GEORGE WASHINGTON UNIVERSITY

ADA UNIVERSITY

COMPUTER SCIENCE AND DATA ANALYTICS

GUIDED RESEARCH I

Report 5

Asiman Mammadzada

Title: Network Intrusion Detection System using Machine Learning

Instructors: Prof. Dr. Stephen Kaisler, Assoc. Prof Dr. Jamaladdin Hasanov

Introduction

In the constantly evolving landscape of cybersecurity, the need to detect and mitigate network intrusions has been more critical. Traditional Network Intrusion Detection Systems (NIDS) have been instrumental in safeguarding networks but are increasingly challenged by sophisticated and newly developed attack patterns. The integration of Machine Learning (ML) into NIDS offers an innovative solution to these challenges, providing adaptive mechanisms to identify, analyze, and respond to potential threats.

The application of ML to NIDS represents combination of two highly dynamic fields: artificial intelligence and cybersecurity. This combination helps the system to learn from historical data and to predict potential intrusions, surpassing the limitations of conventional signature-based detection rules. By identifying patterns and correlations in network traffic, ML-powered NIDS can detect anomalies and unknown attack types that would help escape traditional detection methods.

However, the implementation of ML within NIDS is complex task. Issues related to data quality, model interpretability, and the fine balance between sensitivity present challenges that must be considered.

This report represents the findings in the fascinating intersection of ML and NIDS, mentioning the fundamental principles, methodologies, benefits, and challenges. Through an in-depth examination, I seek to understand how ML enhances the capabilities of NIDS, offering new prospects in network security, and highlighting the practical considerations that must be navigated for successful deployment.

Final Results and Achievements:

As mentioned, traditional Network Intrusion Detection Systems (NIDS) often find difficulties to keep pace with sophisticated attack patterns or detect newly typed patterns. The project has addressed challenge by integrating Machine Learning (ML) into NIDS sponsored dataset from KDDCUP data, specifically employing the Random Forest algorithm, which has proven to be highly effective. Variety of foundational ML algorithms has been tested, such as Logistic Regression, Decision Tree, XGBoost, but Random Forest is the one that produced the most satisfactorily results. Through this approach, the system has achieved a 98% accuracy rate in predicting whether a given network activity is an attack or normal behavior. This initial

classification helps in recognizing potential threats types, specifying the type of the attach and lays the groundwork for further analysis. Once an activity has been found or recognized as an attack, the system's capabilities extend to classifying the specific type of intrusion, achieving an impressive 99% accuracy in the task. The types of attacks identifiable by the system include not only common forms such as back, satan, ipsweep, portsweep, warezclient, and teardrop but also additional attack types like smurf, Neptune, pod, and nmap that has been recently added. The success of the Random Forest algorithm in this context is attributed to its conversion of the rule-based approach into tree-based ML rules, encapsulating the intricacies of network traffic patterns and allowing for distinctions between different behaviors and attack types. This ensemble learning method is known for its robustness, making it an ideal choice for the complex task of intrusion detection. The project's extension to detect a broader range of attack types demonstrates a forward-looking approach that anticipates and adapts to the evolving landscape of cybersecurity. This work represents a significant advancement in network security through the application of ML, contributing a robust and comprehensive tool to the arsenal of defenses against cyber threats. By achieving high accuracy in both detecting attacks and identifying their specific types, it stands as a testament to the potential of modern AI techniques in enhancing cybersecurity measures, reflecting the possibilities of a more resilient digital future.

Future Work

As a crucial next step in the project's evolution, it is essential to focus on expanding the dataset utilized for training the model. The current success, though significant, underscores the importance of a well-rounded and comprehensive dataset that includes the majority of attack types. By incorporating a broader labeled attack types, the model's understanding of different intrusion techniques can better be well-developed. This, in turn, would likely lead to even more detection capabilities. The inclusion of a wider variety of attack patterns in the training data would enable the system to generalize better across different scenarios, making it stronger against evolving cyber threats. In re-training the model with this expanded dataset, the expectation is that the results would further improve, pushing the boundaries of what is achievable in the network security through machine learning.