Automatização da Seleção de Modelos Não Supervisionados na Predição de Ataques DDoS

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

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

Resumo. *As técnicas de Aprendizado de Máquina (AM) auxiliam na automatização de diferentes tarefas em cibersegurança. Diante da vasta quantidade de algoritmos e de hiperparâmetros, um dos maiores problemas é identificar o modelo que reduz os erros. Assim, selecionar o algoritmo de AM é crucial para lidar de forma adequada com cada ataque e cenário. Para resolver este problema, este artigo apresenta a técnica ALTO que seleciona autonomamente o detector de outlier que maximiza a separação do tráfego de rede maligno do benigno sem o uso de rótulos (não supervisionado) no contexto de predição de ataques de negação de serviço distribuído (DDoS). Os resultados indicam que a técnica ALTO seleciona modelos capazes de superar a precisão dos modelos configurados manualmente e geram até 500% mais verdadeiros positivos.*

Abstract. *Machine learning (ML) techniques assist in the automation of various cybersecurity tasks. Given the vast number of algorithms and hyperparameters, one of the biggest challenges is identifying the model that minimizes errors. Therefore, selecting the appropriate ML algorithm is crucial for effectively handling each attack and scenario. Thus, this article presents the ALTO technique, which autonomously selects the outlier detector that maximizes the separation of malicious network traffic from benign traffic without the use of labels (unsupervised) in the context of predicting distributed denial-of-service (DDoS) attacks. The results indicate that the ALTO technique selects models capable of outperforming manually configured models and generating up to 500% more true positives.*

1. Introdução

Segundo o Fórum Econômico Mundial [Forum 2024], os ataques cibernéticos estão entre os principais riscos globais para os próximos 8 anos, com crescente volume, sofisticação e frequência. Particularmente, os ataques de negação de serviço distribuído (do inglês, *Distributed Denial of Service* — DDoS) estão entre as ameaças cibernéticas mais prejudiciais, consumindo rapidamente os recursos das vítimas e causando negação de serviço [Jyoti and Behal 2021]. Em 2023, a Google sofreu um ataque DDoS sem precedentes, que durou cerca de dois minutos e atingiu 398 milhões de solicitações por segundos, atingindo seu pico de solicitações em menos de 30 segundos [Kiner and April 2023].

A literatura reforça que a associação da identificação das características mais relevantes [Fernandes 2017] com a utilização de algoritmos de predição de ataques DDoS otimiza a acurácia dos resultados [Liu et al. 2022]. Liu *et al.* (2022) emprega análise de

componentes principais e eliminação recursiva de características com *support vector machines* (SVM), treinada com dados rotulados. No entanto, essas soluções requerem tempo e expertise, que muitas vezes são escassos nas equipes de segurança. Para amenizar essa dependência, a literatura definiu o Aprendizado de Máquina Automatizado (AutoML) como uma área de pesquisa que simplifica o uso e reduz o custo do Aprendizado de Máquina (AM). O AutoML visa identificar e configurar um algoritmo de AM que reduz erros de classificação para um conjunto de dados selecionado [Feurer et al. 2015].

A maioria das soluções para lidar com ataques DDoS requerem tempo e esforços de especialistas para a configuração e calibragem dos modelos de AM. Cerca de 40% das empresas gastam mais de um mês no processo do desenvolvimento e configuração de modelos [Hecht 2019]. Além do tempo necessário para a entrega de um modelo, as soluções anteriormente apresentadas para detecção de ataques DDoS utilizam métodos supervisionados, que precisam de dados rotulados no treinamento. Isto é um problema, pois dados rotulados são custosos, requerendo um responsável por anotar e rotular os dados, o que é desafiador em diferentes contextos como o de redes de computadores.

