

Avaliação de Métodos de Seleção de Características para Ataques de Inundação HTTP

Iuri A. Mundstock¹, Rafael A. Berri¹, Bruno L. Dalmazo¹

¹ Centro de Ciências Computacionais
Universidade Federal do Rio Grande (FURG)

iuri.andrade.mundstock@gmail.com, rafaelberri@furg.br, dalmazo@furg.br

Abstract. *In an increasingly digital and interconnected world, the protection of web servers becomes essential. In this context, the goal of this work is to perform an evaluation of feature selection methods for the detection of anomalies generated by the DoS Slowhttptest attack. As a result, we aim to contribute to the advancement of knowledge in the field of anomaly detection and identify a set of more relevant features.*

Resumo. *Em um mundo cada vez mais digital e conectado, a proteção dos servidores web torna-se essencial. Nesse contexto, o objetivo deste trabalho é realizar uma avaliação de métodos de seleção de características para detecção de anomalias geradas pelo ataque DoS Slowhttptes. Como resultado, busca-se contribuir para o avanço do conhecimento na área de detecção de anomalias e indicar um conjunto de características mais relevantes.*

1. Introdução

Vivemos em um mundo onde o avanço da tecnologia e a crescente digitalização de atividades diárias levam cada vez mais pessoas a se conectar à Internet e utilizar seus serviços. De acordo com o relatório da We Are Social e Hootsuite de 2022, mais de 4,95 bilhões de pessoas em todo o mundo utilizam a internet [Simon 2022]. Diante disso, os servidores web tornaram-se essenciais disponibilizando conteúdo online. Uma vez que, é através desses servidores que são oferecidos uma vasta gama de benefícios, desde a facilidade de acesso e compartilhamento de informações até o fornecimento de serviços em diferentes partes do mundo.

Sem dúvida, é crescente a dependência dos servidores web, tornando-se cada vez mais claro o grande desafio que representa a segurança desses sistemas. Entre as principais ameaças que os afetam, estão os ataques de negação de serviço (DoS) e ataques de negação de serviço distribuídos (DDoS), que têm um grande potencial de causar prejuízos significativos sejam eles. Por essa razão, é indispensável que sejam implementadas medidas efetivas de proteção e análise de tráfego, capazes de detectar tais ataques antes que eles possam causar danos irreparáveis [Dalmazo et al. 2018].

Na literatura sobre segurança de servidores web, existem inúmeros estudos que examinaram os ataques DoS e DDoS [Dalmazo et al. 2021], bem como suas possíveis soluções como [Najafabadi et al. 2017] que se propõe a detectar esses ataques, ou [Bhargava et al. 2022] que utiliza-se do aprendizado de máquina para analisar o tráfego de rede. Ao revisar essas pesquisas, observa-se uma lacuna no estudo e

aplicação de métodos de seleção de características. A seleção de características é um processo essencial na análise de dados, permitindo identificar as características mais relevantes e informativas dentro de um conjunto de dados.

Nesse contexto, o objetivo deste trabalho é investigar e analisar métodos de seleção de características aplicados na detecção de anomalias em servidores web com protocolo HTTP. Serão exploradas diferentes abordagens de seleção de características, com o intuito de identificar quais as diferenças entre os resultados obtidos. As próximas seções detalham os trabalhos relacionados, a proposta, a implementação, algumas considerações finais e perspectivas de trabalhos futuros.

2. Trabalhos Relacionados

Essa seção apresenta a metodologia da revisão sistemática da literatura, utilizada para recuperar trabalhos existentes no contexto de detecção de ataques de negação de serviço (DoS), e ataques de negação de serviço distribuídos (DDoS) em servidores web com o protocolo HTTP.

2.1. Método de Seleção

Nesse contexto, as palavras chaves são "ataques de negação de serviço", "ataques de negação de serviço distribuídos", "web services" e "HTTP". As Tabelas 1 e 2 apresentam e os critérios de seleção e compilam seus resultados, enquanto a Tabela 3 apresenta os artigos resultantes.

