Optimizing DDoS Attack Detection through Machine Learning based on Feature Selection Strategies

Department of Computer Science, Stephen F. Austin State University

[1]Oluwatosin Islamiyat Yusuf, Faculty Mentor: [2*]Dr. Ivancic Christopher

[1]Email: yusufoi@jacks.sfasu.edu, [2*]Email: ivancic@sfasu.edu

ABSTRACT

This study explores the optimization of Distributed Denial of Service (DDoS) attack detection through the utilization of 25 carefully selected machine learning features. Experiments are conducted to assess the effectiveness of prominent machine learning models, including Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM), using the NSL-KDD dataset. This dataset is renowned for its comprehensive coverage of various cyber-attacks, providing labeled network flows, full packet payloads in pcap format, and CSV files containing profiles and labeled flows for machine learning analysis. The investigation focuses on three key metrics: precision, accuracy, and F-score, to evaluate the proficiency of the models in detecting DDoS attacks. The analysis reveals significant proficiency across all models, with Random Forest, XGBoost, SVM, and LSTM achieving average scores of 99.9993%, 99.9778%, 98.9501%, and 99.5858%, respectively. The findings of this study highlight the efficacy of machine learning approaches in enhancing DDoS attack detection capabilities, offering promising avenues for bolstering cybersecurity defenses.

INTRODUCTION

DDoS attacks pose a significant threat to cybersecurity, capable of overwhelming network resources and causing extensive damage. Traditional detection systems often fall short against the sophistication of modern DDoS attacks, necessitating advanced solutions. DDOS attacks use an overload approach in which a massive number of requests are directed at a target computer system or website, making all incoming traffic indecipherable for the attacked system. Extremely high volumes of traffic could make the system slow down or the hacker take over, which deny legal users access and take over the original situation of managing from administrators. This project seeks to address these challenges by optimizing DDoS attack detection through the implementation of machine learning models enhanced by strategic feature selection strategies, aiming to bolster cybersecurity defenses and ensure the resilience of digital infrastructures against the DDoS attacks.

RESEARCH OBJECTIVE

The objective of the paper is to develop and enhance the detection of Distributed Denial of Service (DDoS) attacks using machine learning techniques. The research aims to enhance the accuracy, precision, F-score of DDoS attack detection by applying feature selection approaches on various classifiers such as Random Forest, Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) on NSL-KDD dataset.

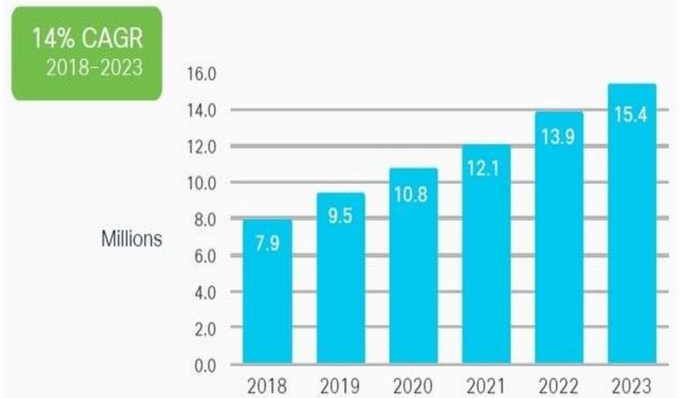

Figure 1. Global DDoS Attack Trends

Figure 1 indicates a significant and steady increase in the frequency of Distributed Denial of Service (DDoS) attacks globally from 2018 through an estimated figure in 2023

METHODOLOGY

RESULTS

Our analysis centered on three primary metrics: precision, accuracy, and F-score. The findings are summarized as follows:

- ❑ The Random Forest model demonstrated precision with an average score of 99.9993%, signifying its exceptional.
- ❑ The XGBoost model followed closely with a score of 99.9778%
- ❑ The result of the LSTM network performance is 99.5858%.
- ❑ The SVM model, while slightly less effective than its counterparts, still performed notably well with a precision of 98.9501%.

RESULTS CONT'D

Figure 2 The screenshot displays classification reports for four different machine learning models: Random Forest, SVM (Support Vector Machine), XGBoost, and LSTM (Long Short-Term Memory). Each model's performance is evaluated based on precision, recall, F1-score, and support metrics for two classes (labeled as "0" and "1"

The heatmap in Figure 3 represents the correlation matrix of the features. Each cell in the grid represents the correlation coefficient between two features. The closer the value is to 1 or -1, the stronger the correlation. Positive values indicate a positive correlation, while negative values indicate a negative correlation.

