os 2479 intervalos normais, o sistema detectou corretamente quase 100% deles (2467). Apesar da quantidade de FPs apresentar a maior quantidade de erros, este tipo de erro é o menos relevante. Isso porque o FP indica que um intervalo malicioso foi identificado como normal. Assim, o FP indica apenas que o sistema deixou de produzir alertas de predição. Contudo para que a predição aconteça é necessário a existência de, ao menos, um VN. A identificação de 214 FP pode ser explicada pelo baixo volume de tráfego de rede originado pela preparação do ataque. Assim, o tráfego malicioso presentes nestes intervalos não é suficiente para indicá-los como *outliers*. Os FNs correspondem ao erro mais grave, pois em 15 vezes o sistema emitiria um alerta sem relação real com o ataque. Contudo 15 FNs correspondem a menos de 1% de todo o *dataset*. Assim, proporcionalmente os FNs são pouco relevantes. A precisão e o *recall* médios ponderados refletem a baixa quantidade de erros, pois o sistema obteve métricas acima de 85%.

No Experimento 2, o comportamento dos VPs, FPs, e FNs foi similar ao do Experimento 1. Dos 560 intervalos normais, o sistema identificou 536 corretamente. Assim, o sistema identificou corretamente 95,71% dos intervalos normais. Assim como no experimento anterior, os FPs representam a maior quantidade de erros. O sistema proposto reconheceu 243 intervalos maliciosos como intervalos normais. Esse total de erros fez com o que as métricas (acurácia, precisão e *recall*) obtidas pelo sistema fossem menores do que as obtidas no primeiro experimento. Isso pode ser explicado pelo fato do tráfego malicioso pouco impactar nos sinais precoces de alerta. Isso faz com que, muitos sinais da preparação dos ataques passem despercebidos, sem representarem *outliers*. Apesar da quantidade de FP, esse erro é menos relevante do que os FNs. Visto que, o sistema conseguiu predizer a ocorrência de ataques DDoS em sete vezes (VN) e com bastante tempo de antecipação (mais de nove minutos). Outro ponto que pode contribuir com a quantidade de FPs é que a documentação não apresenta o início da infecção. Assim, este trabalho rotulou todo o tráfego de *bots* anterior ao início do ataque como intervalo malicioso. Desconsiderando os FP, o sistema PREDICTOR apresentaria 95,77% de acurácia. Por fim, o sistema apresentou apenas 24 FNs, que correspondem a 1% de todo o *dataset*.

Para reforçar a importância dos resultados, este trabalho compara-os com a literatura. A solução proposta aumenta o tempo de predição presente na literatura. Pois os resultados obtidos pelo sistema superam os resultados apresentados em [Rahal et al. 2020] ao predizer o mesmo ataque com 31 minutos e 29 segundos de antecedência. Além disso, no Experimento 1, o sistema PREDICTOR obteve 91,51% com 229 erros totais ($FP + FN$). Essa acurácia supera a obtida em Silva *et al.* (2022) , onde os autores analisaram a mesma captura avaliada no Experimento 1. Isso significa que o sistema PREDICTOR erra menos do que a solução proposta em [Silva et al. 2022]. Além disso, para atingir os resultados apresentados, Silva *et al.* (2022) usaram uma rede neural com muitos parâmetros que consomem muito tempo e recurso para ser configurada e treinada. O modelo utilizado neste trabalho usou apenas dois parâmetros diferentes dos valores padrão definidos na biblioteca, além de consumir menos recursos computacionais e tempo

para realizar as predições. Por fim, ao comparar os resultados do Experimento 1 com os resultados apresentados por de Neira *et al.* (2023), o sistema PREDICTOR prevê o ataque 1 minuto e 29 segundos adiantado e alcança uma acurácia superior, 91,51%, em comparação com acurácia de 84,6% obtida por de Neira *et al.* (2023). Tudo isso é benéfico para a literatura pois aumenta as possibilidades de uso em ambientes reais.

5. Conclusão

Com a constante evolução dos ataques DDoS, detectá-los não é o suficiente para evitar os transtornos causados por eles. Este trabalho apresenta o PREDICTOR, um sistema para auxiliar as equipes de segurança propiciando-lhes mais tempo no combate aos ataques. O sistema usa a teoria dos sinais precoces de alerta aliado com a detecção dos *outliers* para realizar a predição dos ataques. Resultados indicam que a proposta realiza a predição dos ataques com acurácia, precisão e *recall* superiores a 85%. O sistema predisse um ataque DDoS com mais de 30 minutos de antecedência. Trabalhos futuros definem a novas avaliações usando outros *datasets*, a análise de *flows*, a seleção de atributos, o teste de outros indicadores da teoria dos sinais, a avaliação de outros detectores de *outliers* e a avaliação frente a outros tipos de tráfego de rede (*flash-crowds/elephant flows*).

Agradecimentos

Este trabalho foi financiado pelo CNPq (#309129/2017-6 e #432204/2018-0) pela FAPESP (#2018/23098-0 e #2022/06840-0) e pela CAPES (#88887.501287/2020-00).

Referências

Amer, M., Goldstein, M., and Abdennadher, S. (2013). Enhancing one-class support vector machines for unsupervised anomaly detection. In *ACM SIGKDD*, pages 8–15.

Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Halderman, J. A., Invernizzi, L., Kallitsis, M., Kumar, D., Lever, C., Ma, Z., Mason, J., Menscher, D., Seaman, C., Sullivan, N., Thomas, K., and Zhou, Y. (2017). Understanding the Mirai botnet. In *USENIX CSS*, page 1093–1110, USA. USENIX.

Arp, D., Quiring, E., Pendlebury, F., Warnecke, A., Pierazzi, F., Wressnegger, C., Cavallaro, L., and Rieck, K. (2022). Dos and don'ts of machine learning in computer security. In *USENIX*, pages 3971–3988, MA. USENIX Association.

Bedeian, A. G. and Mossholder, K. W. (2000). On the use of the coefficient of variation as a measure of diversity. *ORM*, 3(3):285–297.

Biggs, R., Carpenter, S. R., and Brock, W. A. (2009). Turning back from the brink: Detecting an impending regime shift in time to avert it. *PNAS*, 106(3):826–831.

Bouke, M. A., Abdullah, A., ALshatebi, S. H., Abdullah, M. T., and Atigh, H. E. (2023). An intelligent DDoS attack detection tree-based model using Gini index feature selection method. *MICPRO*, 98:104823.

Bury, T. M., Bauch, C. T., and Anand, M. (2020). Detecting and distinguishing tipping points using spectral early warning signals. *J. R. Soc.*, 17(170).

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):1–20.

de Neira, A. B., Borges, L. F., de Araújo, A. M., and Nogueira, M. (2023). Engenharia de sinais precoces de alerta para a predição de ataques DDoS. In *WGRS*, page 14. SBC.

Devi, D., Biswas, S. K., and Purkayastha, B. (2019). Learning in presence of class imbalance and class overlapping by using One-Class SVM and undersampling technique. *Connection Science*, 31(2):105–142.

Dietzel, C., Feldmann, A., and King, T. (2016). Blackholing at IXPs: On the effectiveness of DDoS mitigation in the wild. In *PAM*, pages 319–332, Cham. Springer.

Feng, Y., Akiyama, H., Lu, L., and Sakurai, K. (2018). Feature selection for machine learning-based early detection of distributed cyber attacks. In *DASC*, page 8. IEEE.

Garcia, S., Grill, M., Stiborek, J., and Zunino, A. (2014). An empirical comparison of botnet detection methods. *Computers & Security*, 45:100–123.

Guttal, V. and Jayaprakash, C. (2008). Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecology Letters*, 11(5):450–460.

Joanes, D. N. and Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *J. R. Stat. Soc*, 47(1):183–189.

Jyoti, N. and Behal, S. (2021). A meta-evaluation of machine learning techniques for detection of DDoS attacks. In *INDIACom*, pages 522–526, India. IEEE.

Kivalov, S. and Strelkovskaya, I. (2022). Detection and prediction of DDoS cyber attacks using spline functions. In *TCSET*, page 4, UA.

Liu, Y., Zhang, J., Sarabi, A., Liu, M., Karir, M., and Bailey, M. (2015). Predicting cyber security incidents using feature-based characterization of network-level malicious activities. In *IWSPA*, page 3–9, USA. ACM.

Machaka, P., Ajayi, O., Maluleke, H., Kahenga, F., Bagula, A., and Kyamakya, K. (2021). Modelling DDoS attacks in IoT networks using machine learning. *arXiv*, pages 1–20.

Muhammad, A., Asad, M., and Javed, A. R. (2020). Robust early stage botnet detection using machine learning. In *ICCWS*, pages 1–6, Pakistan. IEEE.

Muller, K.-R., Mika, S., Ratsch, G., Tsuda, K., and Scholkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE TNN*, 12(2):181–201.

Rahal, B. M., Santos, A., and Nogueira, M. (2020). A distributed architecture for DDoS prediction and bot detection. *IEEE Access*, 8:159756–159772.

Said, D. (2023). Quantum computing and machine learning for cybersecurity: Distributed denial of service (DDoS) attack detection on smart micro-grid. *Energies*, 16(8).

Sharafaldin, I., Lashkari, A. H., Hakak, S., and Ghorbani, A. A. (2019). Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In *ICCST*.

Silva, G. L. F. M., de Neira, A. B., and Nogueira, M. (2022). A deep learning-based system for ddos attack anticipation. In *LATINCOM*, pages 1–6.

Wichtlhuber, M., Strehle, E., Kopp, D., Prepens, L., Stegmueller, S., Rubina, A., Dietzel, C., and Hohlfeld, O. (2022). IXP scrubber: learning from blackholing traffic for ml-driven DDoS detection at scale. In *SIGCOMM*, pages 707–722.

Yoachimik, O., Desgats, J., and Forster, A. (2023). Cloudflare mitigates record-breaking 71 million request-per-second DDoS attack. Acesso em: 05/2023. https://blog.cloudflare.com/cloudflare-mitigates-record-breaking-71-million-request-per-second-ddos-attack/.

Zhong, L., Cheng, L., Xu, H., Wu, Y., Chen, Y., and Li, M. (2017). Segmentation of individual trees from TLS and MLS data. *IEEE J-STARS*, 10(2):774–787.