IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Cloud Network Traffic Classification and Intrusion Detection System using Deep Learning

Mala K[1]

[1]**Assistant Professor, Department of Information Science and Engineering, Channabasaweshwar Institute of Technology Gubbi Tumakur, Karnataka**

Annapurna H S[2]

[2]**Professor and Head, Department of Information Science and Engineering, Sri Siddhartha Academy of Higher Education, Tumkur, Karnataka**

Abstract: A potent technique for identifying Internet of Things (IoT) assaults and identifying novel forms of intrusion to get access to a more secure network is the application of deep learning in a variety of models. Where in there is a high False Positive Rate(FPR) in Network Intrusion Detection System(NIDS),with decent prediction rate. To reduce the FPR and give scalable solution, , we recommend using a Deep Learning (DL) model. By placing the model on the cloud, we can increase the NIDS's responsiveness during periods of high load, hence boosting availability. Because the model is installed as a micro service and is operating on Docker containers on the cloud instance, which can be accessed by REST APIs. According to the testing results, Long Short-Term Memory (LSTM) with two layers and Deep Neural Networks (DNN) with five hidden layers performed with a minimum accuracy of 88.75% and a maximum accuracy of 95.02%. The Random Forest technique in standard machine learning algorithms has achieved 86% accuracy. Network intrusions are becoming fraudulent and sophisticated. For any firm, having an appropriate network intrusion detection system nearby is the most important component.

Keywords: FPR, DNN, DL, Dockers, LSTM

1. Introduction

The demand for efficiency and resilience when using different approaches to learn data has grown owing to the explosive growth of traffic data. As of present scenario contemporary times, network traffic anomaly detection has emerged as a critical element of our cyberspace defense [1]. For instance, the stability of the full network environment is greatly impacted by hardware upgrades and communication protocol updates [2,3]. However, because network attack scenarios and their associated attack methods have grown significantly more complicated, the methods currently employed for intrusion detection will eventually become antiquated. It is imperative that traffic anomaly detection technology always be one step ahead of the attackers, primarily however, attackers are already well-versed in and have access to the most advanced modern techniques. Launching new research is crucial [4–7]. Recent years have seen A spike in curiosity in deep learning model research, giving special consideration to the speech, image, and, and natural language [8–11]. Through the network, cloud computing may provide consumers with a variety of assets in the form of services. The fundamental idea behind cloud computing is that "everything can be a kind of service and will be provided to users in the form of lease" [12–14]. But the rapid advancement and widespread use of cloud computing has brought forth some unavoidable new issues. The primary issue is cloud computing security, which the sector is growing increasingly worried about [15–17]. Self-service on demand, Internet connectivity, rapid elastic architecture, virtualized resource pools, measurability, and multiuser capabilities are just a few of the numerous features of cloud computing. While these features offer users a speedier and more convenient computing method, they also present new security risks to cloud computing systems. A deep learning-based anomalous traffic detection approach in a cloud computing setting is suggested tackle with the issue of the inadequate detection performance of the current intrusion detection methods in the context of high-dimensional huge data with uneven class distribution. The following are the contributions: The general regression neural network (GRNN) is coupled with the fuzzy C-means (FCM) algorithm. To increase the anomalous traffic detection system's stability, the FCM algorithm clusters the samples that would be classified in the original space. The GRNN model is then trained using the sample that is closest to the FCM clustering center. and update the center until a stable clustering center is achieved,the FCM-GRNN technique's parameters are optimized through the usage of the global search feature of the modified fruit fly optimization algorithm (MFOA).Additionally, the fruit flies' excellent smell and visual senses are combined with an iterative search utilizing the three-dimensional search approach to find ideal spread value.

