Intrusion Detection System Using Hybrid Model

Imen Chebbi, Ahlem Ben Younes and Leila Ben Ayed

May 12, 2024

Intrusion Detection System Using Hybrid model

*

Imen Chebbi
FSEGS
University of Sfax, Tunisia
chabbiimen@yahoo.fr

Ahlem Ben Younes
ENSIT
University of Tunis, Tunisia
ahlem.benyounes@ensit.rnu.tn

Leila Ben Ayed
ENSI
University of Manouba, Tunisia
leila.benayed@ensi-uma.tn

***Abstract*—Technology and network industries have advanced quickly in recent decades.The expansion of piracy and the compromise of many existing systems has made it essential to develop information security solutions that can identify new threats. In order to address these issues, this work proposes a system called IDS-AI that is based on artificial intelligence techniques. It is capable of detecting recent and dispersed invasions. We employed an auto-encoder for features reduction and two models to assess CNN and SVM in order to evaluate our strategy. The experimental analysis of the dataset demonstrates the suggested model's capability to produce reliable results. On the UNSW-NB15 dataset, our model actually achieves 99.58% and 99.66% accuracy for SVM and CNN, respectively.**

***Index Terms*—Intrusion Detection System, SVM, CNN, UNSW-NB15.**

I. INTRODUCTION

As more people get access to the internet, protecting online identity against threats to secrecy, integrity and accessibility becomes a problem that needs to be solved on a constant basis. In order to identify and classify suspicious activity, a network intrusion detection system (IDS) is a tool that locates and recognizes irregular network traffic.It is an important part of network security and acts as the first line of defense against possible attacks by alerting an administrator or the appropriate individuals to potentially dangerous network behavior. In multiple scholarly works, several artificial intelligence (AI) strategies are put forth for an effective network intrusion detection system (IDS). Finding network traffic abnormalities is growing more challenging as more people gain access to the internet. According to [1] , around 29.3 billion networked devices, or about two-thirds of the world's population, will be online by the beginning of 2023. Online user privacy and security must be protected, and unsafe network conduct must be rapidly discovered and reported. In a word, network intrusion refers to unauthorized, perhaps harmful action on a digital network. An incursion might damage a computer network confidentiality, integrity, and accessibility, which could lead to issues including system compromise and privacy violations. Worms, traffic flooding, buffer overflow attacks, spoofing, and denial-of-service (DoS) attacks are a few examples of network assaults. The two fundamental subcategories of intrusion detection systems are anomaly-based and signature-based systems. Signature-based systems often employ known attack signatures to spot illegal activity. On the other hand, anomaly-based systems are mostly employed in situations when attack signatures are unknown to detect unexpected network behavior. These detection systems make use of several methods, ranging from deep learning models to statistics, to automatically detect suspicious network activities.

This work makes the following noteworthy contributions:
- Data quality: Before applying machine learning algorithms, data must be processed. We used the UNSW-NB15 dataset in this paper.
- Based on Artificial Intelligence approaches, we propose a system for detecting modern and distributed intrusions: IDS-AI .
- We utilized an auto-encoder to reduce the number of features.
- A comparative of the UNSW-NB15 dataset processing with two models Convolutional Neural Network (CNN), Support Vector Machine (SVM).
- A collection of publicly available Python programs for analyzing the dataset and reproducing the studies.
- Comparing our Model with other works in the literature.

The findings of the research serve as the basis for evaluating this learning approach with other publicly available datasets. The remainder of the paper is organized as follow: In Section 2 present a literature review. In Section 3 we describe our proposed approach. The Section 4 present experimental results of our approach. Finally, Section 5 draws our conclusions and plan for future work.

II. RELATED WORK

The expanding adversary capabilities, which influence the dependability of data communication and networks, make computer security concerns especially challenging to manage. An incursion is used to attempt to breach a security goal while also infecting systems. To protect networks and systems from intrusions, a number of tools and techniques, such as intrusion detection systems (IDS), are created [2-4]. By classifying data activity as either normal or intrusive, a collection of techniques known as intrusion detection can be used to spot undesirable behaviors. Techniques for detecting intrusions that take place inside or outside of a monitored network are known as intrusion detection techniques. As a result, there are two basic detection strategies that can be used. The first method is called misuse detection, and it

Identify applicable funding agency here. If none, delete this.

uses a recognized attack pattern to find intrusion. Anomaly detection or behavioral recognition is the second [3-5], and it is based on a break from a normal paradigm. The hybrid detection strategies goal to boost IDS detection efficiency and precision by combining the benefits of both abuse and anomaly detection.

