IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 63-67

Deep and Machine Learning Approaches for Anomaly-Based Intrusion Detection of Imbalanced Network Traffic

Dr.Yogesh Kumar Sharma[1], Rokade Monika D[2]
[1](Head/Research-Coordinator, Department of Computer Science and Engineering, Shri J.J.T. University, Rajasthan)
[2](Research Scholar, Computer Department, Shri J.J.T. University, Rajasthan)

Abstract: *Basically anomaly detection is the part of intrusion detection system which is categorized in network intrusion detection system as well as host based intrusion detection system. Various existing systems has been developed on synthetic as well as some real time data. It system illustrates some challenges like false positive ratio of system, elevate, low accuracy these are the major challenges of anomaly based intrusion detection system. This work we proposed a deep learning based anomaly intrusion detection system which can eliminate label as well as a label attacks IDS focus on identifying possible incidents or threats, logging information, attempting to stop intrusion or malicious activities, and report it to the management station. In addition, it record information related to observed actions, notify security administrators of significantly observed actions and generate reports. Many Intrusion detection systems can also react to a detected hazard by attempting to prevent it from following. For stopping attack itself, they use numerous response techniques, altering the security surroundings for example reconfiguring a firewall or altering the attack's content. Thus IDS helps in statistical analysis for malicious behavior. In this work we proposed a Deep Learning base intrusion detection system for synthetic as well as real time network environment. Various dataset have been used to evaluate the proposed experimental analysis. The partial implementation of system shows the better results than existing systems. CIDDS-001, KKDCUP99, NSLKDD, ISCX network dataset used for evaluate the system with different algorithms..*

Keywords: *Deep Neural Network (DNN), Random Forest, Anomaly detection, Imbalanced network traffic, Variation Auto-Encoder.*

I. Introduction

Traditional approaches in network intrusion detection follow a signature-based approach, but the employment of anomaly detection approaches supported machine learning techniques are studied heavily for the past twenty years. the continual amendment in the manner attacks square measure showing, the amount of attacks, additionally because the enhancements in the huge knowledge analytics area, build machine learning approaches additional enticing than ever. The intention of this thesis is to point out that mistreatment machine learning within the intrusion detection domain ought to be accompanied with Associate in Nursing analysis of its lustiness against adversaries. Several adversarial techniques have emerged late from the deep learning analysis, largely within the space of image classification. These techniques square measure supported the concept of introducing little changes within the original input file so as to form a machine learning model to misclassify it. This thesis follows a giant knowledge Analytics methodology and explores adversarial machine learning techniques that have emerged from the deep learning domain, against machine learning classifiers used for network intrusion detection.

Intrusion detection system is software for detecting intrusion or malicious activities occurring in a particular network. Computer system goes on a high risk, when it is connected to a network. For large organizations, network security plays a very vital role. Various modifications or soft computing techniques are used such as fuzzy algorithm, genetic algorithm, apriori algorithm, artificial neural network for developing intrusion detection system.Intrusion Detection Systems (IDS) focus on identifying possible incidents or threats, logging information, attempting to stop intrusion or malicious activities, and report it to the management station. In addition, it record information related to observed actions, notify security administrators of significantly observed actions and generate reports. Many Intrusion detection systems can also react to a detected hazard by attempting to prevent it from following. For stopping attack itself, they use numerous response techniques, altering the security surroundings for example reconfiguring a firewall or altering the attack's content. Thus IDS helps in statistical analysis for malicious behavior.

Our goal is to identify novel attacks by unauthorized users in a particular network. If the vulnerability is unknown to the target's administrator or user, we consider an attack to be novel even if the attack or signature pattern is generally known. We are mainly paying attention in four types of remotely launched attacks: denial of

service (DOS), probe, U2R and R2L. A DoS attack is a type of attack in which the hacker or attacker makes a memory resources or computing resources so busy or full to serve rightful networking requests and deny users to access to a system. The examples of Dos attacks are Neptune, apache, ping of death, mail bomb, smurf, UDP storm etc. A remote to user (U2R) attack is an attack in which an attacker or hacker sends packets to a computer system over a particular network, in order to expose the machines weakness and vulnerabilities and abuse rights which a local user would have on the machine which he/she does not have access rights. The examples of U2R attacks are sendmail dictionary, xnsnoop, xlock, guest, phf, etc. A R2L attack is an attack in which attackers exploits a system by starting or accessing a system with normal authorized user account and gain user privileges. The examples of R2L attacks are xterm, perl etc. A probing is an attack in which the hacker scans a networking device or a system for determining weaknesses or vulnerabilities so as to compromise the system.

