A Survey on Intrusion Detection Techniques using Classifier

Deepika Nayak[1], Dr. Sadhna K. Mishra[2]

[1]M. Tech Scholar, Department of CSE, LNCT, Bhopal (India)

[2]Head & Professor, Department of CSE, LNCT, Bhopal (India)

[1]deepikanayakkori@gmail.com, [2]sadhnamanit@yahoo.com

ABSTRACT

In today's word internetworking change the concept of real life and provide various services and advantage based on the internet, due to increasing the popularity and usage of internet, security is a very challenging task. Computer security is two types one is host based and the other one is network based, sometime attack detection is also called intrusion detection. Intrusion detection system offers various tools and techniques for the ensuring to develop and remain more secure of our network. In this paper our aim to study various attacks and threats detection with their proposed techniques using for the network based detection system.

Keywords: - Intrusion Detection System, KDDCUP, DDOS, Classification, Accuracy.

INTRODUCTION

Today many businesses rely on computer networks. These networks fulfill the needs of business, enterprises and government agencies to build knowledgeable, complicated information networks which integrate various technologies such as distributed data storage systems, encryption techniques, voice over IP (VoIP), remote or wireless access and web services [7]. 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.

Intrusion detection system is a software application or a device placed at strategic places on a network to monitor and detect anomalies in network traffic [6]. 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 [9]. The main features of IDS are to raise an alarm when an anomaly is detected. A complementary approach is to take corrective measures when anomalies are detected; such an approach is referred to as an intrusion Prevention System (IPS) [6]. Intrusion detection attacks can be classified into two groups: misuse or signature based and anomaly based intrusion detection [3]. Over the last decade, Tor traffic classification has advanced in its applications in systems like quality of service (QoS) tools or Security information and Event management (SIEM) [7].

A considerable interest have been attracted from researchers and the industries to the study of these technologies and developing classification techniques [6]. NSL_KDD dataset and Defense Advanced Research Projects Agency (DARPA) datasets are widely used as training and testing dataset for intrusion detection system. The NSL_KDD dataset provides 41 features in dataset to train and test the intrusion detection system. But, all 41 features in dataset are not relevant and required for training and testing purpose. If all features are used to train intrusion detection system, model building time unnecessary increased. The relevant feature selection is very essential process in intrusion detection method [2].

The various network attacks can be categorized as User-to-Root (U2R) attacks, Denial-of-Service (DoS) attacks, Remote-to-Local (R2L) attacks and probe attacks. In DoS attack, the attacker interrupts or denies the user access to the server. Examples are Neptune, Ping of Death, Mailbomb, etc. In U2R, the attacker is allowed privileged access by the extension of the root permissions like those of the administrator, the most common example being the buffer overflow attack. In R2L attack, the assailant intrudes the target system illegally, without any permission from the owner. Last of all, probe attack gathers and analyses information with an aim to map the network system, e.g., scanning software like Satan, Mscan and Nmap collect information from the target system such as hostname, service application, IP address and operating system [9].

Recently, numerous researchers have studied the support of machine learning algorithms for IDSs. Machine learning is a field of computer science that trains the computer to think like humans and take actions where required. In simple processing, a computer processes the information based on statements from primary memory. Machine learning techniques try to copy thinking processes such as logical reasoning, intuition, learning from the past, trial and error and generalizations [7].

The standard General Regression Neural Network (GRNN) regresses through a sample dataset to develop a general model for classification. The GRNN consists of a hidden layer of Gaussian neurons. The neurons within the GRNN are kernels created using individual training instances. The kernels are then used to calculate fire strengths (weights), which are a measure of similarity, for candidate instances [4].

Naïve Bayes classifier, known as a conditional probability model, is one of the most useful and efficient learning algorithms. This method works based on the Baye's theorem and also a strong assumption that is defined as Conditional Independence and supposes that the probability of one feature does not have any effect on the probability of the other ones [11].