Preliminary findings show the confusion matrix of 3 models and its relevance towards feature selection.

Random Forest Classification Report:

	precision	recall	f1-score	support
0	0.999897	1.000000	0.999948	19405.000000
1	1.000000	0.999922	0.999961	25744.000000
accuracy	0.999956	0.999956	0.999956	0.999956
macro avg	0.999948	0.999961	0.999955	45149.000000
weighted avg	0.999956	0.999956	0.999956	45149.000000

SVM Classification Report:

	precision	recall	f1-score	support
0	0.998867	0.999382	0.999124	19405.000000
1	0.999534	0.999145	0.999340	25744.000000
accuracy	0.999247	0.999247	0.999247	0.999247
macro avg	0.999200	0.999264	0.999232	45149.000000
weighted avg	0.999247	0.999247	0.999247	45149.000000

XGBoost Classification Report:

	precision	recall	f1-score	support
0	0.999948	1.000000	0.999974	19405.000000
1	1.000000	0.999961	0.999981	25744.000000
accuracy	0.999978	0.999978	0.999978	0.999978
macro avg	0.999974	0.999981	0.999977	45149.000000
weighted avg	0.999978	0.999978	0.999978	45149.000000

LSTM Classification Report:

	precision	recall	f1-score	support
0	0.999639	0.999639	0.999639	19405.00000
1	0.999728	0.999728	0.999728	25744.00000
accuracy	0.999690	0.999690	0.999690	0.99969
macro avg	0.999684	0.999684	0.999684	45149.00000
weighted avg	0.999690	0.999690	0.999690	45149.00000

Figure 2. Model Classification Reports

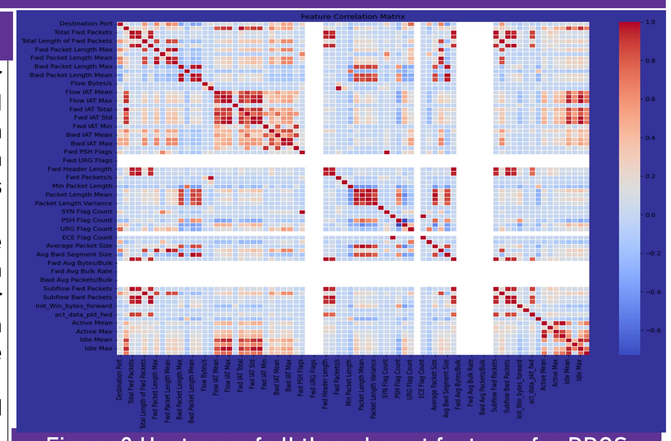

Figure 3 Heatmap of all the relevant features for DDOS

CONCLUSION

The utilization of an intrusion detection system is an effective technique for proactively identifying and mitigating both known and unknown attacks, thereby safeguarding network devices from potential harm caused by attackers. In this study, the significance of using a set of relevant features with an adequate classification learning algorithm for modeling DDOS detection has been demonstrated. Future research should focus on expanding the dataset diversity, exploring additional machine learning algorithms, and integrating the developed models into real-world cybersecurity systems for comprehensive testing and validation

REFERENCES

1. Abdulrahman, Mahmood K. Ibrahem. Evaluation Of DDoS Attacks Detection in A Cicids2017 Dataset Based on Classification Algorithms.
2. Ahmad, I. (2015). Feature selection using particle swarm optimization in intrusion detection. International Journal of Distributed Sensor Networks, 11(10), 806–954.
3. Azmi et al. (2021). Detecting DDS attacks using machine learning algorithm.
4. Khundrakpam, Johnson Singh, and Tanmay De. Efficient Classification of DDoS Attacks Using an Ensemble Feature Selection Algorithm.
5. Mustafa et al. (2023). Intelligent Distributed Denial of Service Attacks Detection (IDDOSAD) Approach.

ACKNOWLEDGEMENTS

I am sincerely grateful to the organizers of the CCSC conference for the invaluable opportunity to present these research findings.

I would like to express my gratitude to my supervisor, Dr. Christopher Ivancic for all his guidance and support through out the project.