DEEP LEARNING APPROACHES FOR ANOMALY DETECTION IN NETWORK TRAFFIC

SRIPADA NSVSC RAMESH
Research Scholar
Dept of Computer Science & Engg
Arni University- H.P

DR. PRASADU PEDDI
Research supervisor
Arni University- H.P

DR. DOLS SANJAY S
Co-Supervioser
Aditya Engineering College,
Surampalem, A P

ABSTRACT

A great deal of attention has been given to deep learning in the field of network and information security. Any intrusion and anomaly in the network can significantly impact many areas, such as security of the private and social data, national security, social and financial concerns, etc. Therefore, network and information security are a broad research domain for which researchers are actively utilizing the functionally improved, emerging deep learning technique and report the improved result. In this review paper, we have analysed several deep learning methods in the area of network anomaly, intrusion detection, network traffic analysis and its classification. We have presented a comprehensive review of widely known deep learning approaches. And then, we conclude with open research challenges and unresolved issue for further study.

Network intrusion detection is a key pillar towards the sustainability and normal operation of information systems. Complex threat patterns and malicious actors are able to cause severe damages to cyber-systems. In this work, we propose novel Deep Learning formulations for detecting threats and alerts on network logs that were acquired by pf Sense, an open-source software that acts as firewall on FreeBSD operating system. Pf Sense integrates several powerful security services such as firewall, URL filtering, and virtual private networking among others. The main goal of this study is to analyse the logs that were acquired by a local installation of pf Sense software, in order to provide a powerful and efficient solution that controls traffic flow based on patterns that are automatically learnt via the proposed, challenging DL architectures.

Keywords: *semi-supervised anomaly detection; deep feature learning; convolutional neural networks; suricata network logs anomaly detection*

INTRODUCTION

Network log management corresponds to the collection, manipulation, analysis and reporting of large volume and velocity data, such as event-logs, audit records, audit trails, etc. Additionally, log management evaluates the firewall capabilities in enhancing net-work security, facilitates the early detection of possible vulnerabilities, and minimizes the resolution time of any suspicious action that corrupts the normal operation of information systems. On this direction, network traffic analysis enables the immediate and deep understanding of several significant network metrics, including the network type, origin, size, destination of variant packets, and the uploading/ downloading speed, among others. Consequently, network traffic analysis provides significant assistance towards the identification of malicious packets and actors within the traffic flow. However, with the increased penetration rates that are reported on modern information systems, sophisticated cybersecurity protection technologies like Machine Learning (ML) and Artificial Intelligence (AI) can be incorporated into the loop. According to our knowledge, this work is out of the first that propose variant

deep feature learning architectures towards the problem of anomaly detection on network logs that were acquired by the pf Sense fire wall, with main goal the Event type classification. The proposed scheme can be efficiently extended-with minor modifications regarding the data pre-processing in detecting multiple threat types from variant information systems. The rest of the paper is structured as follows: Section 2describes the state-of-the-art concerning network intrusion detection systems, along with the most recent Machine Learning/Deep Learning approaches that were reported in the literature; Section 3demonstrates the main formulation of this work, including the proposed architectures: the Convolutional Neural Networks and the Long Short Term Neural Networks towards the multi-class Suricata pf Sense logs anomaly detection; Section 4presents the complete experimental setup, including the data acquisition, the dataset creation, the achieved performance of the proposed schemes and their comparison; finally, Section 5concludes this work and highlights our main future directions.

LITERATURE REVIEW

Shuzhan Wang [2024] Computer network anomaly detection and log analysis, as an important topic in the field of network security, has been a key task to ensure network security and system reliability. First, existing network anomaly detection and log analysis methods are often challenged by high-dimensional data and complex network topologies, resulting in unstable performance and high false-positive rates. In addition, traditional methods are usually difficult to handle time-series data, which is crucial for anomaly detection and log analysis. Therefore, we need a more efficient and accurate method to cope with these problems. To compensate for the shortcomings of current methods, we propose an innovative fusion model that integrates Isolation Forest, GAN (Generative Adversarial Network), and Transformer with each other, and each of them plays a unique role. Isolation Forest is used to quickly identify anomalous data points, and GAN is used to generate synthetic data with the real data distribution characteristics to augment the training dataset, while the Transformer is used for modeling and context extraction on time series data. The synergy of these three components makes our model more accurate and robust in anomaly detection and log analysis tasks. We validate the effectiveness of this fusion model in an extensive experimental evaluation. Experimental results show that our model significantly improves the accuracy of anomaly detection while reducing the false alarm rate, which helps to detect potential network problems in advance.

