International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Efficient Intrusion Detection System Based On Improved Convolution Neural Network

Kokila.S[1],Devipriya.S[2]

Assistant Professor [1], Student [2]

Department of Computer Science and Engineering, Tagore Institute Of Engineering And Technology, Deviyakurichi ,Salem ,Tamil Nadu,India

ABSTRACT :

The intrusion detection systems (IDSs) are essential elements when it comes to the protection of an ICT infrastructure. Intrusion detection systems (IDSs) are widespread systems able to passively or actively control intrusive activities in a defined host and network perimeter. Recently, different IDSs have been proposed by integrating various detection techniques, generic or adapted to a specific domain and to the nature of attacks operating on. This work focus on the network intrusion detection using SVM and neural network. Here SVM classify network behavior into two class first is safe and other is unsafe. Once unsafe network is identified then trained neural network identified attack type of the input sessions. So Whole work is divide into two modules, first is separation of safe and unsafe session from the dataset using SVM. Then in second module identification of type of intrusion is done in unsafe network by EBPNN. This work has proposed SFLANN (Shuffled Frog Leaping and Artificial Neural Network) have three modules first is selection of features from available set of features than second is training of neural network was performed from available set of filtered features. Finally, in third module testing was performed on the trained neural network. Here selection of features was done by Shuffled Frog Leaping Algorithm and training of Error Back Propagation Neural Network was performed. Hence objective of this paper was to reduce number of features with increase intrusion detection accuracy. Experiment was done on real dataset NSL-KDD while comparison was done by existing methods. Results shows that proposed SVM & EBPNN model has increase the precision while accuracy was enhance It was also shows that proposed SFLANN model has increase the precision while accuracy was enhance by 2.03%. This enhancement was achieved by use of SFLANN for initial feature selection model. As this selection of features is done by genetic algorithm, so neural network learning was get improved. Comparison of proposed genetic based feature selection neural network for Subclass Intrusion Detection model was done with Trained Neural Based Subclass Intrusion Detection model.

Keywords – EBPNN , SFLANN

I.INTRODUCTION :

The advent of digital technology has resulted in the generation of vast amounts of data, measured in terabytes every second. Companies across various domains profit by managing resources and transactions over networks. However, ensuring the security and privacy of this big data remains a critical challenge. One of the key issues is detecting and preventing network intrusions with high accuracy and minimal prediction time. Intrusions compromise the confidentiality, integrity, and availability of data resources and services. While firewalls serve as the first line of defense, they are limited to detecting outsider threats and often fail to address complex or insider intrusions, making them insufficient for comprehensive security. To enhance security, Intrusion Detection Systems (IDS) have emerged as essential tools. IDS monitors network traffic and host systems for malicious activities, policy violations, and intrusions. These systems alert administrators when security policies are breached, providing an additional layer of protection. Unlike traditional firewalls, IDS can operate as software or hardware, delivering security through network monitoring, maintaining data integrity, and ensuring service availability. Despite their benefits, IDS technologies face challenges due to the increasing volume of network data and the evolving complexity of attacks, necessitating advancements in their methodologies. Machine Learning (ML) techniques play a significant role in modern IDS by enabling automated analysis of large datasets. ML methods like Support Vector Machines, Bayesian Belief Networks, and Decision Trees facilitate intrusion detection by clustering, predicting, and classifying data elements. However, traditional ML approaches struggle with compatibility issues, high false-positive rates, and limited effectiveness against new attack classes. To overcome these limitations, Deep Learning (DL) has been adopted, offering improved accuracy and computational efficiency. DL algorithms, combined with optimization techniques, enhance IDS performance, as demonstrated in evaluations on datasets such as NSL-KDD and UNSW-NB15. Network attacks can be broadly categorized as active or passive. Passive attacks involve stealthy activities like eavesdropping or traffic analysis, while active attacks aim to disrupt network operations through techniques such as Denial of Service (DoS) or sinkhole attacks. IDS primarily focus on active attacks through strategies of prevention, detection, and mitigation. Prevention aims to block intrusions before they occur, detection identifies ongoing attacks, and mitigation addresses the aftermath of successful intrusions. These strategies are critical for maintaining the security and reliability of network resources. Anomaly-based IDS have gained prominence for their ability to detect previously unknown attacks. By monitoring statistical patterns of network behavior, these systems flag deviations as potential intrusions. However, the dynamic nature of network behavior requires periodic updates to maintain accuracy, adding overhead to the system. Anomaly-based IDS employ techniques like statistical analysis, rule-based expert systems, and clustering to differentiate between normal and malicious activities. Machine Learning enhances these systems by creating adaptable models that improve detection accuracy over time.