Assim, este trabalho propõe a técnica ALTO, um **A**utom**L** para a de**T**ecção de ***O**utliers*. A técnica ALTO utiliza dados não rotulados para a seleção automatizada de algoritmos não-supervisionados. Para selecionar autonomamente o algoritmo mais adequado para predizer ataques DDoS, a técnica compara os algoritmos de detectores de *outliers* candidatos usando os índices de silhueta e Calinski-Harabasz (CH). Esses índices ajudam a diferenciar agrupamentos considerando a coesão e separação dos grupos. A técnica ALTO usa esses índices para caracterizar agrupamentos e selecionar o detector de *outlier* mais adequado sem interação humana. Assim, as equipes de segurança não despendem tempo na seleção de algoritmos e tão pouco na rotulação de dados reais.

A avaliação da técnica ocorreu de forma comparativa com o trabalho de [Lima et al. 2023]. O objetivo de [Lima et al. 2023] era predizer ataques DDoS usando o AM não supervisionado utilizando o *One-Class* SVM configurado manualmente sobre o tráfego de rede pré-processado. Já este trabalho usa a técnica ALTO para selecionar um detector de *outliers* adequado para predizer ataques DDoS. Os resultados preliminares apontam que a técnica aumentou a precisão das predições entre 0,52% e 3,42% e os verdadeiros positivos entre 57% e 500%. Assim, a técnica seleciona algoritmos de AM não supervisionados capazes de predizer ataques DDoS sem usar dados rotulados.

Este trabalho é apresentado da seguinte forma. A Seção 2 apresenta a literatura relacionada. A Seção 3 detalha o funcionamento da técnica ALTO. A Seção 4 apresenta a avaliação. Por fim, a Seção 5 resume as considerações finais e trabalhos futuros.

2. Trabalhos Relacionados

Mohmand *et al.* (2022) propuseram uma técnica para predição de ataques DDoS que seleciona um modelo preditivo similar ao proposto no ALTO. No entanto, a escolha do modelo preditivo requer a rotulação dos dados de treinamento. A dependência de dados rotulados não cobre ataques *zero-day*, impossibilitando prever todos os cenários. Já a técnica ALTO foi projetada para selecionar modelos sem dados rotulados. Poulakis (2020) propõe a seleção de algoritmos não supervisionados que agrupam os dados em *clusters* e avalia modelos em bases offline usando *Grid Search* e *Randomized Search*, que dependem de dados rotulados. Para cada modelo, métricas de validação de *clusters* (e.g., índices *CH*

e *Davies-Bouldin)* são extraídas da base de dados para identificar a técnica de AM que maximize tais índices. A hipótese é que conjuntos de dados semelhantes possuem soluções semelhantes. Ao processar um novo conjunto de dados, a solução compara-o com a base de conhecimento *offline* e sugere o algoritmo que melhor resolveu problemas similares anteriormente. Ao contrário da técnica ALTO, o método de Poulakis (2020) requer conhecimentos prévios para sugerir o algoritmo de AM não supervisionado para novos dados. A técnica ALTO seleciona algoritmos detectores de *outliers*, enquanto Poulakis (2020) sugere algoritmos de agrupamento. Portanto, a técnica ALTO tende a funcionar melhor em ambientes desbalanceados.

3. Técnica ALTO

Esta seção detalha o funcionamento da técnica ALTO, um **A**utom**L** para a de**T**ecção de ***O**utliers*. A técnica ALTO visa encontrar o modelo detector de *outlier* ideal para identificar tráfego de preparação do ataque DDoS, sem que seja necessária a utilização de dados rotulados (Fig. 1). A técnica ALTO recebe e pré-processa tráfego de rede adicionado pelo usuário. Posteriormente, a técnica analisa o tráfego de rede pré-processado para identificar o modelo ideal que será utilizado para predizer os ataques DDoS.