II. Literature Survey

Chang Hoon Kim et. al. [1] proposed a Classifying Malware Using Convolutional Gated Neural Network, this system carried out the Malware or Malicious Software, are an importantthreat to information technology society. Deep Neural Network has been recently achieving a great performance for the tasks of malware detection and classification. In this paper, we propose a convolutional gated recurrent neural network model that iscapable of classifying malware to their respective families. The model is applied to a set of malware divided into 9 different families and that have been proposed during the MicrosoftMalware Classification Challenge in 2015.

V Jitheshet. al. [2] LSTM Recurrent Neural Networks for High Resolution Range Profile Based Radar Target Classification, system proposed Positive and timely identification of targets iscritical in any military scenario. Target identification from backscattered electromagnetic energy is an evolving area. The objective of this paper is to study the applicability of Long Short- Term Memory Recurrent Neural Network (LSTM RNN) forHigh Resolution Range Profile (HRRP) based Radar target classification. Simulated Radar Range Profile data is used here. Three Different Target models are considered in this study. The classification is performed using a LSTM RNN.

Dehua Hong et. al. [3] proposed a system Automatic Modulation Classification using Recurrent Neural Networks which carried Automatic modulation classification (AMC) is one ofthe essential technologies, and also a hard nut to crack in the field of cognitive radio (CR) and non-cooperative communication systems. In this work, we propose a novel AMC method based on the promising recurrent neuralnetwork (RNN), which is shown to have the capability to sufficiently exploit the temporal sequence characteristic of received communication signals. This method resorts to raw signals directly with limited data length, and avoids extractingsignal features manually. The proposed method is compared with a convolutional neural network (CNN) based method and the result indicates the superiority of the proposed one, especially when signal-to-noise ratio (SNR).

MdZahangir Alomet. al. [4] proposed a system Object Recognition using Cellular Simultaneous Recurrent Networks and Convolutional Neural Network, Convolutional Neural Networks (CNNs) have become very popular and have achieved great success in many computer vision tasks – particularly in object recognition. Partially inspired by neuroscience, CNNs share many properties with the visual system of the brain. However, the filtersof convolutional layers play a vital role on overall accuracy of CNNs. In this paper, the Cellular Simultaneous Recurrent Networks (CSRNs) are applied to generate initial filters of Convolutional Networks (CNs) for features extraction and Regularized Extreme Learning Machines (RELM) are used forclassification. Furthermore, Deep Belief Networks (DBN), CNNs with random and Gabor filters are implemented to evaluate the overall performance against the proposed CSRN's filters based CNs with RELM.

Thi Thu Thoung Le et. al. [5] proposed a system Energy disaggregation or NILM is the best solution to reduce our consumption of electricity. Many algorithms in machine learning are applied to this field. However, theclassification results from those algorithms are not as well as expected. In this paper, we propose a new approach to construct a classifier for energy disaggregation with deep learning field. We apply Gated Recurrent Unit (GRU) based on RecurrentNeural Network (RNN) to train our model using UK DALE dataset on this field. Besides, we compare our approach to original RNN on energy disaggregation.