Host-Based Intrusion Detection Systems (HIDS) focus on evaluating internal host activities and collecting data from system logs, file systems, and network events. They are particularly effective in monitoring unauthorized access and insider threats. In contrast, Network-Based Intrusion Detection Systems (NIDS) analyze real-time network traffic to identify suspicious patterns. By monitoring packets at various levels, NIDS ensures comprehensive protection without degrading system performance. Partitioning large-scale networks and deploying IDS at segmented levels further strengthens network security. The increasing complexity of attacks and the exponential growth of network data highlight the need for advanced intrusion detection techniques. IDS leveraging ML and DL algorithms offer the potential to automatically extract valuable insights from massive datasets. By integrating classification, clustering, and predictive analytics, these systems enhance their ability to detect and respond to intrusions effectively. Combining IDS with firewalls and other security measures creates a robust defense against both known and emerging threats. Despite significant advancements, IDS technologies face limitations in achieving 100% security due to the ever-evolving threat landscape. Challenges such as high false-positive rates, inadequate training for new attack types, and computational inefficiencies persist. Addressing these issues requires continuous innovation in IDS methodologies, incorporating optimization algorithms and leveraging advancements in AI and data mining. The future of IDS lies in its ability to adapt to changing network environments and emerging security challenges. The integration of IDS with advanced technologies ensures comprehensive protection for big data environments. By addressing the limitations of traditional security measures and incorporating data mining and DL techniques, IDS can detect intrusions earlier and more accurately. As network traffic grows exponentially, the role of IDS in maintaining the confidentiality, integrity, and availability of data becomes indispensable.

II.RELATED WORKS :

Intrusion detection has become a critical component in safeguarding the interconnected digital world, leading to substantial research efforts in this domain. This section delves into recent advancements, with a particular focus on machine learning (ML) and deep learning techniques for intrusion detection. These methodologies aim to enhance precision, efficiency, and adaptability in identifying network anomalies and safeguarding data integrity. Zhao et al. (2015) proposed an ML-based real-time anomaly detection framework that combines batch and real-time processing using tools like Apache Storm and Hadoop. By employing decision tree (DT), support vector machine (SVM), and naïve Bayesian (NB) techniques, the approach achieved a commendable 90.3% accuracy. Similarly, Ghanem et al. (2015) introduced a hybrid strategy using genetic algorithms, demonstrating 96.1% accuracy on the NSL-KDD dataset. These studies highlighted the potential of ML algorithms to process large datasets and deliver high detection rates efficiently. Al-Yaseen et al. (2015) enhanced intrusion detection with a hybrid model combining modified K-means and multi-level SVM. This approach improved training quality while minimizing time, achieving a 95.71% accuracy rate. Meanwhile, Ji et al. (2016) implemented a multi-level method using DWT and SVM, reaching 95.547% accuracy but falling short compared to neural networks. Khalilian et al. (2016) introduced the DCSTREAM approach for large cluster datasets, outperforming existing methods like ConStream in efficiency. However, further analysis of complex data streams remains a challenge. Farnaaz et al. (2016) utilized random forest classifiers for intrusion detection, showing higher effectiveness compared to traditional classifiers. Despite achieving promising results, models like these often emphasize accuracy, as noted by Kala et al. (2017), who advocated for incremental learning and attack updates in future ML models. Park et al. (2017) further refined feature extraction through character-level image transformation, although their approach added computational overhead with minimal performance gains. Repalle et al. (2017) advanced intrusion detection with AI-powered virtual analysts, transforming unsupervised data into supervised formats for improved model adaptability. Raman et al. (2017) developed a hyper graph-based genetic algorithm model using SVMs, achieving 96.72% accuracy on NSL-KDD data. However, both models revealed limitations in handling large datasets and complex input vectors. Innovative semi-supervised frameworks like Yao et al. (2018) integrated hierarchical k-means to improve detection accuracy, achieving a remarkable 99.3% accuracy for known attacks. Meanwhile, Yerima & Sezer (2018) introduced Droid fusion for Android intrusion detection using multiple classifiers. This method reduced training times but struggled with multi-class attack detection.