Yung-Chung Wang [2023] The prevalence of internet usage leads to diverse internet traffic, which may contain information about various types of internet attacks. In recent years, many researchers have applied deep learning technology to intrusion detection systems and obtained fairly strong recognition results. However, most experiments have used old datasets, so they could not reflect the latest attack information. In this paper, a current state of the CSE-CIC-IDS2018 dataset and standard evaluation metrics has been employed to evaluate the proposed mechanism. After pre-processing the dataset, six models—deep neural network (DNN), convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), CNN +

RNN and CNN + LSTM—were constructed to judge whether network traffic comprised a malicious attack. In addition, multi-classification experiments were conducted to sort traffic into benign traffic and six categories of malicious attacks: BruteForce, Denial-of-service (DoS), Web Attacks, Infiltration, Botnet, and Distributed denial-of-service (DDoS). Each model showed a high accuracy in various experiments, and their multi-class classification accuracy were above 98%. Compared with the intrusion detection system (IDS) of other papers, the proposed model effectively improves the detection performance.

Prathamesh Kulkarn [2022] A sudden spike or dip in a metric is an anomalous behaviour and both the cases needs attention. Detection of anomaly can be solved by supervised learning algorithms if we have information on anomalous behaviour before modelling, but initially without feedback it's difficult to identify that points. Anomaly detection is important and finds its application in various domains like detection of fraudulent bank transactions, network intrusion detection, sudden rise/drop in sales, change in customer behaviour, etc. So we model this as an unsupervised problem using algorithms like Isolation Forest, One class SVM and LSTM. Here we are identifying anomalies using isolation forest.

Konstantina Fotiadou [2021] Network intrusion detection is a key pillar towards the sustainability and normal operation of information systems. Complex threat patterns and malicious actors are able to cause severe damages to cyber-systems. In this work, we propose novel Deep Learning formulations for detecting threats and alerts on network logs that were acquired by pf Sense, an open-source software that acts as firewall on FreeBSD operating system. pfSense integrates several powerful security services such as firewall, URL filtering, and virtual private networking among others. The main goal of this study is to analyse the logs that were acquired by a local installation of pf Sense software, in order to provide a powerful and efficient solution that controls traffic flow based on patterns that are automatically learnt via the proposed, challenging DL architectures. For this purpose, we exploit the Convolutional Neural Networks (CNNs), and the Long Short Term Memory Networks (LSTMs) in order to construct robust multi-class classifiers, able to assign each new network log instance that reaches our system into its corresponding category. The performance of our scheme is evaluated by conducting several quantitative experiments, and by comparing to state-of-the-art formulations.

Anomaly Detection

Anomaly detection in network traffic classification in IoT refers to the process of identifying abnormal or suspicious patterns in network traffic behavior. It involves analyzing network traffic data to detect deviations from normal behavior that may indicate potential security breaches, intrusions, or abnormal activities. This type of detection is important because it helps identify and mitigate unknown or emerging threats in IoT networks. By accurately detecting anomalies in network traffic, administrators can take prompt action to investigate and respond to potential security incidents, ensuring the overall security and integrity of IoT devices, data, and infrastructure.