Yong Zhang et. al. [6] proposed a new architecture termed Comprehensive Attention Recurrent Neural Networks (CA-RNN) which can store preceding, succeeding and local contexts of any position in a sequenceis developed. The bidirectional recurrent neural networks (BRNN) are used to access the past and future information while a convolutional layer is employed to capture local information.The standard RNN is also replaced by two recently emerged RNN variants, namely long short-term memory (LSTM) and gated recurrent unit (GRU), to enhance the effectiveness ofthe new architecture. Another salient feature of the proposed model is that it can be trained end-to-end without any human intervention. It is very easy to be implemented. We conduct

Asmaa Salem et. al. [7] proposed a system Text Dissimilarities Predictions using Convolutional Neural Networks and Clustering, system carried out the analyze text segments of some long text andfind segments which have a different stylometry in comparison to the other. We developed two-steps method: (1) clustering of segments, and (2) classification of segments using convolutional neural networks. The method was tested on ten Arabic and ten English long texts.

Zhang, Jia-Dong et.al. [8] Proposed a system MOCA: Multi-Objective, Collaborative and Attentive Sentiment Analysis. A multi-objective, collaborative, and attentive framework called MOCA for document-level sentiment analysis. MOCA contains three key characteristics:(1) Attentive model for explicit influence.(2) Collaborative model for implicit influence.(3) Multi-objective optimization. Moca has created a new neural companion lying model based on the multilayer close-run to capture It is included in highly personalized interactions between users and objects.

Luo, Wenhan, et al. [9] proposed Trajectories as Topics: Multi-Object Tracking by Topic Discovery. An alternative approach to temporal data association by clustering detection instances, where each cluster corresponds to a unique object. One item is shown As a set of visual words and we consider it a representation. When discriminated, similar examples for the same object lead to similar samples Between different objects. Variation of object appearance is modeled as the dynamics of word co-occurrence and handled by updating the cluster parameters across the sequence in the dynamical clustering procedure.

Selvi, S. Thamarai, et al.[10] System a hybrid text categorization model that combines both Rocchio algorithm and Random Forest algorithm to perform Multilabel text categorization. Stop word remover and word Stemmer has been used to exceed the limit Rocchio algorithm. Text categorization using Rocchio algorithm and random forest algorithm. These methods find correlation between training Data sets and only related specifications filter required Classification.

III. Proposed System Overview

In our proposed system there are two phases training as well as testing modules, in training system first creates the rules or policies and testing modules detect the malicious behavior of test packets using proposed deep learning approach.

Training Phase: In this Phase, Genetic algorithm is used where, we first initialize the chromosomes and group of chromosomes we say as population is created. Once the population is created crossover is applied to obtain new generation of chromosomes. Mutation is applied for updating bit value of attributes of chromosomes randomly. The fitness function will define the fitness value of each chromosome and a selection criterion is applied for selected optimal rules. When variation is completed then Genetic algorithm will get terminated. The outputs of genetic algorithm are genetic rules. The output of genetic algorithm that is genetic rules is given as an input to fuzzy logic. In this phase probability of each attribute is calculated which is used for classification of data as attack or normal.

Testing Phase: In this Phase, Fuzzy rules are given as an input to the Neural Network algorithm for the classification of sub attack. Here system collect the network traffic data using PacketXLib and Wincap Driver. On each instance neural network algorithm will be applied. Transfer function will be used for calculating each node weight .Using Defined threshold , sub attacks can be classified.
In this proposed research work we discussed a methodology intrusion detection system using fuzzy genetic algorithm and neural network. The Expected outcome of our proposed approach is

- To implement and measure the performance of our system for that we will use the standard dataset of DARPA organization and online network dataset.
- To detect any type of malicious connection with known or unknown signature.
- To generates its own rules depending on the real-time behavior of the packet.
- To detect the subtype attack of its master class.
- To improve the Detection Rate of R2L and U2R attacks.
- To detect network based and host based attack.

IV. Algorithms Design

The proposed approach has work on multiple dataset for training as well as testing modules, below is the algorithm we used for test the network packets.

