International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 10, No. 3, 2023, pp. 17-21.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

A Sophisticated Cybersecurity Intrusion Identification Model Using Deep Learning

Grace Mokoena [a] , Dr. Jonas Nilsson [b]

[a] *University of Pretoria, South Africa*.
[b] *University of Pretoria, South Africa*.

Abstract

The data is vulnerable to numerous attacks during transmission within the network environment. Identifying breaches in network communications is becoming more critical. Scientists employ machine learning methodologies to develop efficient Intrusion Detection Systems (IDS). This paper presents an IDS incorporating preprocessing techniques and a Deep Learning (DL) framework for detecting Denial of Service (DoS) attacks. The research evaluated the proposed model utilizing the dataset commonly referenced in academic literature. The study implemented preprocessing techniques, including feature deletion, random subset choice, choosing features, duplication elimination, and normalizing on the database. Enhanced recognition efficiency was achieved for both training and testing assessments. The test results indicate that the Conventional Neural Network (CNN)-based inception-like model had the highest accuracy, with 98% for binary classification and 97% for multiclass classification among the offered models. The deductive time of the suggested framework for different test data sizes appears favorable compared to baseline models with fewer parameters that can be trained. The proposed IDS system, along with the preprocessing techniques, yields better outcomes when compared with contemporary studies.

Keywords: Cybersecurity, Intrusion Identification, Deep Learning, Security.

DOI: 10.71086/IAJSE/V10I3/IAJSE1026
Received: 17 June 2023/Revised: 21 July 2023/Accepted: 11 August 2023/Published: 28 September 2023

Introduction

The demand for cybersecurity and defense against many cyber-attack forms has escalated significantly (Li & Liu, 2021). The primary cause is the widespread adoption of the Internet of Things (IoT), the significant expansion of networked computers, and the vast array of pertinent applications utilized by individuals or organizations for personal or commercial purposes (Asghari et al., 2019). Across vast networks, cyberattacks—including Denial-of-service (DoS) attacks, computer spyware, and unwanted access—have caused permanent harm and financial losses (Salim et al., 2020). A single ransomware virus inflicted substantial losses on numerous companies and sectors, including finance, healthcare, energy, and educational institutions, resulting in an estimated loss of 7 billion dollars. Additional figures show that, on average, a data breach incurs a cost of 3.8 million USD for the afflicted firm and 8.3 million USD for the United States. Responding every day to modern needs in the digital sphere, the demand for security and defense against various cyber-attack types is growing.

Usually speaking, a cybersecurity system consists of a computer safety system and a network safety system. Although many techniques—including firewalls and encryption—are meant to reduce Internet-based cyberattacks, an Intrusion Detection System (IDS) (Kaja et al., 2019) is more suited to protecting the computer system from outside threats. Finding several types of negative network contact and system usage for prevention is the primary goal of an IDS. Conventional remedies, including antivirus software, fall short in adequately handling the tasks. By watching and analyzing daily operations to find safety concerns or threats—including DoS attacks—an IDS detects hazardous cyber-attack behaviors on a network. An IDS helps identify, evaluate, and classify illegal system actions like deletion, unauthorized access, and changes. Developing an efficient IDS that is so important in modern network safety depends on identifying various cyber-attacks and anomalies within a network, therefore improving system protection.

Related Works

While daily activity is under observation and evaluation to find security concerns or risks, a system to detect intrusions is used to pinpoint harmful cyber-attack activity on a network. Much cybersecurity research has been conducted to identify and stop cyberattacks or incursions. One often used technology in the field of cybersecurity is signature-based network IDS (Otoum & Nayak, 2021). This method is acknowledged and has been somewhat popular, including recent income. Particularly in identifying unknown or zero-day threats, the anomaly-based approach helps the signature-based method spot hitherto unnoticed attacks. This approach tracks network traffic and uses relevant security data to detect attack patterns. Safety event patterns are investigated using several data mining and machine learning techniques to guide decisions (Shi et al., 2017). The main drawback of the anomaly-based approach is its possibility of producing high false alarm rates by identifying hitherto unnoticed system events as anomalies. Priority should be placed on reducing the false favorable rates of an IDS. To solve these issues, a sound artificial intelligence-based detecting system is needed.

The paper presented an online machine learning approach dubbed Extreme Learning Machine (ELM) (Luo et al., 2018) for attack detection using computing to accomplish complex and resource-intensive processing jobs like classification and inspections. Data is gathered on edge nodes and saved on cloud servers.

The researchers have demonstrated that sophisticated education, learning, and categorization processes are executed in the cloud with the assistance of computation. The architecture requires extra resources to facilitate parallel processing when managing extensive data collections. The research reviewed neural-based methodologies for network attack administration, wherein the authors aggregated systems for identifying intrusions utilizing Neural Networks, Deep Learning (DL) (Al-Garadi et al., 2020), and weightless neural systems and assessed them using data sets. They determine the time difficulty and efficacy of the evaluated method and give an evaluation. The study employed a framework for DL for IDS to enhance its effectiveness. The proposed method examined the behavioral and content properties of the data to ascertain the features and data gathered from the network. The system minimizes the building time of the suggested model by utilizing numerous machines and simultaneous training methodologies.