As a result, intrusion detection research remains active and relevant. Nowadays, ML approaches are being used to improve computer security and intrusion detection. Numerous research contributions investigate how machine learning techniques are used in intrusion detection to improve training and data quality and provide dependable, accurate IDS. On the other hand, data are not always collected in an ordered manner. Before being examined, unstructured data needs to be processed. This operation is critical for improving data quality and drawing precise conclusions. Data quality controls are implemented prior to initiating the training and classification operations [6]. Furthermore, feature selection is a desired strategy that seeks to pick the relevant features in order to reduce modeling computational costs and improve predictive model performance [6, 7].

III. Our Suggested Method

In this part, we will demonstrate our IDS-AI technique. Figure 1 depicts the proposed method, which is divided into three main phases: data quality, unsupervised feature learning, and supervised classification. The first is used to remove discrepancies. (Infinity, NaN, and null values). The second employs a deep autoencoder to accomplish unsupervised learning. The auto-encoder is used to reduce the dimensionality of unlabeled data and create a new feature representation. The output of the first phase is then used in the third part to perform supervised learning classification with a Support Vector Machine (SVM) or Convolutional Neural Network. (CNN). More information is given in the subsections that follow.

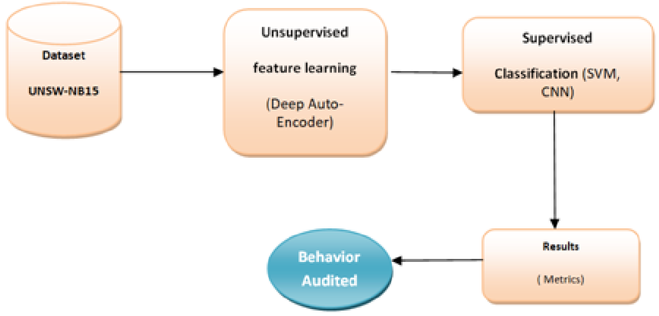

Fig. 1. Our Proposed approach IDS-AI.

Phase1 : Data quality process. The gathering and preparation of data is the phase's main objective. The system then starts a process to gather and assemble the required data from networks. Following data collection, network traffic is subjected to a particular data preparation. Incompatible data types are disregarded during the data preprocessing step after the data have been reviewed. The data has also been cleaned, and the cleaned data has been stored. For this research, we made use of the UNSW-NB15 dataset.

Phase2 : Unsupervised feature learning. When there are several characteristics present, the complexity of the constructed model may increase due to redundancy and noise, which can decrease learner performance. Dimensionality reduction can help to tackle these issues. A feature extraction technique is used to make the necessary adjustments in order to extract the most important features from unlabeled data input. The low-dimensional data generated will then be categorized, allowing the classifier to perform better overall. We utilized an auto-encoder in this phase because it is an unsupervised neural network approach. Its primary purpose is to compress its original input and recreate it as an equivalent reconstructed output.

Phase3 : Supervised classification. The class vector from the initial dataset will be mixed with the newly derived feature space at this stage. As a result, we have a smaller, labeled training set that can be fed into the classifier of our choosing. Several supervised learning algorithms exist, including classification rule neural networks, decision trees, Random Forest, Convolutional Neural Network, and Deep Neural Network . In our work, we chose two supervised classifiers for this task: SVM and CNN.

IV. EXPERIMENTAL RESULTS

TensorFlow was employed as the backend in experiments on Windows 10, while Python and Keras were used for encoding. On the training platform, we train our model, then change the decision criteria and compute the effectiveness measures on the validation set. The production set, a dataset empty of labels, is then used to detect any abnormalities. The final phase is critical because when utilized in practice, the model will create predictions on real-time data without labels against which to compare them. It is always a good practice to hold out a production set to confirm that the model is not performing erratically on this unknown, labelless production set. Table 1 displays the testing results of our model for the UNSW-NB15 dataset [8].

TABLE I
Overall performance (UNSW-NB15 dataset)

Model	Accuracy	F1 score	Recall	Precision
SVM	0.9958	0.9956	0.9962	0.9979
CNN	0.9966	0.9982	0.9978	0.9986

Several studies have evaluated network intruder detection techniques, particularly those based on machine learning and deep learning techniques, using the UNSW-NB15 dataset. We will compare our method to those of four other researchers in this section: Breiman [9], Singh et al. [10], Kabir et al. [11], and Du et al. [12]. Table 2 compares our top UNSW-NB15 dataset results to those of other writers.