2. Literature Survey

Towards overcoming these restrictions caused by high traffic, researchers used cloud computing technologies to accelerate computation. In the recent few years, sophisticated intrusion detection algorithm has incorporated to compute, especially with the Hadoop Distributed File System (HDFS). Hadoop uses the parallel processing concept known as Map Reduce which is the open-source software framework for distributing the processing and storing of large amounts of data. It can quickly analyze huge amounts of network traffic to predict malignant activity or intrusion detection. HDFS is commonly used for handling data related to traffic within the network by offering scalable and dependable data storage. To leverage cloud computing for intrusion detection analysis, new parallel processing paradigm-aligned and there is a necessity to develop model. As an alternative, It is necessary to integrate the computational model into the cloud computing environment. All algorithms that require access to the input data to function with iterative or linear computations are hampered by platforms like Hadoop and Map Reduce, which typically distribute processing among dozens or more computers. As such, the architecture of cloud computing does not support the direct use of many widely used computational algorithms. Researchers are working on developing cloud-based, solutions for scalable performing machine learning. In a way to get around this restriction, Furthermore there is a importance to detect the cloud based intrusion detection by using MLAs[1]. Identifying Uninvited Parties An overview of intrusion detection history was given by Kemmerer and Vigna. From late 1970s till 1900s the system administrators would manually discover and analyze the data, wherein 2000s the real time solutions were available. The need for classification and frameworks has increased in tandem with the variety and complexity of intrusions. IDSes were defined by Debar et al. And in any systems that handle data coming in from the network that needs to be secured. To incorporate taxonomy-based method to understanding them. Intrusion detection is the technique of keeping an eye on activity in computers and networks .They went on to discuss existing intrusion detection algorithms by utilizing two key ideas: Signature-Based Intrusion Detection and Anomaly based intrusion detection. The section "Using MLAs in intrusion detection" will cover these ideas. Techniques for detecting Approaches to intrusion detection fall into two primary categories: Anomaly-based and signature-based. Signature-based methods, often known to be knowledgebased or intrusion-based methods, make use of databases containing historical attack and system vulnerabilities that are known to exist. While signature-based methods can effectively prevent known assaults, they are susceptible to unknown ones. Attack signatures must be updated frequently so it may be helpful to lessen this restriction. But this could involve a lot of overhead and resources. Techniques based on abnormality or behavior used to defend against unknown attackers. By contrasting intrusion attempts with regular network activity, they identify deviations (i.e. commands or traffic). Although there is a huge of overlap between these two methods and they are frequently considered to be similar in the literature, we propose they are not the same. From here on, behavior-based approaches that don't necessarily compare well to anomaly- and primitive-based methods will be defined as that involve initially training a system to generate a basic profile, then using that profile to search for abnormalities. For instance, an administrator may set up rules in behavior-base detection systems that, if broken, would result in warnings. In actuality, however the two kinds of approaches are frequently interchangeable, it's crucial to recognize their little variations. "Around" detection behavior after efficient detection algorithm are developed, the next concern is what sort of behavior the system takes on following detection. IDSes are rarely only reactive systems that respond only after the reality is checked. Wherein Debar et al., for instance, discussed measures upon discovery, Behavior "around" detection is preferred as it provides a more accurate representation of a system's possible responses prior to and following detection.. Thus, the earliest taxonomies of intrusion strategies was developed by Halme and Bauer, who separated the techniques into six categories: detection, preemption, deterrent, deflection, prevention, and countermeasures. The last three strategies—deflection, detection, and countermeasures—are active ways to safeguard system components, while the first three—prevention, preemption, and deterrence—are passive ones meant to fend off attacks. Several of these strategies are adaptable and may be applied before, during, or after attacks at various stages of the procedure. In spite of the nomenclature or deployment sequence, intrusion detection systems (IDSs) are essential technologies that can identify and respond to many forms of attacks. Researchers discovered early real-time IDSes in the late 1990s, and many of them used Halme and Bauer's six approaches, which combined signature and anamoly-based methodologies. But as time went on and traffic kept growing unabated, these methods became unworkable in real-time, so To improve IDS analysis and computation, researchers are turning to cloud computing. Detecting intrusions using MLAs Several researchers have argued that MLAs are important for intrusion detection. Even though during the course of the preceding decades, the approaches and techniques previously outlined significantly improved intrusion detection techniques. MLAs could initially appear as simple choice for enhancing these systems. After example, in other seemingly related fields of computer science, such spam-detection, MLAs serve as the foundation for anomaly identification. MLAs, however, cannot be directly used in intrusion detection. In their overview of the difficulties MLAs face in intrusion detection, Sommer and Paxson suggested out that machine learning algorithms are best at identifying behavior that resembles previously observed patterns. It defies the broad description of anomaly-based intrusion detection methods, which aim to find new types of attacks. Furthermore, the greater cost of errors reduces MLAs' ability to recognize intrusions effectively, limited availability of training data and significant fluctuations in the input data. In spite of these significant obstacles, after resolving problems, researchers developed MLAs for malicious activity detection. Below, we'll discuss recent advancements in the realm of MLAs in intrusion detection. Emphasizing the work that is effectively resolving the issues Sommer and Paxson first rose in 2010. The two MLA types that are typically used in this research are clustering and classification algorithms, both methods are appropriate for intrusion detection methods. Clustering techniques unsupervised machine learning that doesn't rely on training data is called clustering. or models of classification. Rather, it finds comparable patterns in input datasets by dividing them into clusters based on shared attributes. To find these patterns, similarity measurements (such the Euclidean distance) are frequently used. A widely used, straightforward clustering algorithm is the k-means algorithm. The algorithm computes the average distance between the initial cluster centers, or centroids, and every other point in the system after first choosing the first cluster centers. After then, this process can be repeated endlessly, creating new centroids and moving data points until the average distance is reduced and there is not necessary to relocation. Numerous fields can make use of clustering methods, including encompassing, among other fields, computer vision, search, geostatistics, market segmentation, and medicine. They can be helpful in supplying initial setups for other algorithms, And they are also commonly used as preparatory measures. However, clustering methods are used to address the significant issue of network traffic classification, which is more pertinent to our aims. Traffic classification was previously achieved solely through payload- and port-based methods. But as applications now frequently employ encryption, masquerading, and dynamic port numbers, clustering methods have been employed to take advantage of the unique features of applications in behavior-based approaches.

