ISSN: 2581-3404 (Online)

International Journal of Innovative Research in Technology and Management (IJIRTM), Volume-3, Issue-3, 2019

Network Based Intrusion Detection using Attribute Selection and Attribute Classification: Survey & Discussions

Kiran Vanpure[1], Prakash Mishra[2]

[1]M. Tech Scholar, Department of IT, PCST, Indore (India)

[2]Assistant Professor, Department of CSE, PCST, Indore (India)

[1] kiranvanpure777@gmail.com, [2]prakash.mishra@patelcollege.com

ABSTRACT

In recent trends, the services based on the network and sharing the information through network has tremendous growth. Because of this, there may be chance for the existence of intrusions in the network. In order to harden the systems against intrusion, network security is the most important aspect. The traditional techniques are utilized for protecting the information in the network. Among these, intrusion detection system is the most significant methods. Even though, they had several advantages in detecting the intrusions. Still it has some issues such as inaccurate classification results, increased false alarm rate, etc. In this article we presents the comparative literature study including the various classification methods for the network based intrusion detection system.

Keywords: Intrusion Detection, Denial of Services, Neural Network, Classification, Clustering.

INTRODUCTION

Information guarantee is a concern of serious global concern. The intricacy and openness of client/server technology pooled with Internet contain fetch about enormous profit to the progressive society; meanwhile, the hurriedly increasing, openness, high complexity, and increasing accessibility of the networks not only escort to exploitation of vulnerabilities in the communication protocol stack but moreover enlarge the risk of existing information security system. Network attack vulnerability depends upon the expensive information hacking by attackers. The attacker intrude the network system or system server and creating network dump, malicious activity, modification, data theft, flood or denial the system process. Network system affecting the lack of attacks and thus require to intellectual intrusion detection model to protect the network system [1].

The Internet has become a crucial part of everyday communication via social media interaction, e-mail, e-learning, etc. Besides, small and large corporations have extended their consumer base by providing direct customer marketing, internet shopping and inter-company correspondence using basic Internet communication. No doubt, some risks are incurred owing to the use of ineffective and inefficient security tools inviting intrusions from Internet hackers. Thus, it is evident that the prevention technologies in place like malware removal programs, antivirus programs and firewalls, fail to provide absolute protection since attackers employ newer techniques for assaulting the network as well as its users [9]. In recent years, feature selection is often used in intrusion detection system, because it can enhance the accuracy of classifier. Therefore, the use feature selection to get the best subset features, making the detection faster and more efficient. Although researchers found the best solution which significantly shortens work time, they did not verify the accuracy and performance on a real detection system and did not clearly describe the steps on traffic collection [8].

Nowadays, the Internet is accessible from all over the place. With the growing number of electronic devices connected to the web, computer network security could be endangered. This problem has raised the question on how to effectively defend the computer network from internal and external attacks [9]. Intrusion Detection System (IDS) could be the first line defense mechanism to detect intrusion before computer network is endangered. This section covers the discussion of the definition of an IDS, types of IDS and the software used for detection.

Today, the Internet faces threats from intelligent, automated and sophisticated malicious codes that are on the rise. It could be seen in the past that computer worms have the capability to disperse on their own, without human involvement and have the record of launching the worst attacks on computer networks. To provide defense against worms, intrusion detection systems are mostly employed that make use of self-replicating behavior of worms for detecting the signatures and patterns of malicious codes in the network. By the parameters they use for detection, these systems can be categorized as anomaly-based and signature-based systems.

"Intrusion Detection Systems" (IDS) became extremely vital due to the fact that networks might be threatened by both internal and external intruders' attacks. It is defined as a detection system located to observe and monitor computer networks requests. These have been in use since the 1980's. The threats can have harmful damages such as: Denial of service (DoS) which causes prevention of legitimate users from using network resources by streaming irrelevant heavy traffic. Malware also could also cause harm, where attackers use malicious software to foul up systems. IDS is evolving as a response to current and future attacks from internal and external intruders. Most IDS have been developed and implemented to be strict system but they have suffered from the problem of "false positive" or "false negative" alarms. This high rate of false detection causes IDS to lose credibility in practical large scale systems [14].

The rest of this paper is organized as follows in the first section we describe an introduction of about the content based image classification introduction. In section II we discuss about the wavelet transform function. In section III we discuss about the texture classification. In section IV we present rich literature for content based image classification. In section V we discuss about the problem formulation and statement as we getting from the rich literature survey, finally in section VI we conclude the about our paper which is based on the literature survey and specify the future scope.