((distributed denial of service OR DDoS) OR (denial of service OR DoS))
AND (web services AND HTTP)

Tabela 1. Critérios de Inclusão (IC) e Exclusão (EC)

Critério de inclusão	
IC1	Período de Publicação entre 2015-2023(maio)
IC2	Ter acesso aberto
Critério de Exclusão	
EC1	Ser um survey
EC2	Não abordar detecção de ataques DDoS ou DoS

Tabela 2. Estudos selecionados

Resultados	Total
IEEE Xplore	92
Não incluídos por IC1, IC2	38
Excluídos por EC1, EC2	36
Artigos selecionados	18

2.2. Discussão sobre as limitações do estado-da-arte

Ao revisitar os artigos selecionados exibidos na Tabela 3, observa-se uma lacuna sobre estudos comparativos de métodos de seleção de características para melhorar a detecção de anomalias no tráfego de rede. Essa observação sugere que identificar um conjunto de características relevantes pode ser um objeto de interesse para alavancar futuras pesquisas. A aplicação de métodos de seleção de características

pode auxiliar na identificação das características mais relevantes e discriminativas para a detecção de anomalias, reduzindo a dimensionalidade dos dados e melhorando a eficiência e precisão dos sistemas de detecção de intrusão. Portanto, explorar a utilização de métodos de seleção de características na detecção de anomalias é uma abordagem promissora que vai de encontro com necessidade crescente de eficiência e economia de recursos.

Tabela 3. Trabalhos selecionados

N	Título	Ano
01	Web proxy based detection and protection mechanisms against client based HTTP attacks	2015
02	User Behavior Anomaly Detection for Application Layer DDoS Attacks	2017
03	New sensing technique for detecting application layer DDoS attacks targeting back-end database resources	2017
04	Modelling Behavioural Dynamics for Asymmetric Application Layer DDoS Detection	2021
05	Implementation of Machine Learning Based DDOS Attack Detection System	2022
06	HTTP Low and Slow DoS Attack Detection using LSTM based deep learning	2022
07	How Secure are Web Servers? An Empirical Study of Slow HTTP DoS Attacks and Detection	2016
08	Framework for Preprocessing and Feature Extraction from Weblogs for Identification of HTTP Flood Request Attacks	2018
09	Detection of HTTP Attacks using Machine Learning	2022
10	Detection of DoS/DDoS Attack against HTTP Servers Using Naive Bayesian	2015
11	Detection Methods of Slow Read DoS Using Full Packet Capture Data	2020
12	CloudZombie: Launching and Detecting Slow-Read Distributed Denial of Service Attacks from the Cloud	2015
13	Behaviour analysis of HTTP based slow denial of service attack	2017
14	Anomaly detection for web server log reduction: A simple yet efficient crawling based approach	2016
15	Analysis and detection of application-independent slow Denial of Service cyber attacks	2021
16	Adaptive behaviour pattern based botnet detection using traffic analysis and flow intervals	2017
17	A Novel Application Layer DDOS Detecting Method Based on Cluster Method	2018
18	A first look at HTTP(S) intrusion detection using NetFlow/IPFIX	2015

3. Proposta

O objetivo deste trabalho é analisar, comparar e avaliar os métodos de seleção de características para determinar se eles podem aprimorar a eficiência da detecção de anomalias em redes. Foi realizado um estudo sobre os diferentes métodos existentes, e também, conduziu-se um experimento utilizando o conjunto de dados CIC-IDS2017

[Sharafaldin et al. 2018]. A seguir, um classificador *Random Forest* foi aplicado com diferentes métodos de seleção de características para a detecção de anomalias.

