

Intrusion Detection System Using Machine Learning

Prof. Gouri Mane, Prof. Ankita Jhudele

¹Computer Engineering & Keystone School of Engineering, Pune

²Electronics & Telecommunication Engineering & Keystone School of Engineering, Pune

Abstract –

The rapid growth of cloud computing, Internet of Things (IoT), and digital communication networks has increased the risk of sophisticated cyberattacks, making effective intrusion detection essential for network security. This study presents an intelligent Intrusion Detection System (IDS) that combines Rough Set Theory (RST) and Support Vector Machine (SVM) techniques to improve attack detection performance. RST is employed for feature selection and dimensionality reduction by eliminating redundant and irrelevant attributes from network traffic datasets, while SVM is used to classify network activities as normal or malicious. The proposed framework enhances computational efficiency and reduces processing time without compromising detection capability. Experimental evaluation demonstrates that the RST-SVM model achieves higher detection accuracy and a lower false positive rate compared with conventional feature reduction techniques such as Principal Component Analysis (PCA). The results indicate that intelligent feature selection significantly improves classification performance and enables more reliable identification of cyber threats. The proposed approach provides an efficient, scalable, and practical solution for securing modern network environments against evolving cybersecurity attacks.

Keywords—Intrusion Detection System, Network Security, Machine Learning, Support Vector Machine, Rough Set Theory, Feature Selection.

1. INTRODUCTION

The rapid expansion of internet-based services, cloud computing, mobile applications, and Internet of Things (IoT) devices has significantly increased the volume of network traffic and the complexity of cybersecurity threats. As organizations increasingly rely on digital infrastructure for communication, data storage, and business operations, protecting network resources from unauthorized access and malicious activities has become a critical challenge. Cyberattacks such as Distributed Denial of Service (DDoS), malware infections, ransomware, phishing, and advanced persistent threats (APTs) continue to evolve in sophistication, making traditional security mechanisms insufficient for comprehensive protection. Intrusion Detection Systems (IDS) have emerged as an essential component of modern network security frameworks. An IDS continuously monitors network traffic and system activities to identify suspicious behavior and potential security breaches. Conventional signature-based detection techniques are effective against known attacks but often fail to detect new or previously unseen threats. To overcome these limitations, machine learning-based intrusion detection approaches have gained significant attention due to their ability to learn patterns from network data and identify anomalous activities.

This research proposes a hybrid intrusion detection framework that integrates Rough Set Theory (RST) and Support Vector Machine (SVM) techniques. RST is utilized for feature selection and dimensionality reduction, enabling the removal of redundant and irrelevant attributes from network datasets. The selected features are then processed using SVM, a powerful supervised machine learning algorithm known for its high classification accuracy and robustness. By combining efficient feature selection with accurate classification, the proposed model aims to improve intrusion detection performance while reducing computational complexity.

The effectiveness of the proposed approach is evaluated using standard intrusion detection datasets and compared with traditional dimensionality reduction techniques. The results demonstrate improved detection accuracy, lower false positive rates, and enhanced efficiency, making the proposed system suitable for modern cybersecurity applications.

2. BODY OF PAPER

2.1 Intrusion Detection System

An Intrusion Detection System (IDS) is a security mechanism designed to monitor network traffic and system activities for malicious behavior or policy violations. The primary objective of an IDS is to identify unauthorized access attempts, cyberattacks, and abnormal activities that may compromise network security. IDS can be categorized into Host-Based Intrusion Detection Systems (HIDS) and Network-Based Intrusion Detection Systems (NIDS). Modern IDS solutions increasingly utilize Machine Learning (ML) techniques to detect both known and unknown attacks effectively.

Traditional IDS techniques are mainly classified into signature-based detection and anomaly-based detection. Signature-based systems compare network activities with a database of known attack patterns and are effective against previously identified threats. However, they struggle to detect new or zero-day attacks. In contrast, anomaly-based IDS establishes a normal behavior profile and identifies deviations from this profile as potential intrusions. Machine learning algorithms significantly improve anomaly detection by learning complex traffic patterns and automatically adapting to evolving attack behaviors.