3. Methodology:

The proposed DL model switches a regularized multi-layer perception for a fully connected (FNN) in CNN. Unlike FNN, CNN uses convolution as a mathematical operation instead of multiplication or dot product. There are custom hyper parameters for the convolution operation, including the size of the filter, the number which is in filters, and the stages involved in in the algorithm for creating the output matrix. In order to control the decreasing tensor dimensions that are used as input goes through multiple convolution layers, which are later given to input padding. To decrease the size of the sample features across the layers, in between subsequent convolution layers there can be pooling layer. A completely linked layer comes before the categorization output layer that has regularization applied. The UNSW-NB15 dataset, as indicated, will be used to test our model. It was chosen primarily because it represents actual world traffic which is in network that is associated with typical weaknesses and exposures. Many models have been run on the dataset, but the results have been less than ideal, leaving room for improvement. Despite the fact that the original dataset had more than two million case simulations, Using nine attack groups, the architects categorized imputed datasets for training and testing.

Table I lists the various attack kinds along with a brief description of each assault.

Table.1: Attack categories and descriptions

Attack	Short Description
Normal	Benign network traffic
Fuzzers	Malignant traffic related to spams which are semi random program injected automatically, penetrations or port scans
Analysis	Attack related to intercepting and or inspect network traffic through penetrations or scans
Backdoors	Attack pertaining to use of mechanism designed to attack remotely bypassing the existing security
DoS	Attack aimed at flooding network resources making it less available and inaccessible
Exploits	Exploitations through security loop holes in operating system or any other software applications
Generic	Attacks which are explained by block-cipher, brute force or cryptanalysis
Reconnaissance	Vulnerabilities are observed in the target system
Shellcode	to navigate through the system and gain control it is a short code “payload”
Worms	Malicious code that multiplies itself in etwork

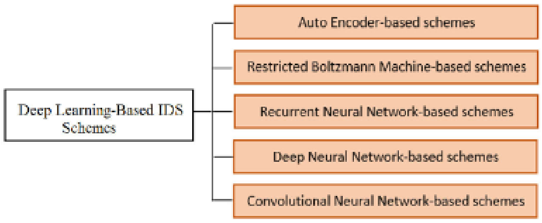
Deep Learning-Based IDS Schemes
Auto Encoder-based schemes
Restricted Boltzmann Machine-based schemes
Recurrent Neural Network-based schemes
Deep Neural Network-based schemes
Convolutional Neural Network-based schemes

Figure:1 Classification of the deep learning based intrusion detection schemes.