A estrutura do trabalho se resume à um estudo detalhado do conjunto de dados CIC-IDS2017 e aplicação dos métodos de seleção de características selecionados. Posteriormente, é utilizado um classificador *Random Forest* para realizar a classificação dos dados com base nas características selecionadas, modelo também utilizado por [Al-gethami and Aljuhani 2022]. Por fim, os resultados obtidos são analisados e discutidos, buscando-se identificar padrões, avaliar a eficiência e fornecer conclusões relevantes. Os detalhes sobre cada uma dessas etapas serão apresentados de forma mais aprofundada ao longo dessa seção.

Nesse contexto, o modelo conceitual ilustra a organização do trabalho na Figura 1. A Figura apresenta o fluxo de dados de forma clara e organizada, proporcionando uma visão geral do processo e facilitando o entendimento da proposta.

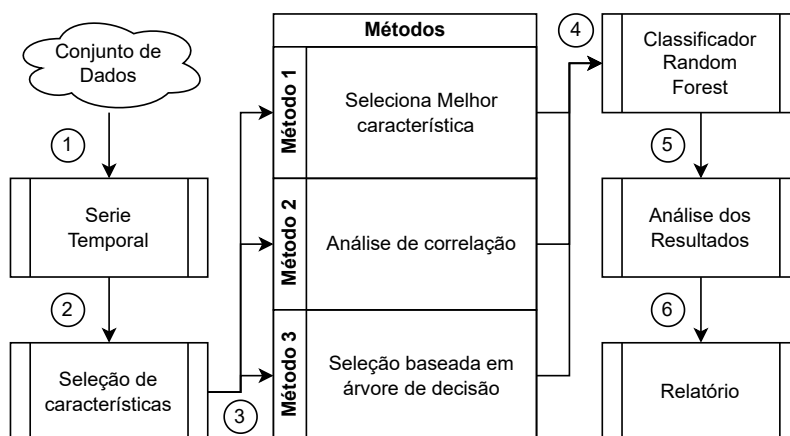


Figura 1. Modelo conceitual

O modelo é alimentado com uma série temporal (fluxo 1). Então, serão aplicados os seguintes métodos e seleção de características: seleção da melhor característica, análise de correlação e seleção baseada em árvores de decisão (fluxo 2). Cada método será avaliado individualmente, levando em consideração critérios como eficácia na seleção das características mais relevantes, capacidade de reduzir a dimensionalidade dos dados e melhorar a performance na detecção de anomalias (fluxo 3).

Depois, o classificador *Random Forest* é aplicado aos resultados obtidos de cada método de seleção de características (fluxo 4). Esse algoritmo utiliza uma abordagem baseada em conjunto de árvores de decisão para classificar os dados e identificar padrões anômalos. Com a aplicação do *Random Forest* aos resultados de cada método de seleção de características, será possível avaliar a eficácia dos métodos para detecção de anomalias e obter uma análise comparativa para determinar qual deles apresenta melhores resultados (fluxo 5).

De posse dos resultados, é realizada uma análise dos dados obtidos. Nessa etapa, serão examinados os padrões identificados, as métricas de desempenho do modelo de detecção de anomalias e as conclusões tiradas a partir desses resultados (fluxo 6). Por fim, será elaborado um relatório detalhado sobre os resultados obtidos ao

longo do estudo. Nesse relatório, serão apresentadas as análises e interpretações dos resultados de cada etapa do processo, incluindo a seleção de características, a aplicação do classificador *Random Forest*.

4. Avaliação

Para a avaliação, foi utilizado o conjunto de dados CIC-IDS2017. Para implementar os métodos de seleção de características e classificação, foi utilizada a linguagem de programação Python com as bibliotecas Pandas, matplotlib, numpy e sklearn. Contudo, vale lembrar que antes da aplicação dos métodos de seleção de características faz-se necessário um tratamento prévio dos dados. Assim, todos dados foram convertidos para reais e descartados os dados redundantes e ID do tráfego, bem como foram separados os diferentes tipos de ataques permanecendo como alvo apenas os ataques do tipo *DoS Slowhttptest*.