II RELATED WORK

In this section we discuss about the literature survey for the host based and network based intrusion detection system using various classification techniques and other techniques.

In [1] a concise, and easy to use statistical learning procedure, abbreviated NASCA, which is a four-stage intrusion detection method that can successfully detect unwanted intrusion to our systems. The model is static, but it can be adapted to a dynamic set up. The classification agent is capable of making decisions in a constantly changing environment and therefore testing the model, while evaluating the network.

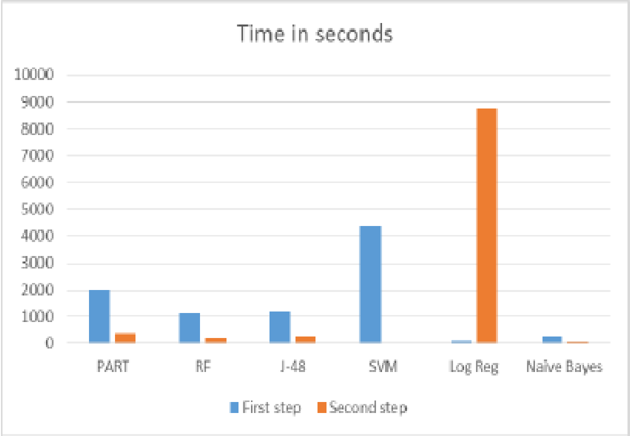
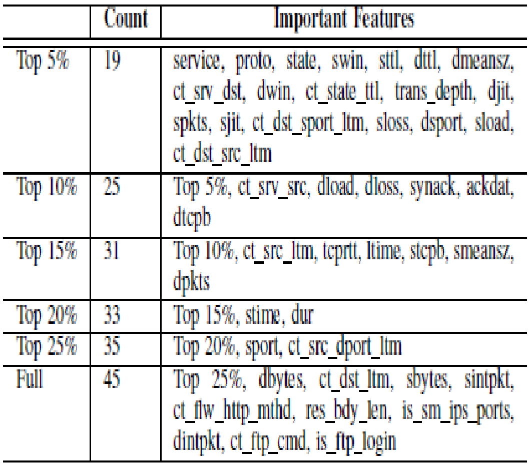


Fig. 1: Time in Seconds to Build a Model.

In [2] a deep learning binomial classifier for Network Intrusion Detection System is proposed and experimentally evaluated using the UNSW-NB15 dataset. Three different experiments were executed in order to determine the optimal activation function, then to select the most important features and finally to test the proposed model on unseen data. The evaluation results demonstrate that the proposed classifier outperforms other models in the literature with 98.99% accuracy and 0.56% false alarm rate on unseen data.



	Count	Important Features
Top 5%	19	service, proto, state, swin, sttl, dttl, dmeansz, ct_srv_dst, dwin, ct_state_ttl, trans_depth, djit, spkts, sjit, ct_dst_sport_ltm, sloss, dsport, sload, ct_dst_src_ltm
Top 10%	25	Top 5%, ct_srv_src, dload, dloss, synack, ackdat, dtcpb
Top 15%	31	Top 10%, ct_src_ltm, tcprtt, ltime, stcpb, smeansz, dpkts
Top 20%	33	Top 15%, stime, dur
Top 25%	35	Top 20%, sport, ct_src_dport_ltm
Full	45	Top 25%, dbytes, ct_dst_ltm, sbytes, sintpkt, ct_flw_http_mthd, res_bdy_len, is_sm_ips_ports, dintpkt, ct_ftp_cmd, is_ftp_login

Fig. 2: Top important features.

In [3] hybrid efficient model used to analyze the optimal features in the data, and it improve the detection rate and time complexity effective. This approach deals with high false and low false negative rate issue, first pre-processed data should be correlation based particle swarm optimization with GR-CR (Gain Ratio & Co-Relation) combination of this approach provide learning based some important subset of features and shows progress in the accuracy and time complexity level.