Similarly, Kabir et al. (2018) utilized the least square support vector machine for binary and multiclass evaluations but faced performance challenges with larger datasets. Recent efforts also include hybrid models like Saleh et al. (2019), which combined KNN and SVM classifiers for 95.77% accuracy. Gao et al. (2019) proposed incremental ELM with adaptive PCA for detecting new cyber-attack types with reduced training times. Despite their success, hybrid models often grapple with architectural complexity and shallow learning issues, as seen in Thi-Nga et al. (2020), whose packet assignment algorithm optimized detection time by 23%. Overall, machine learning approaches in intrusion detection demonstrate significant advancements in efficiency and accuracy. However, challenges like high false alarm rates, computational overhead, and scalability persist, paving the way for further exploration in both machine learning and deep learning methodologies.

III.PROPOSED SYSTEM :

The proposed system introduces an innovative multi-level Intrusion Detection System (IDS) framework combining machine learning and deep learning techniques to enhance anomaly detection accuracy and reduce false alarms. It leverages hybrid models that incorporate clustering, classification, and optimization algorithms, including k-means, SVM, random forests, and neural networks. The system integrates advanced data preprocessing, feature selection, and adaptive learning mechanisms to handle large, dynamic datasets effectively.

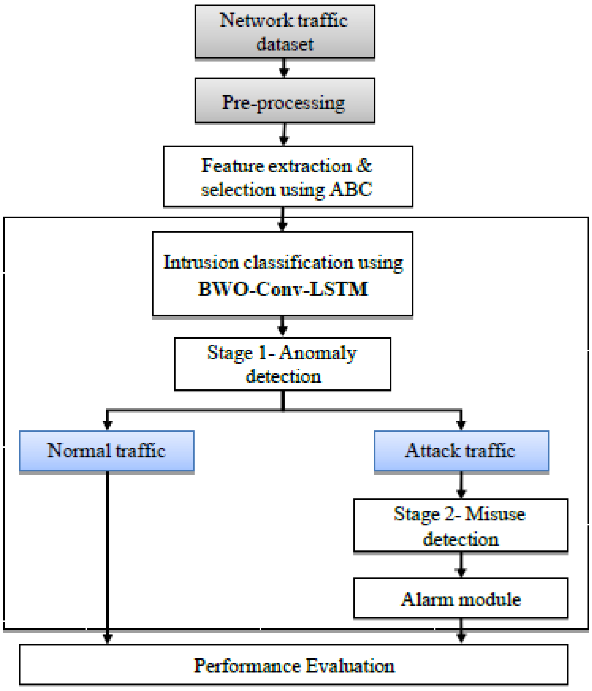
Network traffic dataset
Pre-processing
Feature extraction & selection using ABC
Intrusion classification using BWO-Conv-LSTM
Stage 1- Anomaly detection
Normal traffic
Attack traffic
Stage 2- Misuse detection
Alarm module
Performance Evaluation

