Network attacks prediction using set of machine learning models for supporting decision making

Sidahmed Elandaloussi, Pascale Zaraté

Network attacks prediction using set of machine learning models for supporting decision making

Elandaloussi Sidahmed and Zaraté Pascale

IRIT / Toulouse University
2 Rue du Doyen Gabriel Marty, 31042 Toulouse Cedex 9, France
sid.elandaloussi@irit.fr
Pascale.Zarate@irit.fr, web-page: https://www.irit.fr/~Pascale.Zarate

Abstract:

Over the last few years there has been a notable increase in the extent and impact of network attacks. These attacks aim to compromise the confidentiality, integrity, or availability of data and network resources. Furthermore, decision-making becomes crucial in formulating proactive strategies on prevention or detection tasks in order to respond promptly to these network attacks. Besides, there are many approaches to identifying these attacks and making decisions but machine learning techniques are the most popular and reliable for identifying unknown attackers and achieving complete process automation. In this paper a set of Machine Learning methods is used, in particular boosting algorithms to enhance the attack detection process and to create multiple models and then combine them to produce improved results.

Keywords: Decision making, Artificial Intelligence, Machine Learning set, Networks Attacks, Boosting Models, Bagging models, Aggregation method.

Introduction:

Network attacks refer to unauthorized actions or activities aimed at exploiting vulnerabilities in a computer network. These attacks can have various motives, including gaining unauthorized access to sensitive information, disrupting network operations, or causing damage to the targeted system. There are several types of network attacks (**Denial or distributed of service[1,2,3], Malware attacks[4,5], Phishing attacks[6], Man in the middle attacks[7], Packet sniffing attacks[8]**.....).

Several researchers proposed a variety of solutions for identifying and preventing network attacks. The authors of **[9]** examine the cloud's open secure architecture advantages in brief, and focus on DDoS security threats in the cloud model along with the existing methods to defend against them with their pros and cons. According **[10]** Achieving enhanced network resilience against targeted link attacks can be accomplished by leveraging easily accessible local information and employing a robustness measure that is both quick to compute and reliable**. [11]** Collected network flow data from SQL injection attack on the three most widely-used relational database engines and by using this dataset, various supervised learning-based models were trained with the mentioned datasets and the outcomes indicate the feasibility of detecting SQL injections attacks through NetFlow Version 5-based flow data. Another survey in network attacks **[12]** which focuses on an experimental evaluation of the impact of Denial or distributed of service (DDoS) attacks on communications in multiprotocol label switching (MPLS) networks.

The increase in network attacks poses a significant challenge to decision-making processes. Moreover, decision-makers must remain vigilant due to the intricate and dynamic nature of these threats. Staying informed about emerging attack vectors, continuously evaluating the organization's network security stance, and making informed decisions are essential steps in strengthening defensive measures. **[13]** Defines an approach for an IT infrastructure diagnostic in order to analyze, classify and take problems to closure in a short time face to a multi-criteria decision-making problem. **[14]** propose an evolutionary game decision model for network attack defence based on regret minimization algorithm to optimize the learning mechanism through replication dynamic equations, and an optimal defence decision is formulated, aiming to significantly enhance the convergence and learning efficiency of defence decision algorithms.

Over the years, there have been continuous advancements in the field of cyber attacks and security, with the implementation of a diverse array of techniques aimed at safeguarding data. Today, Intelligence artificial (IA) generates pertinent decisions and that one helps to perform analysis and further predictions **[15]**. Then, many researchers have demonstrated the importance of network attack prediction with different IA techniques. **[16]** propose an effective detection technique against DDoS attacks for both SDN data plane and control plane and using a parameter's value threshold to track the existence of the DDoS attack by tracking the average arrival bit rate for switch traffic with an unknown destination address in the data plane. **[17]** They utilize to determine whether the internet traffic is regular for detecting DDoS attacks using machine learning. In order to optimize the network's performance and deliver swift, efficient services to users, enhancements have been implemented **[18]** incorporating artificial intelligence into the realm of network security marks a significant stride forward. By introducing a novel system detection algorithm rooted in the principals of artificial intelligence, this endeavour has proven instrumental in bolstering the capabilities of AI for conducting