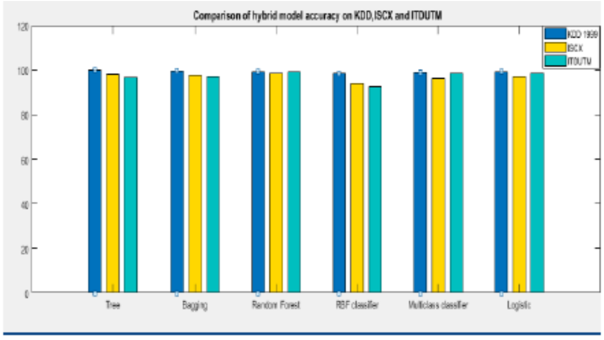


Fig 3: Comparison results of classification.

In [4] a system that enables proactive defenses at the level of a single browsing session. By observing user behavior, it can predict whether they will be exposed to malicious content on the web seconds before the moment of exposure, thus opening a window of opportunity for proactive defenses.

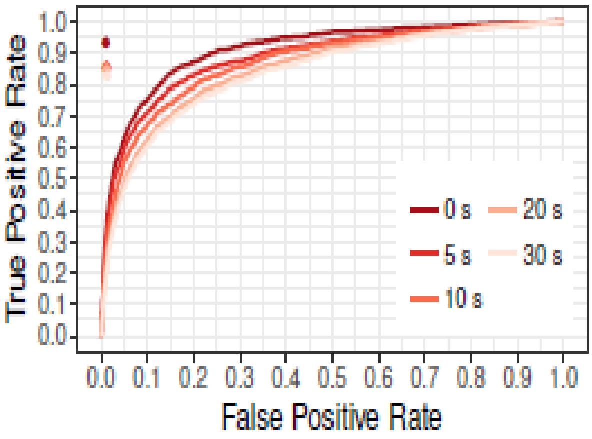


Fig 4: Performance of different time using true and false positive rate.

In [6] they propose an approach to bridge measurement data with manual analysis. They borrow the idea from threat intelligence: we define campaigns using a 4-stage model, and describe each stage using IOCs (indicators of compromise), e.g. URLs and IP addresses. They train a multi-class classifier to extract IOCs and further categorize them into different stages.

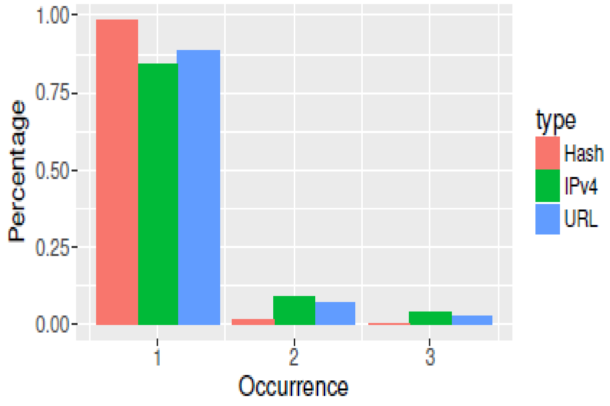


Fig 5: Percentage of the IOC occurrence in technical blogs for file hash, URL and IP address. The distribution of IOC occurrence is highly skewed, and therefore we only show the percentage of the occurrence less than 3.

In [8] some new features to detect the botnet traffic, and we found the best solutions by using feature selection algorithm. These two methods are particle swarm optimization and genetic algorithms, and by using back propagation network as the classifier, they evaluate our subset feature on botnet detection that shows high detection rate, and they validate that own manufactured feature packet transmission time of regularity can be adopted, and the accuracy will change with the t-value.

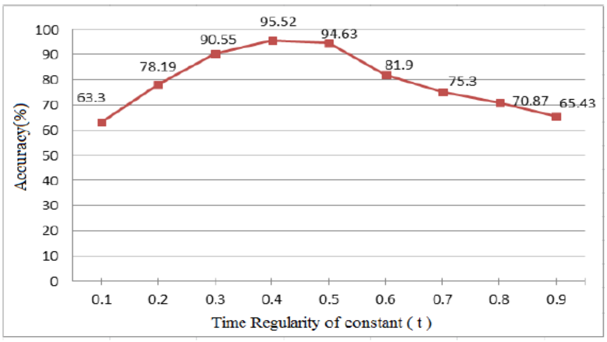


Fig 6: Time regularity of accuracy with t value.

