Study of NSL-KDD Dataset and Application of Machine Learning Methods in Detecting and Classifying Various Types of Network Attacks

Lê Đức Dũng
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam
student490@ictu.edu.vn

***Abstract*—This thesis examines the NSL-KDD dataset and applies machine learning methods to detect and classify various types of network attacks. The author analyzes the dataset to identify its suitability for comparing different intrusion detection models and evaluates the performance of various classification algorithms. The results show that the NSL-KDD dataset is ideal for this purpose, with 41 features used to evaluate intrusion samples. Random Forest algorithm demonstrates high accuracy in both cases with and without feature reduction. The study also highlights the need for feature reduction to improve system performance and reduce detection time.**

***Index Terms*—NSL-KDD, machine learning, intrusion detection, Random Forest, feature reduction**

I. Introduction

The thesis provides an overview of the NSL-KDD dataset, its history, and applications. The NSL-KDD dataset is a well-known dataset used for evaluating intrusion detection systems. It contains 41 features and 5 types of attacks, including normal, DoS, Probing, R2L, and U2R. The dataset is designed to be a more balanced and complete representation of network attacks compared to the original KDD Cup 99 dataset. The author discusses the importance of feature reduction in improving the performance of intrusion detection systems, as many features in the dataset are redundant and do not contribute to the detection process [1]. Earlier ICTU student work on related designs is included here for context [2]. Related published studies provide context for the subject and methods [3], [4], [5], [6], [7], [8], [9], [10]. These contextual works were selected for this English summary; they are not asserted to appear in the original thesis bibliography.

II. Methods

The author analyzes the NSL-KDD dataset using various classification algorithms, including Random Forest, to detect and classify network attacks. The study evaluates the effectiveness of these algorithms in reducing false positives and improving detection speed. The dataset is preprocessed to remove redundant features, and the remaining features are used to train and test the classification models. The Random Forest algorithm is chosen for its ability to handle high-dimensional data and provide accurate predictions. The study also explores the impact of feature reduction on the performance of the classification models [1].

III. Results

The analysis of the NSL-KDD dataset reveals that it is well-suited for comparing different intrusion detection models. The use of all 41 features in the dataset helps in evaluating intrusion samples, reducing detection time and system performance degradation. Random Forest shows high accuracy in both cases with and without feature reduction. The study demonstrates that feature reduction can significantly improve the performance of the classification models, reducing the number of false positives and speeding up the detection process [1].

IV. Discussion

The study suggests that Random Forest is a suitable algorithm for intrusion detection systems, especially when feature reduction is applied. Future research should focus on developing new techniques to enhance the effectiveness of IDS using the NSL-KDD dataset and collecting additional data from different environments to better reflect real-world network threats. The author notes that the NSL-KDD dataset is a valuable resource for researchers and practitioners in the field of network security, providing a comprehensive and balanced representation of network attacks [1].

V. Conclusion

The thesis concludes that the NSL-KDD dataset is ideal for comparing different intrusion detection models. The author recommends further research to improve the effectiveness of IDS and suggests that Random Forest can be a valuable tool for high-speed network applications. The study highlights the importance of feature reduction in improving the performance of intrusion detection systems and reducing the number of false positives [1].

Supervisor: Nguyễn Đức Bình, Thai Nguyen University of Information and Communication Technology (ndbinh@ictu.edu.vn).

References

[1] Lê Đức Dũng, "Study of NSL-KDD Dataset and Application of Machine Learning Methods in Detecting and Classifying Various Types of Network Attacks," internal undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2024.

[2] ICTU Student 102, ”Development of an Intrusion Detection System Based on Data Mining Techniques,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2009.
[3] Xianwei Gao et al., ”An Adaptive Ensemble Machine Learning Model for Intrusion Detection,” IEEE Access, 2019. doi:10.1109/access.2019.2923640.
[4] Surbhi Solanki et al., ”A Survey on Machine Learning based Intrusion Detection System on NSL-KDD Dataset,” International Journal of Computer Applications, 2020. doi:10.5120/ijca2020920343.
[5] Mohammed Zakariah et al., ”Intrusion Detection System with Customized Machine Learning Techniques for NSL-KDD Dataset,” Computers, Materials & Continua, 2023. doi:10.32604/cmc.2023.043752.
[6] Ravipati Rama Devi et al., ”Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD-99 and NSL-KDD Datasets - A Review Paper,” International Journal of Computer Science and Information Technology, 2019. doi:10.5121/ijcsit.2019.11306.
[7] Razan Abdulhammed et al., ”Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection,” Electronics, 2019. doi:10.3390/electronics8030322.
[8] Sandeep Gurung et al., ”Deep Learning Approach on Network Intrusion Detection System using NSL-KDD Dataset,” International Journal of Computer Network and Information Security, 2019. doi:10.5815/ijcnis.2019.03.02.
[9] Shadman Latif et al., ”Investigation of Machine Learning Algorithms for Network Intrusion Detection,” International Journal of Information Engineering and Electronic Business, 2022. doi:10.5815/ijieeb.2022.02.01.
[10] Fuat TÜRK, ”Analysis of Intrusion Detection Systems in UNSW-NB15 and NSL-KDD Datasets with Machine Learning Algorithms,” Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 2023. doi:10.17798/bitlisfen.1240469.