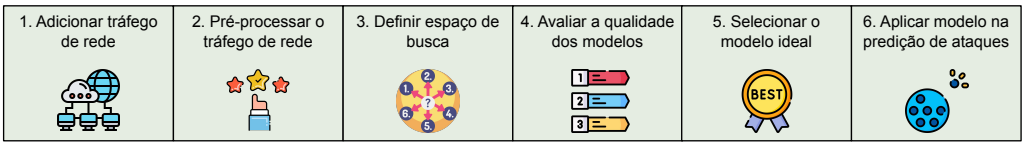

Figura 1. Visão geral da técnica ALTO

3.1. Adicionar e pré-processar o tráfego de rede

A técnica ALTO opera quando o usuário insere o tráfego de rede para a análise do modelo ideal (Etapa 1 da Figura 1). A versão atual da técnica ALTO pressupõe que o tráfego de rede já esteja tratado e preparado para realizar a detecção de *outliers* e predizer os ataques DDoS. Assim, atualmente, a técnica ALTO não aplica nenhum tipo de pré-processamento nos dados adicionados pelo usuário. As versões futuras da técnica ALTO irão identificar e aplicar técnicas de pré-processamento para melhorar a seleção do modelo (Etapa 2 da Figura 1). Realizar a seleção de características sem o uso de dados rotulados e que maximize os resultados facilitará a escolha mais rápida e assertiva do modelo. Além disso, dados faltantes e escalas diferentes impactam negativamente a utilização dos detectores de *outliers*. Portanto, a técnica ALTO futuramente escolherá estratégias de pré-processamento que maximizem os índices de qualidade (Sub. 3.3), visando aumentar a acurácia.

3.2. Definir Espaço de Busca

A Etapa 3 utiliza três detectores de *outliers* para definir o espaço de busca. O espaço de busca contém 934 modelos avaliados pela técnica ALTO. Os algoritmos que compõem o espaço de busca são o *One-Class* SVM, *Stochastic Gradient Descent* (SGD) *One-Class* SVM e o *Local Outilier Factor* (LOF). O *One-Class* SVM é adequado para detecção de *outliers* em ambientes com quantidade desbalanceada de dados [Devi et al. 2019]. O SGD *One-Class* SVM apresenta menor complexidade computacional para detectar *outliers* em comparação com o *One-Class* SVM. Além disso, o SGD *One-Class* SVM é adequado para ambientes com desbalanceamento dos dados [Oliveira et al. 2023]. Por fim, o LOF

calcula a densidade de desvio de um determinado ponto até seus vizinhos, sendo adequado para conjuntos de dados com muitas dimensões [Oliveira et al. 2023].

O espaço de busca da técnica ALTO é formado por 934 modelos produzidos pelas combinações dos parâmetros dos três algoritmos citados. A técnica ALTO varia os parâmetros *kernel* e *nu* para o *One-Class* SVM. Os parâmetros representam respectivamente a forma com que os dados são separados espacialmente e uma estimativa da quantidade de *outliers* presente no conjunto de dados. Os valores para o *kernel* são *linear*, *poly* e *rbf* e os possíveis valores para *nu* são os valores do intervalo $[0,05$ e $0,5]$ com incrementos de $0,001$. Já para o algoritmo SGD *One-Class* SVM, a técnica ALTO varia os parâmetros *learning_rate* e *nu*. O parâmetro *Learning_rate* indica como o algoritmo controla a taxa de atualização dos pesos durante o processo de treinamento usando o SGD. Este parâmetro varia entre as opções *constant*, *optimal*, *invscaling* e *adaptive*. O parâmetro *nu* tem a mesma definição e configuração apresentada para a técnica *One-Class* SVM. Por fim, para o algoritmo LOF, a técnica ALTO varia os parâmetros *algorithm* e *n_neighbors*. O parâmetro *algorithm* representa a forma em que é medida o tráfego malicioso em relação a seus vizinhos. E o parâmetro *n_neighbors* define a quantidade de vizinhos próximos utilizados para calcular a densidade. Os possíveis valores para *algorithm* são *auto*, *ball_tree*, *kd_tree* e *brute*. Os possíveis valores para *n_neighbors* são os inteiros do intervalo $[2, 10]$. Portanto, a técnica ALTO visa encontrar o modelo que otimize a predição de ataques dentre as 934 combinações possíveis, formadas pela associação de cada algoritmo.