II ATTRIBUE SELECTION

Attribute selection is the method of recognizing and eliminating the irrelevant and redundant information from an evaluated system. If we are able to eliminate some of the unrelated features, we can moderate the complexity, by removing irrelevant dimensions and to enhance the performance of the prospective classification procedure. The decision agent is capable of operating not only quicker and with less information, but also to improve the classification accuracy process, [1]. The main idea of the attribute selection procedure is to rank the relevant variables and henceforth to use only the appropriate information in order to perform the classification of the network. Attribute reduction is the process of mapping the existing high-dimensional data onto a lower dimensional space.

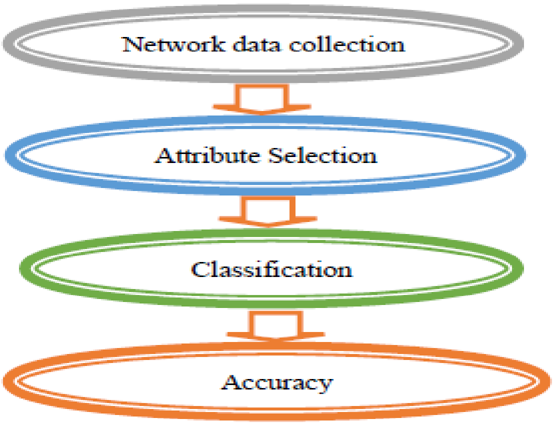


Fig 7: Network Attribute Selection, Classification and Accuracy Procedure.

III CLASSIFICATION PROCESS

The classification method is an essential step of the proposed procedure. It is a two-stage process with objectives accuracy, precision and faster classification. Therefore, in order to accelerate the procedure, supplementary analyses will be completed, only whenever the network is classified as being under attack. The classification is accomplished based on the so called majority score split [1]. As an addition, it can be executed if needed with two major procedures in order to perform the classification or prediction analysis, namely boosting and bagging. On one hand, in boosting, the succeeding trees assign additional weight to instances that were incorrectly classified by earlier trials and at the end, a weighted score is calculated for the classification purposes. On the other hand, in bagging, the succeeding trees are independent from the previous trees, moreover every tree is grown by means of a bootstrap sample. Random forest grows multiple trees and each of them produces a classification with an assigned score for the specific class. As a result, the forest indicates the classification with the highest score.

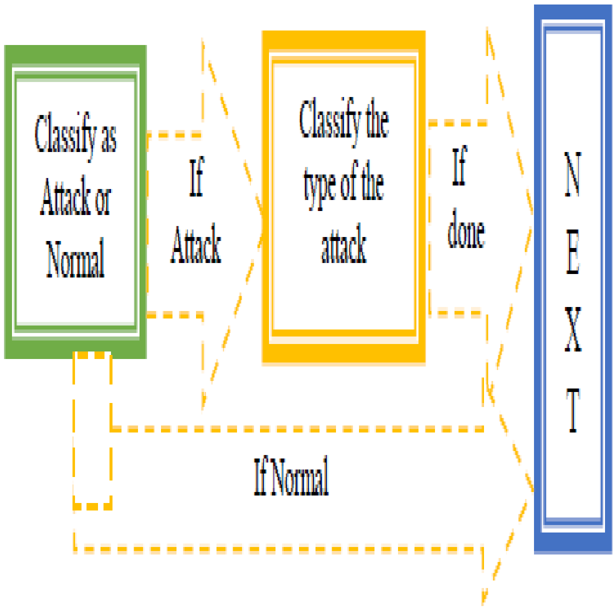


Fig 8: Classification Process.

IV CLASSIFICATION OF ATTACK TYPES

The dataset for the proposed system was selected from the University of California, Irvine Knowledge Discovery in Databases (KDD) Archive. This dataset is a collection of information related to the network gathered over some time-period. This data comprises of various features such as protocol type (UDP, TCP, ICMP), connection duration, type of service (HTTP, FTP, Telnet), total number of bytes transmitted to source host, total number of bytes transmitted to destination host, quantity of urgent packets, if destination and source addresses are same, quantity of wrong packets. There are about 41 attributes and a single target in each record. The value of the target depicts the attack name. For every connection, there are up to 41 features.

Particularly, a connection can be defined as a series of TCP packets that start and end at distinct instants of time and the data flow from the source to target IP address occurs as per a well-defined protocol. The various features are classified into four classes [9]:

- Basic Features: These features can be obtained from packet headers with no inspection of the payload. It comprises of features like the flag, duration, type of protocol and type of service.