rigorous security assessments on network systems. This innovative approach not only enhances the efficiency of security inspections but also sets the stage for a more robust and adaptive defence mechanism against evolving cyber threats. Furthermore, a single machine learning model can make prediction errors based on the accuracy of the training dataset but utilizing a set of technical algorithms can result in the development of a more robust and sophisticated detection technique when compared to alternative methods. **[19]** aimed to evaluate the significance of various classical and set machine learning models in identifying intricate network attacks. This evaluation serves as a valuable guide for selecting robust strategies that can effectively navigate and counteract the challenges posed by advanced network attacks. Several studies **[20,21]** show that set machine learning methods produce satisfactory results, though the evaluation rates may differ compared to other learning techniques. Our study is part of preventive network attacks to support decision makers with using set learning algorithms to achieve more accurate predictions. The remainder of the paper is organized as follows: Section 2 is devoted to discussing a proposed technique and algorithms. The type and methodology used to collect datasets is explained in section 3. In section 4 we show and discuss the obtained results. Finally, the conclusions are presented in Section 6.

Proposed Approach

Set of methods are widely used in various domains, including machine learning competitions, finance, healthcare, and more, due to their ability to improve predictive performance, reduce overfitting, and enhance model robustness.

Boosting is an iterative set method where each model in the ensemble corrects the errors made by the previous ones. Popular algorithms include XGBoost **[22]** and Gradient Boosting Machines **[23]**, which sequentially fit new models to the residuals of the previous models.

Bagging is a powerful set learning technique used primarily in machine learning for improving the stability and accuracy of models.

In our study, we introduce a novel algorithm designed to integrate multiple learning techniques into a unified framework. This algorithm incorporates the assignment of weights to each prediction iteration, enabling it to effectively predict all encountered attacks. By amalgamating various learning methodologies, our approach aims to streamline decision-making processes in handling diverse attack scenarios. Here are the différents steps of our new algorithm:

1. Initially build the first model and execute all ensemble machine learning techniques separately.
2. Calculate error (residuals) for each algorithm.
3. Combines all residuals in the new dataset as a new model (Weak model).
4. Attribute new weight for each algorithm according to previous prediction (the highest weight for the algorithm that gave the best prediction).
5. Execute the same algorithms to this new model.
6. Aggregate results by taking account of the algorithm's weights.
7. Add prediction from this model to the set model and obtain a final prediction.

Data set

Our dataset contains benign and the most up-to-date common attacks, which resembles the true real-world data. It also includes the results of the network traffic analysis with labeled flows based on the timestamp, source, and destination IPs, source and destination ports, protocols and attack (CSV files). Also available is the extracted features definition.

To address a predictive problem using different supervised learning technicals, we typically adhere to three fundamental steps:

Collecting Representative Training Set: This involves gathering a dataset that accurately represents the problem we aim to solve. So, Ensuring the dataset's quality, diversity, and sufficiency is crucial for training robust models.

Selecting and Implementing Boosting Algorithms: Boosting is a machine learning technique that combines multiple weak learners to create a strong learner. Algorithms like AdaBoost, Gradient Boosting, and XGBoost are popular choices due to their effectiveness in handling diverse datasets and improving predictive performance.

Training the Model: During training, the algorithm iteratively learns from the data, adjusting its parameters to minimize prediction errors and improve accuracy. This iterative process continues until the model achieves satisfactory performance or convergence.

