A novel ensemble learning system for cyberattack classification

Óscar Mogollón-Gutiérrez
José Carlos Sancho Núñez
Universidad de Extremadura
{oscarmg, jcsanchon}@unex.es

Mar Ávila Vegas
Universidad de Extremadura
mmavila@unex.es

Andrés Caro Lindo
Universidad de Extremadura
andresc@unex.es

***Abstract*—This article introduces a novel approach to enhancing cybersecurity through AI by analyzing network traffic, proposing a two-stage cyberattack classification model to handle class imbalance with a one-vs-rest strategy. It aims to differentiate between legitimate and illegitimate network traffic, employing binary models in the first phase to separate traffic types, and an ensemble model in the second phase for detailed classification. Utilizing the UNSW-NB15 dataset for performance evaluation, the proposed system demonstrates superior results, achieving an F1 score of 0.912 in binary classification and 0.7754 in multiclass classification, outperforming other contemporary methods by 0.75% and 3.54% respectively in F1 score.**

***Index Terms*—intrusion detection, ensemble learning, UNSW-NB15**

Type of contribution: *Research already published [1]*

I. Introduction

Regardless of the nature of stored data in Information Systems (IS), artificial intelligence (AI) can be applied to guarantee or increase security by implementing Intrusion Detection Systems (IDS). Intrusions into IS aim to exploit vulnerabilities to gain access and subsequently make fraudulent use of the systems. These systems aid in identifying malicious actions or variations through the monitoring and classification of network traffic, which is generated by clients attempting to access these systems.

IDS should deal appropriately with the inherent complexity of network data. Class imbalance is a prevalent problem in the field of network traffic intrusion detection, where one class of network traffic (normal traffic) is often much more prevalent than the other class (intrusive traffic).

Our work proposes an intelligent system for cyberattack classification. It consists of a two-stage cyberattack classification ensemble model addressing class imbalance following an OvR approach. This contribution seeks to mitigate the imbalanced learning in network traffic analysis and classification, which is a significant issue in this research field.

II. Related Works

In recent years, the number of studies related to intrusion detection systems has grown considerably. Ensemble models are presented in the literature. A combined voting model is used where the class prediction is obtained from the probabilities of each classifier [2]. Other approaches perform two-stage classification, where, first, a binary classifier determines whether a sample is an attack or not, and then a second process tries to identify the type of attack using classical algorithms [3].

Several datasets that collect the behavior of simulated environments under different attacks are used in the scientific literature, e.g., KDD99 [4], NSL-KDD [5] or UNSW-NB15 [6] . The datasets prior to UNSW-NB15 had several limitations that affected the effectiveness of an IDS in a real environment, and they were not designed with an IOT perspective in mind.

III. Materials

This section presents a description of the dataset, including classes distribution, classification algorithms used in the experiments, and a set of evaluation metrics that enable to evaluating the proposed two-stage cyberattack classification model.

A. Dataset

The UNSW-NB15 dataset was created by the Cyber Range Lab at the University of New South Wales, Canberra (UNSW) with the goal of simulate a heterogeneous environment of legitimate and real attack traffic [19]. The experiments in this research have been performed with this dataset. This version of the UNSW-NB15 contains a total of 44 characteristics, 40 being numeric and 4 being categorical.

B. Algorithms

The construction of the two-stage ensemble classifier involves four key classification algorithms: K Nearest Neighbors (KNN), which uses a majority voting system among the k nearest observations for classification; Support Vector Machine (SVM), aiming to find the optimal hyperplane for class separation; Decision Trees (DT), employing simple decision rules for prediction; and Multilayer Perceptron (MLP), a basic form of artificial neural network with multiple layers. These experiments were executed using Python 3.8, alongside libraries such as Pandas, Numpy, Scikit-Learn, and Imbalanced-learn. The optimal hyperparameters for each algorithm were determined via grid search and 5-Fold cross-validation techniques.

C. Evaluation metrics

In this study, several metrics are chosen to evaluate the performance of intrusion detection models. These are accuracy, precision, recall/detection rate (DR), F1-score and false alarm rate/false positive rate (FAR/FPR), which are calculated as a function of true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs).

IV. Experimental Design

A. Phase 1: Cleaning, Preprocessing and Normalization

As usual, the data must be cleaned and normalized prior to the application of classification algorithms. To this end, the numerical features have been normalized by calculating the standard score so that each feature follows a Gaussian distribution (mean zero and unit variance). On the other hand, categorical features have been encoded using numerical labels.

B. Phase 2: Classifier Models Generation

The next step is a generation of ten binary specialized classification models because there are ten different categories in the dataset. Therefore, each of the generated models can distinguish between samples corresponding to one type of traffic and the rest of the samples. Obtaining each classification model requires the creation of a specific dataset. This model generation process is organized into several subphases to be carried out for each type of traffic collected in UNSW-NB15.

In Phase 2.1, training and test sets are selected for each binary classification model. The training sets used for the generation of binary models are generated in a balanced way.

In Phase 2.2, for hyperparameter tuning, grid search and cross-validation parameters are configured. Next, in Phase 2.3, a model is trained with each combination of parameters. Each generated binary model is evaluated with a specific test set. It consists of samples of the target traffic (positive class) and the rest (negative class) from the original test set. Both Phase 2.2. and Phase 2.3. should be applied for each of the four algorithms studied.