- Content Features: The payload of actual TCP packets can be accessed using domain knowledge, and it comprises of features like the number of login attempts failed.

- Traffic Features based on time: These are designed for capturing properties which mature in a temporal window of 2 seconds. Examples include connections to a single host in 2 seconds.

VI CONCLUSIONS AND FUTURE WORK

In the ancient times, there were few intruders and so the user can manage them easily from the known or unknown attacks [3]. In recent years the security becomes the most serious problem in issues of securing data or information year over year. Because the intruders introduce a new variety of intrusions in the market, so that user can't manage their computer system or network. In this paper we discuss about the rich literature survey for the intrusion detection system using various classification techniques, in the near future we improve the detection ratio for the vairous intruder.

.REFERENCES:-

[1] Zheni Stefanova, Kandethody Ramachandran, "Network Attribute Selection, Classification and Accuracy (NASCA) Procedure for Intrusion Detection Systems", IEEE 2017. pp 1-7.

[2] Malek Al-Zewairi, Sufyan Almajali, Arafat Awajan, "Experimental Evaluation of a Multi-Layer Feed-Forward Artificial Neural Network Classifier for Network Intrusion Detection System", International Conference on New Trends in Computing Sciences, IEEE 2017. pp 167-172.

[3] Sivasangari Gopal, Sathya M, " A Feature Selection for Intrusion Detection System Using a Hybrid Efficient Model", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2018. pp 1917-1929.

[4] Mahmood Sharif, Jumpei Urakawa, Nicolas Christin, Ayumu Kubota, Akira Yamada, "Predicting Impending Exposure to Malicious Content from User Behavior", In Proceedings of 2018 ACM SIGSAC Conference on Computer & Communications Security, 2018. pp 1487-1501.

[5] Win, Thu Yein, Tianfield, Huaglory, Quentin Mair, " Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing", IEEE Transactions on Big Data (99), 2017. pp. 1-15.

[6] Ziyun Zhu, Tudor Dumitras, "ChainSmith: Automatically Learning the Semantics of Malicious Campaigns by Mining Threat Intelligence Reports", IEEE 2016, pp 1-15.

[7] Kasun Amarasinghe, Milos Manic, "Improving User Trust on Deep Neural Networks based Intrusion Detection Systems", Accepted version of the paper appearing in the proceedings of the 44th Annual Conference of the IEEE Industrial Electronics Society, 2018. pp 1-8.

[8] Jen-Li Liao, Kuan-Cheng Lin, Jyh-Yih Hsu, "Botnet detection and feature analysis using Back propagation neural network with bio-inspired algorithms", Int. J. Cognitive Performance Support, Vol. 1, No. 2, 2018, pp 132-142.

[9] Rashidah Funke, Olanrewaju, Burhan Ul Islam Khan, Athaur Rahman Najeeb, Ku Nor Afiza Ku Zahir, Sabahat Hussain, "Snort-Based Smart and Swift Intrusion Detection System", Indian Journal of Science and Technology, Vol 11, 2018. pp 1-9.

[10] Xiao Wang, Quan Zhou, Jacob Harer, Gavin Brown, Shangran Qiu, Zhi Dou, John Wang, Alan Hinton, Carlos Aguayo Gonzalez, Peter Chin, "Deep learning-based classification and anomaly detection of side-channel signals", Proc. of SPIE, 2018. pp 1-9.

[11] Hossam Faris, Alao M. Al-Zoubi, Ali Asghar Heidari, Ibrahim Aljarah, Majdi Mafarja, Mohammad A. Hassonah, Hamido Fujita, "An Intelligent System for Spam Detection and Identification of the most Relevant Features based on Evolutionary Random Weight Networks", Information Fusion (2018), pp 1-28.

[12] Kasun Amarasinghe, Kevin Kenney, Milos Manic, "Toward Explainable Deep Neural

Network based Anomaly Detection", IEEE 2018, pp 1-8.

[13] Gavin Watson, " A Comparison of Header and Deep Packet Features when Detecting Network Intrusions", 2018, pp 1-10.

[14] Ahmed Elsherif, " Automatic Intrusion Detection System Using Deep Recurrent Neural Network Paradigm", JISCR, 2018. pp 28-41.

[15] Ruby Sharma, Sandeep Chaurasia, " An Integrated Perceptron Kernel Classifier for Intrusion Detection System", I. J. Computer Network and Information Security, 2018, pp 11-20.