Algorithm 1 : Recurrent Neural Network
Input: Training Rules Tr[], Test Instances Ts[], Threshold T.
Output : Weight w=0.0

Step 1 : Read each test instance from (TsInstnace from Ts)
Step 2 : TsIns $= \sum_{k=0}^{n}\{Ak \dots An\}$
Step 3 : Read each train instance from (TrInstnace from Tr)
Step 4 : TrIns $= \sum_{j=0}^{n}\{Aj \dots \dots Am\}$
Step 5 : **w** = WeightCalc(TsIns, TrIns)
Step 6 : if (w >= T)
Step 7 : Forward feed layer to input layer for feedback FeedLayer[] ← {Tsf,w}
Step 8 : optimized feed layer weight, Cweigt ← FeedLayer[0]
Step 9 : Return Cweight

Algorithm 2 : Pattern Matching Algorithm
Input : Feature of BK rules TrainF[], features if test recordTestF []
Output : highest Similarity weight for class label
Step1: for all (T in TrainF [] !=null) do
Step 2.items [] split(T)
Step 3. items1 [] split(TestF)
Step 4.w = classifyToAll(Train,TestF[], Label)
Step 5. Return w;

V. Results and Discussions

NSLKDD, KDDCUP 99 from DARPA organization. ISCX dataset, Network real traffic dataset. (Using packetX and winpcap), WSNTrace Dataset (WSN data generated using NS2 simulation environment) all dataset has taken for test the application.The below figure 1 shows the detection rate for different attacks as well as error rate.

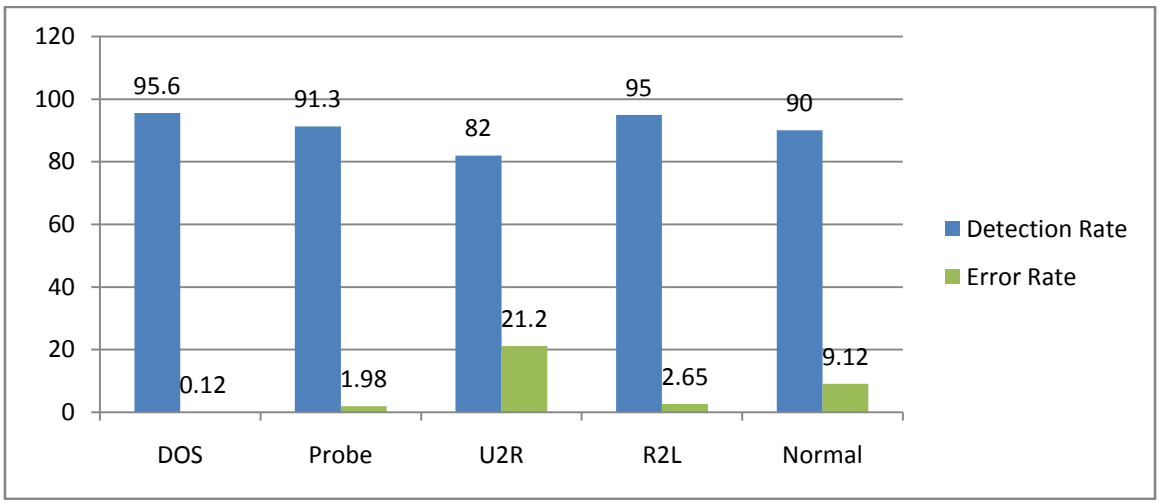

Figure 1:Performance evaluation of proposed system

The above figure 1 shows the performance evaluation of system, overall analysis shows each attack type of system. The below table 1 shows the various experiment evaluation according to various input data size.

Table 1 : Detection rate of various systems

Packet size	DOS		Probe		U2R		R2L	
	TP	FN	TP	FN	TP	FN	TP	FN
1000	99%	0.13	91%	2%	70%	30%	98%	2%
2000	99%	0.15	91%	2%	84%	16%	97%	3%
5000	99%	0.11	92%	1%	87%	13%	98%	2%
10000	99%	0.15	92%	2%	71%	29%	99%	1%
20000	99%	0.17	91%	2%	76%	24%	97%	3%
50000	99%	0.19	91%	2%	77%	23%	99%	1%
100000	99%	0.17	91%	2%	67%	33%	98%	2%