Anomaly Detection Based on Log Analysis

Log analysis is an important research area in the field of network anomaly detection

and log analysis. Log files record the activities and events of systems such as network devices, servers, applications, etc., and thus provide important information about the state of the system. Researchers use the information in log files to detect anomalous behavior. A common approach is to use rule-based log analysis. This approach relies on predefined sets of rules for detecting events that do not match those rules. For example, if unauthorized access attempts or unusual system error messages appear in the logs, a system administrator can trigger an alert using a rule. Another approach is to use natural language processing (NLP) techniques to analyze log text. Researchers

Network Intrusion Detection Systems

Nowadays the complexity of cyber-attacks and malicious events has grown tremendously. For this purpose, the design and exploitation of robust Intrusion detection systems becomes a fundamental entity towards the protection and sustainability of the Information and Communication Tecnology infrastructures. IDSs provide the means to early detect malicious cyber-activities and design efficient mitigation actions against them. Recently, numerous interesting intrusion detection approaches have been proposed in the related literature. Generally, network intrusion detection methodologies are divided into two main categories (i) the signature-based, and (ii) the anomaly detection-based approaches. Signature-based intrusion detection approaches focus on detecting pre-defined malicious threats, for instance network traffic for a series of malicious packet sequences, or bytes. The key benefits of the specific approaches lie into the fact that signatures can be efficiently extracted, under the constraint that precise prior knowledge regarding their patterns or structure exists.

Anomaly Detection in network

An anomaly detection is a process that finds unusual patterns in the traffic in the network. It aims to identify potential security breaches or other issues that could affect the network. An anomaly detection solution provides various advantages to network administrators. It can help them identify potential security threats such as malware and hacking attempts. It can also help them identify system failures that could cause downtime. In addition, it can help them plan their resources and improve the performance of their networks. Unfortunately, there are some limitations to network anomaly detection. One of the most challenging factors is distinguishing between abnormal and normal traffic. Since the behavior of the network can vary depending on various factors, such as the time of day and the user's behavior, it is not easy to define a standard for normalization. Unfortunately, one of the biggest limitations of an anomaly detection solution is the high number of false positives. This can be caused by the mistake of identifying legitimate traffic as anomalous. It can be very time-consuming and costly to investigate. Also, it can't effectively detect attacks that are designed to evade detection.

METHODOLOGY

The design concept of this model aims to enable computer network anomaly detection and log analysis to achieve superior performance and accuracy in increasingly complex threat environments by integrating diverse technologies. Its core idea lies in organically combining these components to more effectively capture and respond to a variety of network security challenges, ensuring the reliability and security of computer networks. This

innovative deep learning approach represents an important contribution to the field of cybersecurity, providing a powerful tool and methodology to address future cyber threats. First, the raw computer network dataset is fed into the Isolation Forest component. The task of the Isolation Forest is to quickly separate normal network traffic data from potentially anomalous data points by constructing a randomized binary tree. The purpose of this step is to generate an anomaly score for each data point, where higher scores indicate potentially anomalous data points. Also, the works in conjunction with the Isolation Forest component, which consists of two parts, the generator and the discriminator. The task of the generator is to generate synthetic data similar to normal network traffic data to approximate the distribution of real data. The discriminator, on the other hand, evaluates the similarity between the generated data and the real data and pushes the generator to generate more realistic data. This process is iterative, and the generator continuously improves the quality of the generated data, enhancing the fidelity of the data while helping to capture potentially anomalous data features.

RESULT & DISCUSSION

Without feature selection

Table 1 Evaluation without feature selection

Algorithm	Accuracy	Precision	Recall	F1-Score
SVM	92	93	89	91
Random Forest	94	96	91	93
ANN	95	94	93	93

With feature selection

Table 2 Evaluation with feature selection

Algorithm	Accuracy	Precision	Recall	F1-Score
SVM	96	97	95	96
Random Forest	97	98	96	97
ANN	98	97	97	97