Figure 1: System Architecture of proposed system

IV. MODULES :

The implementation of the proposed IDS model begins with data preprocessing, which involves label transformation, duplication removal, and data normalization. Label transformation converts symbolic class labels into numerical representations, facilitating both binary and multi-class classifications. For binary classification, labels are assigned as 0 for normal and 1 for attack, while multi-class classification uses codes like 000 for normal and 001 for DoS. Duplication removal ensures unbiased learning by eliminating redundant data, particularly in datasets like ISCX-IDS. Data normalization scales features to a range of [0, 1] using the formula \(X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \), stabilizing the dataset for effective learning. Feature selection is implemented using the Artificial Bee Colony (ABC) algorithm to identify the most informative features and discard irrelevant ones. This optimization step enhances classification accuracy while reducing computational complexity. The ABC algorithm, inspired by the foraging behavior of honey bees, was chosen for its robustness, faster convergence, and adaptability. By selecting optimal features, the ABC algorithm improves the efficiency and accuracy of hybrid deep learning classifiers such as Conv-LSTM, which analyze complex patterns in intrusion detection datasets. The Conv-LSTM architecture is further optimized using the Bidirectional Weighted Optimization (BWO) algorithm. BWO initializes optimal weights for hidden layers, enabling the network to learn efficiently from the data. This optimization improves both the accuracy and training speed of the classifier. The Conv-LSTM architecture itself combines convolutional layers for spatial feature extraction with LSTM layers for temporal pattern recognition, making it highly effective for intrusion detection tasks. To ensure scalability and handle large-scale datasets, the Conv-LSTM model is integrated with the MapReduce framework. This distributed programming model processes data in parallel across multiple nodes. During the Map phase, data batches and their labels are processed by mappers, which train the system using BWO-optimized weights. The updated weights are then aggregated in the Reduce phase, resulting in global weight updates. This distributed approach accelerates training and enhances the model’s scalability. The proposed system is evaluated on multiple IDS datasets such as NSL-KDD, UNSW-NB15, and ISCX-IDS. Performance metrics like accuracy, precision, recall, and F1-score are used to assess its effectiveness. Comparative studies highlight the superior performance of the MapReduce-based BWO-Conv-LSTM system in detecting intrusions and minimizing false positives. The integration of MapReduce ensures that the system is scalable, efficient, and capable of handling the computational demands of modern intrusion detection environments.

V.RESULTS AND DISCUSSION :

The results demonstrate the effectiveness of the proposed MapReduce-based BWO-Conv-LSTM intrusion detection system across multiple datasets, including NSL-KDD, UNSW-NB15, and ISCX-IDS. The system achieves high classification accuracy, precision, recall, and F1-score, showcasing its ability to accurately detect normal and malicious activities while minimizing false positives. The integration of the Artificial Bee Colony algorithm for feature selection significantly reduces computational complexity and enhances classification performance. Additionally, the use of the MapReduce framework ensures scalability and efficiency in handling large-scale datasets, with distributed training accelerating the learning process. Comparative evaluations with other state-of-the-art methods confirm the superiority of the proposed approach in terms of detection accuracy, computational efficiency, and adaptability to various intrusion detection scenarios.

VI.CONCLUSION :

The proposed IDS model demonstrates an effective solution for intrusion detection by integrating advanced preprocessing, feature selection, and a hybrid deep learning architecture. The use of ABC for feature selection and BWO optimization enhances classification accuracy while reducing complexity. Integration with the MapReduce framework ensures scalability and efficient processing of large-scale datasets. The system's performance on benchmark IDS datasets confirms its robustness and ability to minimize false positives. Overall, the model offers a scalable, accurate, and efficient approach to modern intrusion detection challenges.

REFERENCE :

1. Dean, J., & Ghemawat, S. (2008). "MapReduce: Simplified Data Processing on Large Clusters." Communications of the ACM, 51(1), 107-113. Karaboga, D., & Basturk, B. (2007). "A Powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm." Journal of Global Optimization, 39(3), 459-471. Lyu, X., Zhou, C., & He, J. (2020). "A Hybrid Intrusion Detection System Using CNN and LSTM." Journal of Network and Computer Applications, 164, 102692. Moustafa, N., & Slay, J. (2015). "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems." Military Communications and Information Systems Conference (MilCIS), 1-6. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). "A Detailed Analysis of the KDD CUP 99 Data Set." Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 1-6. Roy, S., Chetry, P., & Sharma, D. K. (2020). "Optimization Techniques in Deep Learning: An Overview." International Journal of Artificial Intelligence and Applications, 11(4), 53-66. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). "A Deep Learning Approach to Network Intrusion Detection." IEEE Transactions on Emerging Topics in Computational Intelligence, 2(1), 41-50. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." Advances in Neural Information Processing Systems (NIPS), 25, 1097-1105. Lecun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning." Nature, 521(7553), 436-444.