Conclusion and Future Work

Using the most recent dataset of simulated network traffic and adding pertinent elements as well as prevalent cyber security vulnerabilities and exposures, this article addressed network intrusion detection systems. When compared to the outcomes of comparable deep learning-based network IDSs, the suggested deep learning classification architecture in conjunction with the semi-dynamic hyper parameter tuning strategy showed appreciable improvements to multiclass models. The models demonstrated that, for both the prepartitioned and user-defined multiclass classification, our suggested method achieved an overall accuracy of 95.4% and 95.6%. While the suggested approaches have yielded encouraging outcomes, We admit that there is space for improvement, namely with regard to feature reduction techniques. To improve model classification with the UNSW-NB15 dataset and increase our models' resistance to zero-day attacks, future work should involve transfer learning with pertinent existing datasets. We will look into bootstrapping strategies to create a balanced dataset for training a multiclass classification model in addition to transfer learning. In order to provide adaptive and resilient network intrusion detection systems that accurately identify common vulnerabilities and exposures as well as zero-day network behavioral aspects that lower the chance of compromise, deep learning anomaly detection models will be added to cyber security infrastructures.

References

[1] Z. H. A. N. G. Yong-dong, C. H. E. N. Si-yang, P. E. N. G. Yu-he, and Y. A. N. G. Jian, “A survey of deep learning based network intrusion detection,” Journal of Guangzhou University Natural Science Edition, vol. 18, no. 3, pp. 17– 26, 2019.

[2] J. Huang, W. Zhang, W. Huang, W. Huang, L. Wang, and Y. Luo, “High-resolution fiber optic seismic sensor array for intrusion detection of subway tunnel,” in 2018 Asia Communications and Photonics Conference (ACP), pp. 1–3, Hangzhou, China, October 2018.

[3] C. Deng and H. Qiao, “Network security intrusion detection system based on incremental improved convolutional neural network model,” in International Conference on Communication and Electronics Systems., pp. 1–5, Coimbatore, India, 2016.

[4] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, “Network anomaly detection: methods, systems and tools,” IEEE Communication Surveys and Tutorials, vol. 16, no. 1, pp. 303– 336, 2014.

[5] N. El Moussaid and A. Toumanari, “Overview of intrusion detection using data-mining and the features selection,” in International Conference on Multimedia Computing and Systems, pp. 1269–1273, Marrakech, Morocco, 2014.

[6] A. Milenkoski, M. Vieira, S. Kounev, A. Avritzer, and B. D. Payne, “Evaluating computer intrusion detection systems,” ACM Computing Surveys, vol. 48, no. 1, pp. 1–41, 2015.

[7] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges,” Cybersecurity, vol. 2, no. 1, pp. 1–22, 2019.

[8] R. Domingues, M. Filippone, P. Michiardi, and J. Zouaoui, “A comparative evaluation of outlier detection algorithms: experiments and analyses,” Pattern Recognition, vol. 74, no. 4, pp. 406–421, 2018.

[9] Y. Mehmood, M. A. Shibli, U. Habiba, and R. Masood, “Intrusion detection system in cloud computing: challenges and opportunities,” in 2nd National Conference on Information Assurance, pp. 59–66, Rawalpindi, Pakistan, 2013.

[10] A. Drewek Ossowicka, M. Pietrołaj, and J. Rumiński, “A survey of neural networks usage for intrusion detection systems,” Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 1, pp. 497–514, 2021.

[11] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho,“A survey of network-based intrusion detection data sets,” Computers & Security, vol. 86, no. 6, pp. 147–167, 2019.

[12] A. Bakshi and Sunanda, A comparative analysis of different intrusion detection techniques in cloud computing, Springer, Singapore, 2019.

[13] S. G. Kene and D. P. Theng, “A review on intrusion detection techniques for cloud computing and security challenges,” in 2nd international conference on electronics and communication systems, pp. 227–232, Coimbatore, India, 2015.

[14] N. Keegan, S. Y. Ji, A. Chaudhary, C. Concolato, B. Yu, and D. H. Jeong, “A survey of cloud-based network intrusion detection analysis,” Human-centric Computing and Information Sciences, vol. 6, no. 1, pp. 1–16, 2016.

[15] A. Aburomman and M. B. I. Reaz, “A survey of intrusion detection systems based on ensemble and hybrid classifiers,” Computers & Security, vol. 65, no. 4, pp. 135–152, 2017.

[16] A. Nisioti, A. Mylonas, P. D. Yoo, and V. Katos, “From intrusion detection to attacker attribution: a comprehensive survey of unsupervised methods,” IEEE Communication Surveys and Tutorials, vol. 20, no. 4, pp. 3369–3388, 2018.

[17] L. N. Tidjon, M. Frappier, and A. Mammar, “Intrusion detection systems: a cross-domain overview,” IEEE Communication Surveys and Tutorials, vol. 21, no. 4, pp. 3639–3681, 2019.