Discussion & Result

This section describes the experimental results. The machine learning models and proposed models are implemented in python language. This work utilised XGBoost, Bagging, GradientBoost Machine learning approaches to predict the Network attacks and to validate our proposed aggregating algorithm. The metrics Accuracy, precision, racall and F1 measure were employed to evaluate, revealing distinct patterns for the XGBoost, Bagging, GradientBoost and aggregating algo as depicted in the following Table1:

Models	Accuracy	Precision	Recall	F1
XGBoost	95.17	94.51	92.19	91.05
Bagging	93.21	92.33	91.11	90/07
GradienBoost	95.06	94.43	91.94	90.65
Aggregating Algo	**99.84**	**98.67**	**97.95**	**97.83**

Table1. Comparing network attacks prediction results

The dataset was divided into two subsets: one for training and the other for testing. Specifically, three quarters of the data were allocated for training purposes, while one quarter was reserved for testing the model's performance. The accuracy measures the proportion of all samples that are correctly classified by the model. It gives an overall assessment of the model's correctness. Recall is the ratio of correctly classified positive samples (attacks) to the total number of actual positive samples. It indicates the model's ability to capture all positive instances. Precision quantifies the proportion of correctly classified positive samples (attacks) among all samples that the model classified as positive. It reflects the model's accuracy in identifying true positive cases without misclassifying negatives. F1-Measure, also known as the F1-Score, represents the harmonic mean of precision and recall. It combines both precision and recall into a single metric, providing a balanced assessment of a classifier's performance.