3.3. Avaliar a Qualidade dos Modelos

A técnica ALTO aplica todos os 934 modelos ao tráfego de rede fornecido pelo usuário para identificar o modelo ideal (Etapa 4). Ela avalia os resultados de detecção de *outliers* usando índices de qualidade: o índice de silhueta e o índice de CH. O índice de silhueta mede a coesão e separação dos dados, variando de -1 (dados mal agrupados) a $+1$ (alta coesão e separação). O termo *a* indica a distância média entre uma amostra e os pontos do mesmo agrupamento, enquanto o termo *b* representa a distância mínima entre pontos de diferentes agrupamentos. O índice de CH avalia a validade do agrupamento com base na soma da média dos quadrados entre e dentro dos agrupamentos (Eq. 2) [Liu et al. 2010]. Os resultados do CH variam sem um intervalo fixo. Maiores valores indicam agrupamentos com maior separação e compacidade. O termo SSw indica variância total nos agrupamentos e o termo SSb a variância total intra agrupamento. O termo N representa a quantidade de amostras, K é a quantidade de agrupamentos.

$$Índice\ de\ silhueta = \frac{b-a}{\max(a,b)} \qquad (1) \qquad Índice\ CHI = \left(\frac{\text{SSb}}{\text{SSw}}\right) \times \left(\frac{N-K}{K-1}\right) \qquad (2)$$

A técnica ALTO usa os índices de silhueta e CH, pois em cenários reais, obter os rótulos para treinar e avaliar as soluções é custoso ou até inviável. Além disso, usar rótulos pode limitar a atuação das soluções de predição de ataques DDoS. Esses índices foram projetados para avaliar a qualidade dos modelos sem a utilização dos rótulos e reais e a literatura respalda a utilização deles [Poulakis 2020]. Portanto, a hipótese que embasa a técnica ALTO é que ao utilizar esses índices, a técnica ALTO consegue escolher modelos que maximizem a separação do tráfego normal do tráfego de preparação dos ataques DDoS, representados pelos *outliers*. Pois, quanto mais coesos e separados os dados estiverem, maior será os resultados dos índices de silhueta e CH.

Tabela 1. Resultados da técnica ALTO em comparação com a literatura

Experimento	Verdadeiro Positivo	Acurácia	Precisão	*Recall*
Exp 1 (CTU-13)	**12**	89,96%	**86,04%**	89,96%
[Lima et al. 2023] (CTU-13)	2	**91,51%**	85,59%	**91,51%**
Exp 2 (CICDDoS2019)	**11**	**67,04%**	**56,41%**	**67,04%**
[Lima et al. 2023] (CICDDoS2019)	7	67,04%	54,54%	**67,04%**

3.4. Selecionar e Aplicar o Modelo na Predição de Ataques

Para selecionar o modelo ideal, a técnica ALTO usa apenas o tráfego de rede inserido pelo usuário, aplicando a função "*fit_predict*" dos 934 modelos. Com o resultado da identificação de *outliers*, a técnica calcula os índices de silhueta e CH para todos os modelos. Como o índice CH não possui limite superior, é necessário normalizar os resultados para que ele não impacte negativamente na seleção do modelo ideal (Etapa 5). Assim, os índices de silhueta e CH são normalizados para valores entre 0 e 1, onde 0 representa os piores modelos e 1 os melhores. Com os índices de silhueta e CH calculados e normalizados para todos os 934 modelos, a técnica ALTO pode identificar o modelo ideal. Para isso, a técnica ALTO calcula a média aritmética simples dos índices de silhueta e CH para todos os modelos. Portanto, o modelo que apresentar a maior média será o modelo sugerido para o usuário aplicar em sua tarefa de predição de ataques DDoS. Assim, a decisão do modelo ideal fica pautada sobre dois índices de qualidade que não usam dados rotulados.

4. Avaliação

Esta seção avalia a técnica ALTO em dois experimentos. Este trabalho evoluiu o estudo de [Lima et al. 2023], que visava prever ataques DDoS usando AM não supervisionado com o *One-Class* SVM configurado manualmente sobre o tráfego de rede pré-processado. Portanto, o objetivo desta avaliação é usar a técnica ALTO para selecionar um algoritmo de AM não supervisionado, especificamente um detector de *outliers* que supere ou iguale os resultados obtidos manualmente em [Lima et al. 2023]. Para obter uma comparação justa, segue-se o pré-processamento e as métricas de avaliação, acurácia, precisão média ponderada e *recall* médio ponderado definidas em [Lima et al. 2023]. Por fim, os experimentos foram realizados no Google Colab, com resultados e código online[1].