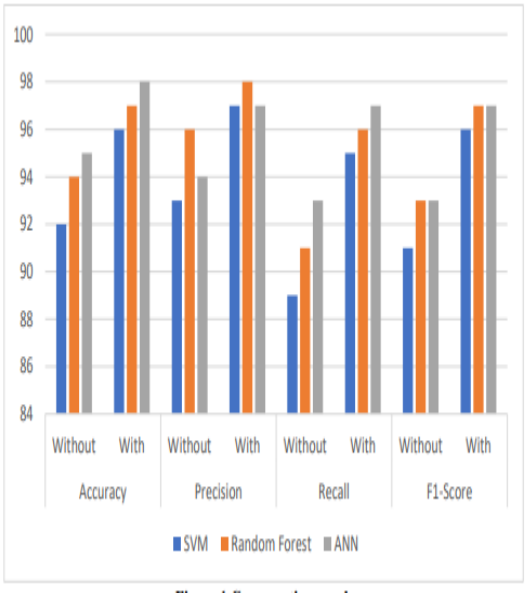

Figure 1 Comparative graph

Although the three algorithms did well when it came to detecting network anomalies without feature selection, they performed significantly better when it came to capturing the most relevant data with the selected features. The improvements in the accuracy, recall, F1-score, and precision of the three algorithms were significant. The results indicate that selecting the right features can significantly improve deep

learning and machine learning algorithms' performance when detecting anomalies in networks.

CONCLUSION

Advancements in network traffic classification for IoT using deep learning techniques have shown promising results in various areas, including application awareness, accuracy improvement, malicious traffic detection, anomaly detection, botnet detection, intrusion detection, and zero-day attack detection. These advancements have brought several strengths to the field. DF model, and others demonstrate superior performance, versatility, and fast detection speed. They exhibit strengths like effective feature extraction, robustness in handling encrypted traffic, enhanced trustworthiness, and improved representation learning. Recurrent neural networks, reinforcement learning, transfer learning, federated learning, parallel processing, real-time anomaly detection, high-performance computing, and supervised learning for labeling anomalies are some of the directions that should be pursued. Additionally, refinement of models, exploration of new architectures and techniques, expansion of datasets, addressing privacy and explain ability concerns, differentiation of attack subcategories, reliability testing, generalization to diverse environments, optimization of classification for minority classes, and evaluation on real IoT network data are key areas for future research. By addressing these limitations and pursuing these avenues of research, advancements in network traffic classification in IoT using deep learning techniques can lead to improved security, privacy, trustworthiness, and defense against various threats and vulnerabilities.

Reference

1. *Shuzhan Wang [2024], "Deep Learning-based Anomaly Detection and Log Analysis for Computer Networks", Journal of Information and Computing, ISSN 3006-0931,vol. 2,issue.(2),pages.34-63*
2. *Konstantina Fotiadou [2021], "Network Traffic Anomaly Detection via Deep Learning", Information, ISSN:2078-2489,vol.12,issue.(5),https://doi.org/ 10.3390/info12050215*
3. *Yung-Chung Wang [2023], "Network Anomaly Intrusion Detection Based on Deep Learning Approach", Sensors, ISSN 1424-8220,Vol.23,issue.(4), https:// doi.org/10.3390/s23042171*
4. *Prathamesh Kulkarn [2022], "Anomaly detection in network traffic using unsupervised machine learning approach", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, ISSN:2321-2004,Vol.10,Issue.4*
5. *Yongtao Guan [2022], "Optimal Frameworks for Detecting Anomalies in Sensor-Intensive Heterogeneous Networks," INFORMS Journal on Computing, INFORMS,ISSN: 1526-5528, vol.34,issue.(5),pages.2583-2610,*
6. *Yasmeen S. Almutairi [2022], "Network Intrusion Detection Using Machine Learning Techniques", Advances in Science and Technology Research Journal, ISSN:2299–8624,Vol.16,issue.(3),pages.193–206*
7. *Vrushali V. Kondhalkar [2022], "Network Intrusion Detection System using Machine Learning", International Research Journal of Engineering and Technology, ISSN: 2395-0056,Volume.09,Issue.03*
8. *Vipin Das [2010], "Network Intrusion Detection System Based On Machine Learning Algorithms", International Journal of Computer Science & Information Technology (IJCSIT),ISSN 0975-9646,Vol 2,issue.6*
9. *P. Sudharsanarao [2019], "Detecting And Preventing Of Dos Attacks By Dynamic Path Identifier Networks", International Journal of Computer Science Trends and Technology,ISSN: 2347-8578 ,Volume.7,Issue.3*

10. *Olivia Jullian [2023], "Deep-Learning Based Detection for Cyber-Attacks in IoT Networks: A Distributed Attack Detection Framework", Journal of Network and Systems Management,ISSN : 1573-7705,vol.31*