The research analyzed feature selection approaches and their efficacy on recently utilized databases. This study provides concise details about several strategies. The dataset of DoS assaults and the outcomes of several methods were analyzed using association mapping. The duration of feature selection was compared to the training duration using the identical datasets. A performance assessment is provided for several kinds of attacks across multiple databases derived from the feature gathering of the DoS database.

Proposed IDS for DoS Assaults

This section delineates the preparation methods conducted on the dataset. The research introduced the suggested DL algorithms.

Preprocessing Procedures

Fig. 1 illustrates the preprocessing procedure for acquiring the dataset for use in the IDS. Initially, certain features deemed insignificant in the decision-making process are discarded. A random choice is employed to diminish the dataset's size. Duplicate information is eliminated to enhance training efficacy. Min-max and logarithmic normalizing were executed to generate two new sub-datasets for assessment.

The CIC-DDoS-2019 data set contains 51M records encompassing 12 distinct attack types. Upon analysis of the data set, it is determined that certain records contain exclusively zero values. Significant disparities exist between the lowest and highest values, and when assessing network traffic, many attributes are not essential for identifying an attack. The sample counts of the assault kinds inside the 88-featured database. The database comprises 51M cases, including 48M DoS assaults and 3M reputable traffic occurrences. Existing duplicates in this database have been eliminated.

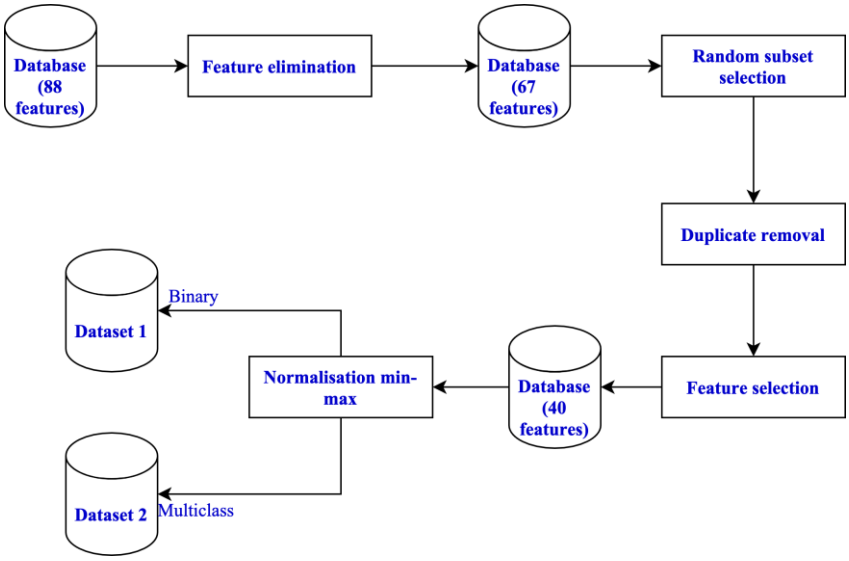


Fig. 1: Preprocessing model

Proposed IDS models

The research conducted performance evaluations of DL models utilizing three distinct baseline models alongside the suggested approach, employing sub-datasets derived from the dataset analyses. The characteristics selected are partitioned into training and testing databases using random selection. Alongside the proposed model, the research analyzed other standard DL methods. In DL designs, selecting appropriate hyperparameters includes network dimensions, layer kinds, and function activation. The types and sizes of layers considerably influence the accuracy of DL systems. The research conducted experiments with several layer classes in the suggested approach, including dense, convolutional, and Long Short Term Memory (LSTM) layers.

Results and Findings

The research conducted studies using a computer equipped with a microprocessor featuring eight cores, 16 GB of primary memory, and a graphics card with 8 GB of storage. The field tests used Ubuntu, Python, and the Keras toolkit to create and train the suggested DL algorithms. The research used the Adam optimization approach, with 500 epochs, a learning rate of $3e^{-4}$, and a batch size of 32. The most effective networks were preserved while evaluating all tested hyperparameters utilizing the Model Checkpoint object from the Keras toolkit.

The Deep Neural Network (DNN) achieved an accuracy of approximately 97%. The Convolutional Neural Network (CNN) baseline approach, configured with 64 neurons in the ConviD levels, achieved an accuracy of roughly 96%. The LSTM foundation model, including two Long Short Term Memory (LSTM) layers, had an accuracy of 95%. CNN-based models yielded more favorable outcomes, even in the most basic scenarios. The suggested model, which

decomposes the input characteristics into bytes, achieved 98% in multi-class assessment. The research employed the same model to achieve the categorization outcome. There is a need to minimize the calculation time of the detection method to facilitate the analysis of a more significant number of packages per minute. Thus, examining the deductive times of the candidate model aids in identifying the real-time constraints of the networks. DNN and CNN algorithms yield detection outcomes more rapidly than models based on LSTM. The overall training duration of a model is contingent upon the number of periods and the amount of computational equipment utilized. The training durations per epoch for the Dense, CNN, and LSTM-based systems using the test hardware are roughly 6 seconds, 12 seconds, and 30 seconds, respectively. Fig. 2 shows the accuracy, and Fig. 3 shows the response time analysis.