O Experimento 1 usa a base de dados CTU-13 [Garcia et al. 2014]. A captura 51, simulação de tráfego de botnet contida no CTU-13, reporta 8803 segundos do tráfego da rede, em que 1546 desses segundos contém tráfego malicioso, com 46 milhões de pacotes e 41 GB de dados sem o *payload* dos pacotes. Os pesquisadores lançaram ataques de inundação *User Datagram Protocol* e *Internet Control Message Protocol* e o primeiro ataque ocorreu no segundo 5632. Na técnica ALTO, um detector de *outlier* é selecionado a partir de um conjunto de dados, que neste experimento compreende um terço da base de dados (até o segundo 2934). Enquanto os dados de teste estão entre os segundos 2934 e 5631, onde 216 segundos contêm tráfego malicioso. A Tabela 1 apresenta os resultados do Experimento 1. A técnica ALTO selecionou o *One-Class* SVM com *poly* e 0,5 como parâmetros para *kernel* e *nu*. Foram identificados 12 verdadeiros positivos com o modelo selecionado, com acurácia e *recall* ponderado de 89,96% e precisão de 86,04%.

O Experimento 2 usou a base de dados CICDDoS2019 [Sharafaldin et al. 2019]. O CICDDoS2019 disponibiliza mais de 61 milhões de pacotes, relacionado ao primeiro

[1]https://github.com/1995-Matheus-Lima/sbseg-2024-tecnica-ALTO

ataque DDoS. O ataque é do tipo *Portmap*, possui mais de 27 GB de dados referente à preparação e condução do ataque e às ações de usuários normais. O ataque analisado durou 540 segundos e iniciou-se no segundo 1484 da captura. Seguindo a mesma divisão de treinamento e teste do Experimento 1, este estudo utilizou um terço de todo o conjunto de dados (até o segundo 674) para selecionar o modelo ideal. Os dados de teste e de previsão de ataque variam entre os segundos 674 e 1483, sendo 250 segundo tráfego malicioso.

Após a análise dos dados, a técnica ALTO identificou 29 modelos com as maiores médias dos índices de qualidade, sugerindo esses modelos para a predição dos ataques DDoS. A técnica ALTO não possui uma estratégia para desempatar os modelos. Contudo, trabalhos futuros avaliarão a complexidade dos algoritmos, o tempo de execução e outras características dos modelos para criar uma estratégia de priorização e escolha em casos de empate. A Tabela 1 apresenta os resultados obtidos usando o SGD *One-Class* SVM com a opção *adaptive* para o *learning_rate* e 0.474 para o parâmetro *nu*. A acurácia foi de 67,04%, a precisão média ponderada de 56,41% e o *recall* médio ponderado de 67,04%.

5. Considerações Finais e Trabalhos Futuros

Os resultados apresentados reforçam a relevância e o potencial deste trabalho. No Experimento 1, o modelo selecionado pela técnica ALTO aumentou a precisão da predição dos ataques DDoS em 0,52% e a taxa de verdadeiros positivos aumentou em 500%. Em contrapartida, a acurácia da técnica ALTO foi 1,69% menor em relação a [Lima et al. 2023]. No Experimento 2, a acurácia da técnica ALTO foi a mesma de [Lima et al. 2023]. O total de verdadeiros positivos aumentou 57%, de 7 para 11, e a precisão subiu 3,42%. Esse crescimento é relevante, por oferecer mais notificações às equipes de segurança, permitindo antecipar ataques e evitar danos. Além disso, os resultados foram obtidos autonomamente, sem o uso de rótulos. Por fim, o tempo de seleção do modelo ideal para a predição de ataques nas bases de dados CTU-13 e CICDDoS2019 foram, respectivamente, 11 minutos e 46 segundos e 1 minuto e 30 segundos. Esse resultado supera os 29 minutos e 51 segundos e 15 minutos e 21 segundos obtidos por [Brito et al. 2023].