4.1. Métodos Avaliados

A seguir é descrito os métodos avaliados e seus parâmetros de seleção:

- Seleção das melhores características – Para a seleção da melhor características foi utilizada a função 'SelectKBest' do sklearn, selecionando as 3 melhores características. Na função 'SelectKBest' foi utilizada uma função de análise de variância disponível no sklearn, chamada 'f_regression'. Assim, as características selecionadas com o uso desse método foram:

Active Mean Active Max Active Min

- Seleção com análise de correlação – Para seleção de características baseado na análise da correlação foi utilizado o critério padrão de 0.20 de correlação com o alvo. Obtendo, assim, as seguintes colunas a serem utilizadas:

Active Mean Active Max Active Min

- Seleção baseado em árvore de decisão – Para seleção de características baseado em árvore de decisão foi utilizado o algoritmo de classificação DecisionTreeClassifier disponível através do sklearn e utilizando o critério de 0.1 de importância da características para sua seleção. Nesse contexto foram selecionadas 3 características, sendo elas:

Source IP Bwd IAT Std Active Min

4.2. Resultados da Avaliação dos Métodos

Após as características selecionadas, foi feita a separação do dataset com dois conjuntos de dados, um conjunto com as características; e outro com o alvo da pesquisa, ou seja, o *DoS Slowhttptest*. Assim, foi aplicado o algoritmo RandomForestClassifier com a utilização do sklearn ensemble para classificar cada um dos métodos utilizados, obtendo os resultados apresentados na Tabela 4, onde é apresentada a acurácia e tempo de processamento dos métodos de seleção de características.

Tabela 4. Classificação dos métodos

Método	Acurácia	Tempo
Todas características	99.999%	253.35s
Seleção das melhores características	99.747%	31.57s
Seleção com análise de correlação	99.682%	28.14s
Seleção baseado em árvore de decisão	99.682%	28.42s

5. Considerações Finais

Este trabalho comparou e avaliou diferentes métodos de seleção de características para detectar anomalias geradas por ataques de inundação HTTP. Entre os métodos utilizados na pesquisa, a seleção das melhores características obteve uma acurácia melhor. Contudo, mais estudos sobre melhores parâmetros e erros gerados com cada um dos métodos ainda se faz necessário, assim como experimentar outros métodos de seleção de características. Nesse contexto, também deve ser levado em conta o propósito para o qual será utilizado, uma vez que para essa classificação foi utilizado como alvo o ataque *DoS Slowhttptest*, e não qualquer tipo de anomalia presente em redes de computadores.

Referências

- Al-gethami, W. and Aljuhani, A. (2022). Detection of http attacks using machine learning. In *2022 2nd International Conference on Computing and Information Technology (ICCIT)*, pages 344–348.
- Bhargava, R., Pal Singh, Y., and Narawade, N. S. (2022). Implementation of machine learning based ddos attack detection system. In *2022 3rd International Conference for Emerging Technology (INCET)*, pages 1–5.
- Dalmazo, B. L., Marques, J. A., Costa, L. R., Bonfim, M. S., Carvalho, R. N., da Silva, A. S., Fernandes, S., Bordim, J. L., Alchieri, E., Schaeffer-Filho, A., Paschoal Gaspar, L., and Cordeiro, W. (2021). A systematic review on distributed denial of service attack defense mechanisms in programmable networks. *International Journal of Network Management*, 31(6):e2163.
- Dalmazo, B. L., Vilela, J. P., and Curado, M. (2018). Triple-similarity mechanism for alarm management in the cloud. *Computers & Security*, 78:33–42.
- Najafabadi, M. M., Khoshgoftaar, T. M., Calvert, C., and Kemp, C. (2017). User behavior anomaly detection for application layer ddos attacks. In *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, pages 4–16.
- Sharafaldin, I., Lashkari, A. H., and Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, 1:108–116.
- Simon, K. (2022). Digital 2022: October global statshot report. data reportal. Disponível em: <https://www.amper.ag/post/we-are-social-e-hootsuite-digital-2022-resumo-e-relatorio-completo>. Acesso em: 10 de agosto de 2023.