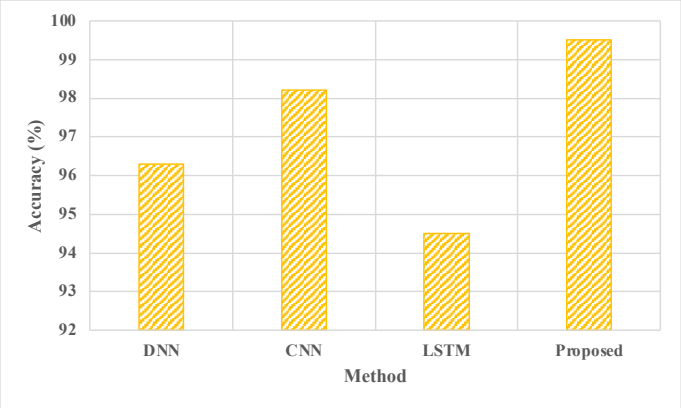


Fig. 2: Accuracy analysis

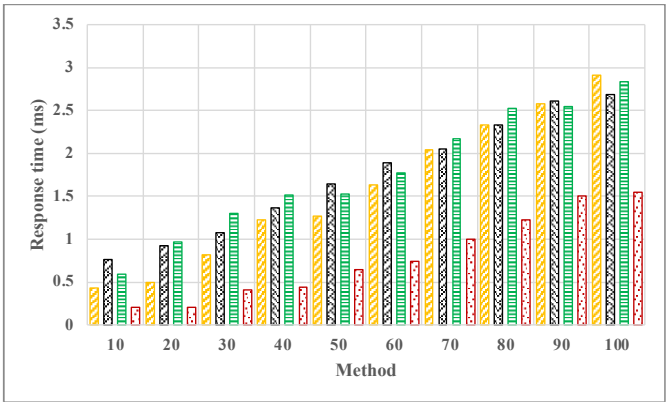


Fig. 3: Response time analysis

Conclusion

This research presents a novel IDS utilizing DL algorithms for DDoS attacks. The research used a dataset comprising 12 classes, including one benign category. The research evaluated multiple DL designs with different units per layer, including DNN, CNN, and LSTM. The research enhanced the system through preprocessing procedures, including feature reduction and selection, identifying 40 significant features from 88. The research acquired a novel homogenous data set by randomly selecting a comparable amount of specimens from each assault type. The research eliminated duplicate information to develop a pristine, non-redundant data set overlooked by most pertinent investigations. This work introduces two novel data sets to the literature that significantly impact the efficacy of training methods derived from the dataset information set. The research implemented min-max normalization methods to assess their impact on efficiency. The research generated relevant data by acquiring a normalized dataset comprising an equal quantity of preprocessed observations from each assault type, devoid of duplicates. The test findings indicate that the proposed model attained accuracy rates of 98% for multi-class and 97% for binary categories. The observations indicate that the suggested CNN-based technique, which incorporates one-dimensional convolutional layers, yields superior results and demonstrates greater accuracy than other contemporary studies. The suggested model exhibits a commendable inference time relative to the original models. The proposed approach effectively identifies diverse attack types for multi- and binary-class assessment.

References

Li, Y., & Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. *Energy Reports*, *7*, 8176-8186.

Asghari, P., Rahmani, A. M., & Javadi, H. H. S. (2019). Internet of Things applications: A systematic review. *Computer Networks*, *148*, 241-261.

Salim, M. M., Rathore, S., & Park, J. H. (2020). Distributed denial of service attacks and its defenses in IoT: a survey. *The Journal of Supercomputing*, *76(7)*, 5320-5363.

Kaja, N., Shaout, A., & Ma, D. (2019). An intelligent intrusion detection system. *Applied Intelligence*, *49*, 3235-3247.

Otoum, Y., & Nayak, A. (2021). As-ids: Anomaly and signature based ids for the internet of things. *Journal of Network and Systems Management*, *29*(3), 23.

Shi, D., Guan, J., Zurada, J., & Manikas, A. (2017). A data-mining approach to identification of risk factors in safety management systems. *Journal of management information systems*, *34*(4), 1054-1081.

Luo, X., Yang, X., Jiang, C., & Ban, X. (2018). Timeliness online regularized extreme learning machine. *International Journal of Machine Learning and Cybernetics*, *9*, 465-476.

Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. *IEEE communications surveys & tutorials*, *22*(3), 1646-1685.