TABLE II
COMPARISON WITH OTHER WORKS AVAILABLE IN THE LITERATURE.

Type	ACC(%)
OUR MODEL (SVM)	99.58
OUR MODEL (CNN)	99.66
RF [9]	74.35
FGRNN [10]	99.50
Kabir [11]	96.24
UMAP-RF [12]	92.60

It can be shown that the accuracy of IDS-AI is higher than that of other models. In some instances, the results are global and do not specify which assaults were detected, and some of the models being compared do not include cross validation. Although some additional trials are required to compare our results to those given in the literature, the general results obtained demonstrate the suitability of using IDS-AI to detect intrusions in the dataset. The usage of these methods, while somewhat lowering the produced performance, keeps the findings benchmarked with other related study results.

V. Conclusion and Future Works

Cyberattacks have become more frequent and sophisticated as a result of the expansion of Internet access. In order to improve the security of systems and data, a collection of contemporary techniques known as intrusion detection are used to monitor them. In this study, we show a dependable network intrusion detection method based on artificial intelligence. Preprocessing is being used to improve the discovery rate and precision of IDS based on data heterogeneity. For good data quality, a feature selection procedure based on the entropy choice method is handled before building the model. A revolutionary method is validated by the offered solutions, which guarantee accurate efficiency. The following datasets are used to compare the results: UNSW-NB15. As a consequence, when compared to current models, the new recommended network intrusion detection approach offers various benefits and high accuracy. Future studies will employ additional successful methods to raise our system's detection rate and precision.

References

[1] Cisco Annual Internet Report (2018–2023). URL https://www.cisco.com/c/en/us/solutions/ collateral/executive-perspectives/annual-internetreport/white-paper-c11-741490.html (2018).

[2] G. Fernandes, J. J. P. C. Rodrigues, and L. F. Carvalho, “A comprehensive survey on network anomaly detection,” Telecommunication Systems, vol. 70, pp. 447–489, (2019).

[3] A. Guezzaz, A. Asimi, Z. Tbatou, Y. Asimi, and Y. Sadqi, “A global intrusion detection system using pcapsocks sniffer and multilayer perceptron classifier,” International Journal on Network Security, vol. 21, no. 3, pp. 438–450, (2019).

[4] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges,” Cybersecurity, vol. 2, (2019).

[5] A. Guezzaz, Y. Asimi, M. Azrour, and A. Asimi, “Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection,” Big Data Mining and Analytics, vol. 4, no. 1, pp. 18–24, (2021).

[6] M. Rostami, K. Berahmand, E. Nasiri, and S. Forouzandeh, “Review of swarm intelli-gence-based feature selection methods,” Engineering Applications of Artificial Intelli-gence, vol. 100, Article ID 104210, (2021).

[7] H. Alazzam, A. Sharieh, and K. E. Sabri, “A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer,” Expert Systems with Applications, vol. 148, Article ID 113249, (2020).

[8] Moustafa N. and Slay J., “The evaluation of network anomaly detection systems: statistical analysis of the unsw-nb15 data set and the comparison with the kdd99 data set,” Information Security Journal: A Global Perspective, vol. 25, no. 1-3, pp. 18–31, (2016).

[9] Breiman L., Random forests Machine learning 45(1):5–32. https://doi.org/10.1023/A:1010933404324, (2020).

[10] Singh P., Jaykumar P. J., Pankaj A., Mitra R., Edge-Detect: Edge-centric Network Intru-sionDetection using Deep Neural Network, 2021 IEEE 18th Annual Consumer Communications Networking Conference (CCNC), DOI: 10.1109/CCNC49032.2021.9369469, (2021).

[11] Kabir M.H., Rajib M.S., Rahman A.S.M.T., Rahman M.M., Dey S.K., Network Intrusion Detection Using UNSW-NB15 Dataset: Stacking Machine Learning Based Approach, 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), (2022).

[12] Du, X.; Cheng, C.; Wang, Y.; Han, Z. Research on Network Attack Traffic Detection HybridAlgorithm Based on UMAP-RF. Algorithms 2022, 15, 238. https://doi.org/10.3390/ a15070238, (2022).