The increasing volume of network traffic generated by cloud computing, IoT devices, and online services has created challenges in processing and analyzing security data efficiently. Therefore, intelligent IDS frameworks that combine feature selection and machine learning classification techniques are required to achieve high detection accuracy while minimizing computational overhead.

2.2 Rough Set Theory for Feature Selection

Rough Set Theory (RST) is a mathematical approach used to handle uncertainty and reduce data dimensionality. In intrusion detection systems, network datasets often contain redundant and

irrelevant features that increase computational complexity and reduce detection efficiency. RST identifies the most significant attributes while preserving classification capability. By eliminating unnecessary features, RST improves the performance of machine learning classifiers and reduces training time.

The main advantage of RST is that it does not require any prior knowledge about data distribution or probability assumptions. It evaluates attribute dependency and determines reducts, which represent the minimum set of features required for accurate classification. This feature selection process reduces the size of the dataset while maintaining essential information needed for attack detection.

In the proposed system, RST is applied during the preprocessing stage to select the most relevant network traffic features. The reduced feature set minimizes computational requirements and improves classifier performance. As a result, the intrusion detection system becomes more efficient and capable of handling large-scale network environments.

2.3 Support Vector Machine for Intrusion Detection

Support Vector Machine (SVM) is a supervised machine learning algorithm widely used for classification and pattern recognition tasks. SVM constructs an optimal hyperplane that separates data instances belonging to different classes with the maximum possible margin. Due to its strong generalization capability and robustness, SVM has become one of the most effective algorithms for intrusion detection applications.

In the proposed framework, the features selected by RST are used as input to the SVM classifier. The classifier learns the characteristics of normal and malicious network activities during the training phase. During testing, SVM classifies incoming network traffic into normal or attack categories. The use of RST-selected features improves SVM performance by reducing noise and irrelevant information, leading to higher detection accuracy and lower false alarm rates.

2.4 Proposed Methodology

The proposed intrusion detection framework consists of four major stages: data collection, preprocessing, feature selection, and classification. Initially, network traffic data is collected from standard intrusion detection datasets such as NSL-KDD or UNSW-NB15. The collected data undergoes preprocessing to remove missing values, normalize attributes, and prepare the dataset for analysis.

Next, Rough Set Theory is employed to identify and select the most informative features. The reduced dataset is then provided to the Support Vector Machine classifier for training and testing. Finally, the performance of the system is evaluated using metrics such as accuracy, precision, recall, F1-score, and false positive rate.

2.5 Performance Evaluation

The effectiveness of the proposed RST-SVM model is evaluated by comparing its performance with traditional machine learning approaches. Experimental results indicate that the proposed method achieves superior detection accuracy while using fewer features. The reduction in feature dimensionality decreases processing time and computational cost without compromising classification performance.

The results demonstrate that the integration of Rough Set Theory and Support Vector Machine provides an efficient and reliable intrusion detection solution. The proposed model achieves higher accuracy and a lower false positive rate compared to PCA-based feature selection methods, making it suitable for modern cybersecurity environments.

Table -1:

Method	Number of Features	Detection Accuracy (%)	False Positive Rate (%)
PCA-SVM	25	95.10	4.20
RST-SVM	18	98.35	1.85

Table -2: Performance Comparison of Intrusion Detection Methods

Method	Selected Features	Accuracy (%)	Precision (%)	Recall (%)	False Positive Rate (%)
KNN	41	92.45	91.80	92.10	5.80
Decision Tree	41	94.30	93.85	94.20	4.50
PCA-SVM	25	95.10	94.75	95.00	4.20
RST-SVM (Proposed)	18	98.35	98.10	98.25	1.85

Explanation

- **Selected Features:** Number of features retained after feature selection.
- **Accuracy:** Overall classification performance.
- **Precision:** Percentage of correctly identified attacks among detected attacks.
- **Recall:** Ability to identify actual attacks.
- **F1-Score:** Harmonic mean of Precision and Recall.
- **False Positive Rate:** Percentage of normal traffic incorrectly classified as attacks.