Once the four models have been generated (one for each algorithm) for each type of traffic collected, in Phase 2.4, F1-score is considered to select the binary model with the best performance.

C. Phase 3: Construction of the Final Model

In Phase 3, with the binary models already generated, and after evaluating the individual performance through the evaluation metrics, construction of two-stage ensemble is carried out, organized in two phases, as shown in Fig. 1. The first one is responsible for identifying the behavior as Normal or as an Attack, using the binary model trained for this purpose. In a second level of detection, the possible threat is classified as one of the nine possible attacks collected in UNSW-NB15.

This classification task is performed using an ensemble model composed of the binary classifiers of the nine attack types. Each classifier outputs the probability of belonging to its class and the rest. Taking into consideration the former probability, the class with the highest value is taken as the final prediction.

V. Results and discussion

In binary classification, our model exhibits exceptional performance in binary classification, with a notable accuracy of 92.92% and an F1-score of 91.2%. In multiclass classification, proposed two-stage ensemble model demonstrates a strong capability with an accuracy of 76.05% and an F1-score of 77.54%, outperforming other contemporary works.

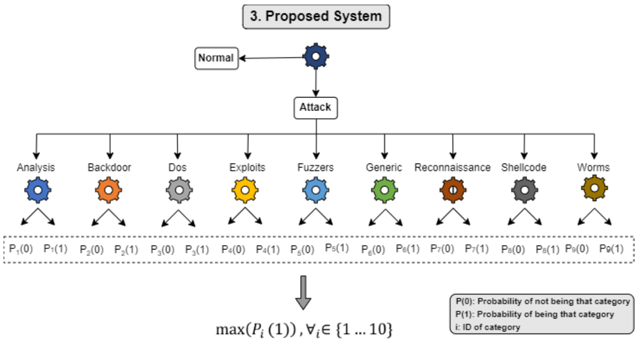
3. Proposed System
Normal
Attack
Analysis
Backdoor
Dos
Exploits
Fuzzers
Generic
Reconnaissance
Shellcode
Worms
$P_1(0)$ $P_1(1)$ $P_2(0)$ $P_2(1)$ $P_3(0)$ $P_3(1)$ $P_4(0)$ $P_4(1)$ $P_5(0)$ $P_5(1)$ $P_6(0)$ $P_6(1)$ $P_7(0)$ $P_7(1)$ $P_8(0)$ $P_8(1)$ $P_9(0)$ $P_9(1)$
$\max(P_i(1)), \forall_i \in \{1 \dots 10\}$
P(0): Probability of not being that category
P(1): Probability of being that category
i: ID of category

Figure 1. Two-Stage Ensemble Classifier

VI. Conclusions

This paper proposes a novel method of constructing an attack detection model on the UNSW-NB15 dataset. Our approach, based on an ensemble of binary expert models, can differentiate several types of network traffic. Obtained results demonstrates that the proposed model generates good results even in unbalanced datasets, improving other state-of-art contributions and demonstrating its efficiency in terms of model generation.

Acknowledgements

Cabe destacar que esta iniciativa se realiza en el marco de los fondos del Plan de Recuperación, Transformación y Resiliencia, financiados por la Unión Europea (Next Generation) en el marco del proyecto con referencia C109/23 ”Proyecto estratégico UEx (Escuela Politécnica de Cáceres)-INCIBE”.

References

[1] O. Mogollon-Gutierrez, J. C. Sancho Nuñez, M. Avila Vegas, and A. Caro Lindo, “A novel ensemble learning system for cyberattack classification,” *Intelligent Automation and Soft Computing*, vol. 37, no. 2, p. 1691–1709, 2023. [Online]. Available: http://dx.doi.org/10.32604/iasc.2023.039255

[2] A. V. Elijah, A. Abdullah, N. JhanJhi, M. Supramaniam, and B. Abdullateef, “Ensemble and deep-learning methods for two-class and multi-attack anomaly intrusion detection: An empirical study,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 9, 2019. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2019.0100969

[3] M. Souhail et. al., “Network based intrusion detection using the unsw-nb15 dataset,” *International Journal of Computing and Digital Systems*, vol. 8, no. 5, p. 477–487, Jan. 2019. [Online]. Available: http://dx.doi.org/10.12785/IJCDS/080505

[4] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the kdd cup 99 data set,” in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*. IEEE, Jul. 2009. [Online]. Available: http://dx.doi.org/10.1109/CISDA.2009.5356528

[5] S. Choudhary and N. Kesswani, “Analysis of kdd-cup’99, nsl-kdd and unsw-nb15 datasets using deep learning in iot,” *Procedia Computer Science*, vol. 167, p. 1561–1573, 2020. [Online]. Available: http://dx.doi.org/10.1016/j.procs.2020.03.367

[6] N. Moustafa and J. Slay, “Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set),” in *2015 Military Communications and Information Systems Conference (MilCIS)*. IEEE, Nov. 2015. [Online]. Available: http://dx.doi.org/10.1109/MilCIS.2015.7348942