Esta pesquisa em andamento possui trabalhos futuros que endossam a relevância e guiam a evolução da proposta. Além de adicionar técnicas de preprocessamento, o uso de novos índices para complementar a avaliação também é um trabalho futuro. Novos índices como o S_Dbw [Halkidi and Vazirgiannis 2001] e o CDbw [Halkidi and Vazirgiannis 2008] são evoluções dos índices clássicos. Esses e outros índices modernos podem incrementar os resultados obtidos pela técnica ALTO. Aumentar o espaço de busca e adicionar uma função de otimização são importantes para trabalhos futuros. Atualmente, o espaço de busca da técnica ALTO compreende 934 modelos baseados em três algoritmos. Contudo, existem outros algoritmos como o *Isolation Forest*. Portanto, adicionar novos algoritmos aumenta a quantidade de modelos avaliados e possibilita o aumento na acurácia do modelo selecionado. Além disso, a versão atual da técnica ALTO avalia várias combinações de parâmetros. Adicionar funções de otimização como a bayesiana [Feurer et al. 2015] pode aumentar o espaço de busca e a acurácia, diminuindo a quantidade de avaliações realizadas pela técnica.

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

Brito, D., de Neira, A. B., Borges, L. F., and Nogueira, M. (2023). An autonomous system for predicting ddos attacks on local area networks and the internet. In *2023 IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1–6. IEEE.

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.

Fernandes, C. A. F. S. (2017). Algoritmo do tipo filter-wrapper de seleção de features para utilização na seleção de genes. Master's thesis, Universidade de Coimbra.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J. T., Blum, M., and Hutter, F. (2015). Efficient and robust automated machine learning. In *NIPS*, page 2755–2763, USA. MIT Press.

Forum, W. E. (2024). The global risks report 2024 Acesso em: 06/2024. https://abrir.link/kLZLQ.

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

Halkidi, M. and Vazirgiannis, M. (2001). Clustering validity assessment: finding the optimal partitioning of a data set. In *ICDM*, pages 187–194.

Halkidi, M. and Vazirgiannis, M. (2008). A density-based cluster validity approach using multi-representatives. *Pattern Recognit. Lett.*, 29(6):773–786.

Hecht, L. (2019). Add it up: How long does a machine learning deployment take? https://acesse.dev/2cDZb.

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.

Kiner, E. and April, T. (2023). Google mitigated the largest DDoS attack to date, peaking above 398 million rps Acesso 10/23. https://tinyurl.com/5xb2kux3.

Lima, M. H., de Neira, A. B., Borges, L. F., and Nogueira, M. (2023). Predição não-supervisionada de ataques ddos por sinais precoces e one-class svm. In *SBSeg*, pages 403–416. SBC.

Liu, Y., Li, Z., Xiong, H., Gao, X., and Wu, J. (2010). Understanding of internal clustering validation measures. In *2010 IEEE ICDM*, pages 911–916. IEEE.

Liu, Z., Qian, L., and Tang, S. (2022). The prediction of DDoS attack by machine learning. In *ECNCT*, volume 12167, pages 681–686. SPIE.

Mohmand, M. I., Hussain, H., Khan, A. A., Ullah, U., Zakarya, M., Ahmed, A., Raza, M., Rahman, I. U., Haleem, M., et al. (2022). A machine learning-based classification and prediction technique for DDoS attacks. *IEEE Access*, 10:21443–21454.

Oliveira, J. M., Almeida, J., Macedo, D., and Nogueira, J. M. (2023). Comparative analysis of unsupervised machine learning algorithms for anomaly detection in network data. In *2023 LATINCOM*, pages 1–6. IEEE.

Poulakis, G. (2020). Unsupervised AutoML: a study on automated machine learning in the context of clustering. Master's thesis, Πανεπιστήμιο Πειραιώς.

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*.