VI. Conclusion

The proposed system describe a common drawback that affects machine learning, much, is that the unbalanced category distribution drawback as a results of disproportionate categories. This study was taken off to style associate degree economical anomaly based intrusion detection system from the unbalanced network intrusion dataset and to check other ways of treating original unbalanced category distributions. to check solutions, we tend to used various metrics like preciseness, Recall, mean value, Detection Rate, and the Combined metric along side Accuracy. The experiments make sure that RNN along side down-sampling and sophistication balancer cause effective and comparable ends up in terms of accuracy. Moreover, Random Forest with less range of samples achieved high accuracy. RF is an effective methodology that has the flexibility to estimate missing information and maintain accuracy once an oversized proportion of the information is missing. Also, it's able to balance the error in unbalanced datasets. Thus, RF will be more practical for time period information fusion and applications for smaller sample sizes. Finally, the combined metric is in a position to provide higher insight on the analysis of varied systems and choosing the simplest among them.

References

[1]. Kim CH, Kabanga EK, Kang SJ. Classifying malware using convolutional gated neural network. In2018 20th International Conference on Advanced Communication Technology (ICACT) 2018 Feb 11 (pp. 40-44). IEEE.

[2]. Jithesh V, Sagayaraj MJ, Srinivasa KG. LSTM recurrent neural networks for high resolution range profile based radar target classification. In Computational Intelligence & Communication Technology (CICT), 2017 3rd International Conference on 2017 Feb 9 (pp. 1-6). IEEE.

[3]. Hong D, Zhang Z, Xu X. Automatic modulation classification using recurrent neural networks. In Computer and Communications (ICCC), 2017 3rd IEEE International Conference on 2017 Dec 13 (pp. 695-700). IEEE.

[4]. Alom MZ, Alam M, Taha TM, Iftekharuddin KM. Object recognition using cellular simultaneous recurrent networks and convolutional neural network. In2017 International Joint Conference on Neural Networks (IJCNN) 2017 May 14 (pp. 2873-2880). IEEE.

[5]. Kim J, Kim H. Classification performance using gated recurrent unit recurrent neural network on energy disaggregation. InMachine Learning and Cybernetics (ICMLC), 2016 International Conference on 2016 Jul 10 (Vol. 1, pp. 105-110).IEEE.

[6]. Zhang Y, Er MJ, Venkatesan R, Wang N, Pratama M. Sentiment classification using comprehensive attention recurrent models. InNeural Networks (IJCNN), 2016 International Joint Conference on 2016 Jul 24 (pp. 1562-1569). IEEE.

[7]. Salem A, Almarimi A, Andrejková G. Text Dissimilarities Predictions Using Convolutional Neural Networks and Clustering. In2018 World Symposium on Digital Intelligence for Systems and Machines (DISA) 2018 Aug 23 (pp. 343-347). IEEE.

[8]. Zhang, Jia-Dong, and Chi-Yin Chow. "MOCA: Multi-Objective, Collaborative and Attentive Sentiment Analysis." IEEE Access (2019).

[9]. Luo, Wenhan, et al. "Trajectories as Topics: Multi-Object Tracking by Topic Discovery." IEEE Transactions on Image Processing 28.1 (2019): 240-252.

[10]. Selvi, S. Thamarai, et al. "Text categorization using Rocchio algorithm and random forest algorithm." Advanced Computing (ICoAC), 2016 Eighth International Conference on.IEEE, 2017.

[11]. Dr. Yogesh Kumar Sharma, "Critical Study of Software Models Used Cloud Application Development", E-ISSN: 2227-524X, Volume No. 7, Issue No. 3.29, pp. 514-518, 2018.

[12]. Dr. Yogesh Kumar Sharma, "Effective Multilayered Energy Harvesting and Aggregation in underwater Acoustic Networks for Performance Enhancement", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN(Online)- 2278-8875, ISSN(Print)- 2320-3765, Volume No. 6, Issue No. 7, pp. 5821-5829, 07 July 2017.

[13]. Dr. Yogesh Kumar Sharma, "Performance Evaluation of Delay Tolerant Networks Routing Protocols Under Varying Time of Live", International Journal of Advance Research in Computer Science, ISSN- 0976-5697, Volume No. 8, Issue No. 1, pp. 299-302, Jan-Feb 2017.