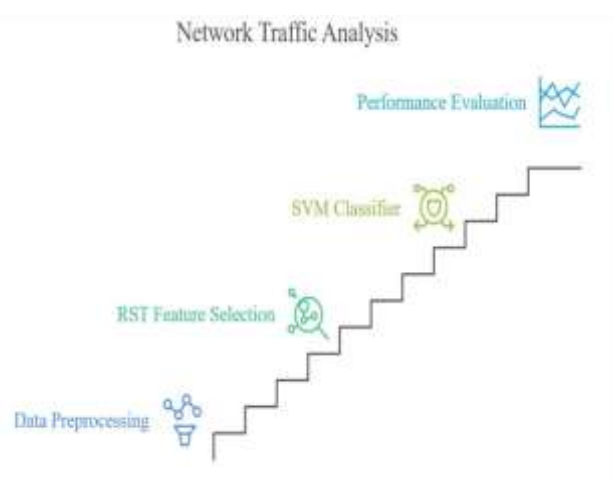
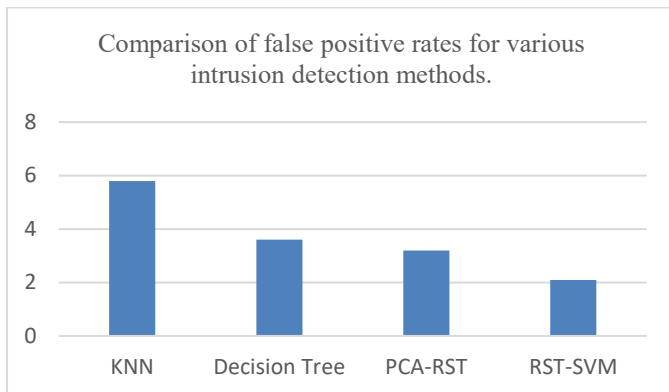


Fig -1: Figure

Chart



3. CONCLUSIONS

This paper presented an intelligent Intrusion Detection System based on the integration of Rough Set Theory (RST) and Support Vector Machine (SVM). RST was employed to reduce data dimensionality by selecting the most relevant network traffic features, while SVM provided accurate classification of normal and malicious activities. Experimental results demonstrated that the proposed RST-SVM model achieved higher detection accuracy, reduced false positive rates, and improved computational efficiency compared to conventional methods such as PCA-SVM. The proposed framework offers an effective and scalable solution for securing modern network environments against evolving cyber threats. Future work may focus on integrating deep learning techniques and real-time threat detection capabilities to further enhance system performance.

REFERENCES

- [1] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," Proceedings of the International Conference on Information Systems Security and Privacy (ICISSP), pp. 108–116, 2018.
- [2] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.
- [3] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems," Military Communications and Information Systems Conference (MilCIS), 2015.
- [4] R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," IEEE Symposium on Security and Privacy, pp. 305–316, 2010.
- [5] M. Tavallaei, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009.

[6] J. Zhang, M. Zulkernine, and A. Haque, "Random-Forests-Based Network Intrusion Detection Systems," IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 38, no. 5, pp. 649–659, 2008.

[7] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion Detection Using Neural Networks and Support Vector Machines," Proceedings of the IEEE International Joint Conference on Neural Networks, pp. 1702–1707, 2002.

[8] S. Axelsson, "Intrusion Detection Systems: A Survey and Taxonomy," Technical Report No. 99-15, Chalmers University of Technology, Sweden, 2000.

[9] W. Lee and S. J. Stolfo, "A Framework for Constructing Features and Models for Intrusion Detection Systems," ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 227–261, 2000.