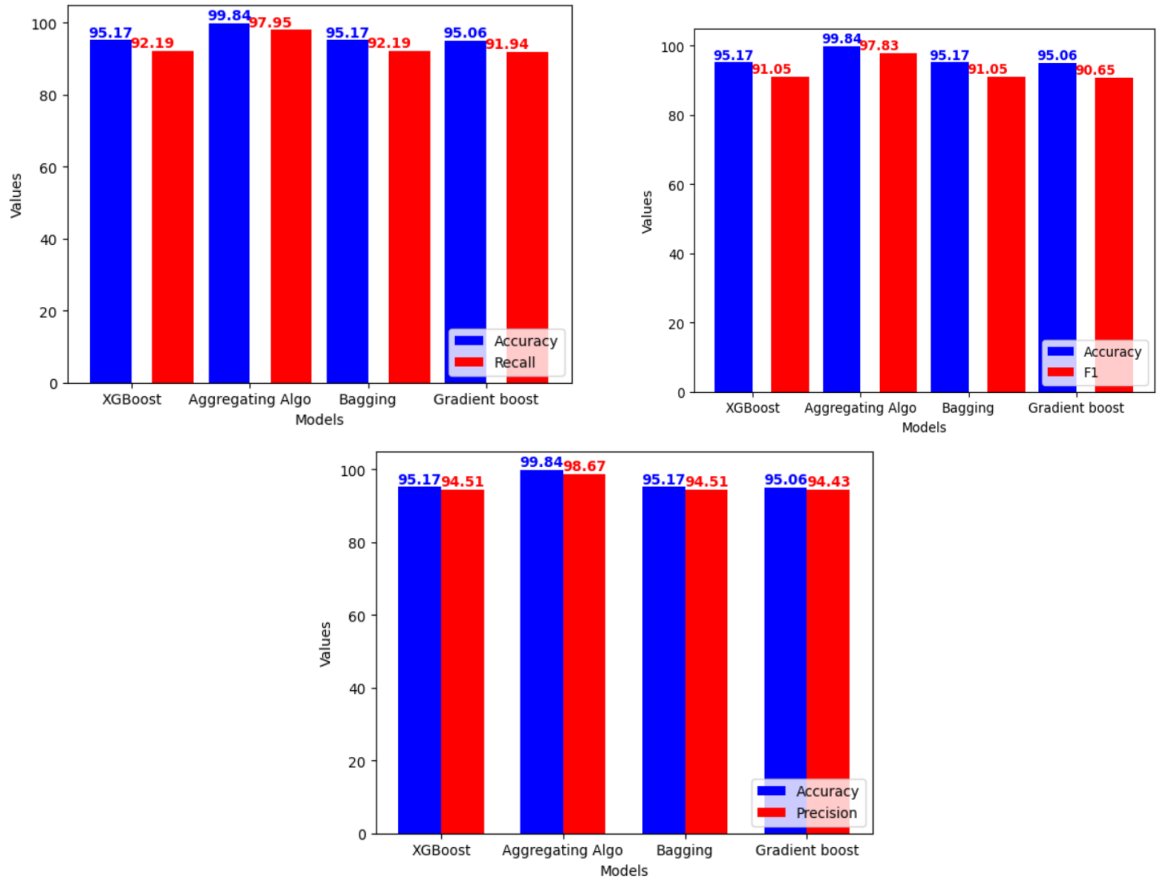

Figure1: Models Metrics results

The experimental results show that the proposed method efficiency by obtaining high precision, attacks detection rate compared to other related techniques as depicted in Figure1. Our approach gives a precision of 98,67% and Accuracy of 99,84%, while the best prediction of our implemented set machine learning model gives a precision of 94,51% and an accuracy of 95,17 with a precision deviation of 4,16%. As result, we observe that when combined, these diverse algorithms can provide more robust predictions than any single algorithm alone

Conclusion

In this study, we were pointed out that set machine learning solutions offer a highly promising approach by effectively identifying potential unknown attackers and facilitating complete process automation. In addition, we developed a new model that combines a sequential and parallel set machine learning method to benefit from both advantages, which has enabled us to achieve very promising results.

In this survey, the metrics used to evaluate the proposed system are Accuracy, Precision, Recall, F1-Measure and they give very promising values compared with another ML technique like XGBoost, Bagging, GradientBoost and can successfully accomplish classification, and prediction tasks in order to support technician for making a final decision about the attacks.

As future work, we plan to integrate unsupervised learning algorithms into our approach and provide further explanation on how our method calculates the weights assigned to these models. Additionally, we aim to elucidate how these weights can be incorporated within an unsupervised learning framework to enhance the effectiveness of our approach.

References:

1. Anderson Bergamini de Neira a,∗ , Burak Kantarci b , Michele Nogueira,2023 “Distributed denial of service attack prediction: Challenges, open issues and opportunities”
2. Raj Kumar Batchu, Hari Seetha, 2022. “An integrated approach explaining the detection of distributed denial of service attacks”
3. Mohammad Kamrul Hasana,∗ , A.K.M. Ahasan Habiba,∗ , Shayla Islamb,∗ , Nurhizam Safiea , Siti Norul Huda Sheikh Abdullaha , Bishwajeet Pandeyc, 2023, The 3rd International Conference on Power and Electrical Engineering (ICPEE 2022) 29–31 December, Singapore, “DDoS: Distributed denial of service attack in communication standard vulnerabilities in smart grid applications and cyber security with recent developments”
4. Sheng Xu, Haicheng Tu, Yongxiang Xia, 2022. ”Resilience enhancement of renewable cyber–physical power system against malware attacks”
5. Hemant Rathorea,∗ , Animesh Sasana , Sanjay K. Sahaya , Mohit Sewak b, Pattern Recognition Letters Volume 164, December 2022, Pages 119-125 “Defending malware detection models against evasion based adversarial attacks”
6. Mohan Krishnamurthy, Eric S. Seagren, Raven Alder, Aaron W. Bayles, Josh Burke, Skip Carter, Eli Faskha, How to Cheat at Securing Linux 2008, Pages 203-247, “Chapter 7 - Network Analysis, Troubleshooting, and Packet Sniffing”
7. Mohamed Faisal Elrawy a,b,∗,1 , Lenos Hadjidemetriou b,2 , Christos Laoudias b,2 , Maria K. Michael, Sustainable Energy, Grids and Networks 36 (2023) 101167 ,”Detecting and classifying man-in-the-middle attacks in the private area network of smart grids”
8. Neminath Hubballi∗ , Nikhil Tripathi, Journal of Information Security and Applications 35 (2017) 32–43: “An event based technique for detecting spoofed IP packets”
9. Andrew Carlina , Mohammad Hammoudehb , Omar Aldabbas, The International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015), “Defence for Distributed Denial of Service Attacks in Cloud Computing”
10. Marco Tomassini, Physica A 615 (2023) 128563 ,“Designing robust scale-free networks under targeted link attack using local information”
11. Ignacio Samuel Crespo-Martínez, Adrián Campazas-Vega, Ángel Manuel Guerrero-Higueras b , Virginia Riego-DelCastillo b , Claudia Álvarez-Aparicio b , Camino Fernández-Llamas,

Computers & Security 127 (2023) 103093 “SQL injection attack detection in network flow data”
12. Béla Genge, Christos Siaterlis, international journal of critical infrastructure protection 6 (2013) 87–95 ,”Analysis of the effects of distributed denial-of-service attacks on MPLS networks”
13. S. elandaloussi, N. Taghezout International Journal of Decision Support System IJDSST A Text Mining Approach agent based DSS for IT Infrastructure Maintennace
14. Hui Jin 1 , Senlei Zhang 1 , Bin Zhang, Shuqin Dong, Xiaohu Liu, Hengwei Zhang ⇑ , Jinglei Tan ⇑, Journal of King Saud University – Computer and Information Sciences 35 (2023) 292–302, “Evolutionary game decision-making method for network attack and defense based on regret minimization algorithm”
15. P. Krishna Kishore * , S. Ramamoorthy, V.N. Rajavarman, International Journal of Intelligent Networks 4 (2023) 38–45, “ARTP: Anomaly based real time prevention of Distributed Denial of Service attacks on the web using machine learning approach”
16. Waheed G. Gadallah a,b,∗, Hosny M. Ibrahim a, Nagwa M. Omar , Computers & Security 137 (2024) 103588, “A deep learning technique to detect distributed denial of service attacks in software-defined networks”
17. Yonghong Wang a,b , Xiaofeng Wang a,b,* , Mazeyanti Mohd Ariffin c , Masoumeh Abolfathi d , Abdulmajeed Alqhatani e , Laila Almutairi, Computers and Electrical Engineering 108 (2023) 108655, “Attack detection analysis in software-defined networks using various machine learning method”
18. T. Hua, L. Li, T. Guarda, I. Lopes, and Á. Rocha, “Computer network security technology based on artificial intelligence,” *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 5, pp. 6021–6028, 2019.
19. Dhanya K. A.a , Sulakshan Vajipayajulab, Kartik Srinivasanc , Anjali Tibrewalc , T. Senthil Kumard, T. Gireesh Kumard, on of Network Attacks using Machine Learning ,Procedia Computer Science 218 (2023) 57–66, “Detection of Network Attacks using Machine Learning and Deep Learning Models”
20. Muhammad Nasir Amin a,* , Bawar Iftikhar b , Kaffayatullah Khan a , Muhammad Faisal Javed b , Abdullah Mohammad AbuArab a , Muhammad Faisal Rehman , Structures 50 (2023) 745–757 , “Prediction model for rice husk ash concrete using AI approach: Boosting and bagging algorithms”
21. Daniel Asante Otchere [a b], Tarek Omar Arbi Ganat [a b], Jude Oghenerurie Ojero [c], Bennet Nii Tackie-Otoo [a b], Mohamed Yassir Taki , “Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions”, Journal of Petroleum Science and Engineering, Volume 208, Part E, January 2022, 109244
22. Zhen Sun [a], Yalin Li [a 1], Yuxi Yang [b], Li Su [c], Shijie Xie [a, ”Splitting tensile strength of basalt fiber reinforced coral aggregate concrete: Optimized XGBoost models and experimental validation”, Construction and Building Materials, volume 416, 16 February 2024, 135133
23. Clément Dombry, Jean-Jil Duchamps, « Infinitesimal gradient boosting », Stochastic Processes and their Applications,Volume 170, April 2024, 104310