Support Vector Machines (SVM) is a machine learning algorithm that learns to classify data using points labelled training examples falling into one or two classes. The SVM algorithm builds a model that can predict if a new example falls into one category or the other [6]. Least square support vector machines (LS-SVM) are the modified version of support vector machines. LS-SVM has been used different purposes such as for adaptive communication channel equalization, to study the nonlinear time series prediction on Morlet Wavelet kernel function for facial gender classification and for measurement of soluble solids content of rice vinegars [24]. Although LS-SVM is significant, it has not yet being used for detecting intrusions [8].

Genetic Algorithm (GA) approach with an enhanced starting populace and determination administrator, to proficiently identify different sorts of system intrusions. GA is utilized to enhance the inquiry of assault situations in review records, because of its great adjust investigation/misuse; it gives the subset of potential attacks which are available in the review document in a sensible handling time [10].

The rest of this paper is organized as follows in the first section we describe a introduction of about Intrusion detection system and their techniques. in section II we discuss about the rich literature survey for the about Intrusion detection system and various researchers techniques for the same. In section III we discuss about the problem formulation and statement as we getting from the rich literature survey, In section IV we show a table for KDDCUP Features Dataset, finally in section V 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 previous work done in the field of intrusion detection system using various techniques such as some classification techniques, evolutionary techniques and optimization methods, these sections further describe various author research work for the security of system and after that we formulate a problem for the solution in future work. Gaby Abou Haidar, Charbel Boustany et al. emphasizes the importance of anomaly-based intrusion detection techniques [1], the important outcomes of these systems, latest developed methods and what is expected from the future experiments in this field. Moreover, the technique of learning user profiles effects in detecting intrusions will be discussed. D.P.Gaikwad, Ravindra C. Thool et al. The Bagging method of ensemble with REPTree as base class is used to implement intrusion detection system. The relevant features from NSL_KDD dataset are selected to improve the classification accuracy and reduce the false positive rate [2]. James Brown, Mohd Anwar, Gerry Dozier et al. implemented an Evolutionary General Regression Neural Network (E-GRNN) as a two-class classifier for intrusion detection based on features of application layer protocols (e.g., http, ftp, smtp, etc.) used in simulated network traffic activities. Elike Hodo, Xavier Bellekens, Ephraim Iorkyase, Andrew Hamilton, Christos Tachtatzis, Robert Atkinson et al. focuses on the classification of Tor traffic and nonTor traffic to expose the activities within Tor traffic that minimizes the protection of users. A study to compare the reliability and efficiency of Artificial Neural Network and Support vector machine in detecting nonTor traffic in UNB-CIC Tor Network Traffic dataset is presented in this paper [6]. Enamul Kabir, Jiankun Hu, Hua Wang, Guangping Zhuo et al. proposes a novel approach for intrusion detection system based on sampling with Least Square Support Vector Machine (LS-SVM). Decision making is performed in two stages. In the first stage, the whole dataset is divided into some predetermined arbitrary subgroups. The proposed algorithm selects representative samples from these subgroups such that the samples reflect the entire dataset. L. Khalvati, M. Keshtgary, N. Rikhtegar et al. A hybrid approach is proposed towards achieving a high performance. In fact, the important goal of this paper is to generate an efficient training dataset. In order to exploit the strength of clustering and feature selection, an intensive focus on intrusion detection combines the two, so the proposed method is using these techniques as well. Pratham Harshit Rajmahanty, S. Ganapathy et al. presents many decision tree algorithms which are proposed by various researchers in the past for effective decision making on intrusion detection systems and some other decision making systems. Moreover, a comparative analysis also made in this paper for demonstrating their capability [12].

III PROBLEM STATEMENT

Intrusion detection process is very complex process in network security. In current network security scenario various types of attack family are available some are known family and some are unknown family. The family of know attack or malware detection used some well know technique such as signature based technique and rule based technique. There are various techniques we can apply for the pre-processing and classification of datasets for the enhanced the security system. The soft computing and data mining approach of network malware classification technique suffered from classification rate and false alarm generation, the detection rate is very high as well as very challenging task to detect some unknown attack because sometime our current system is not able to detect a very newly generated attack, there are various issues for the further discussion and implementation work some of defined below:

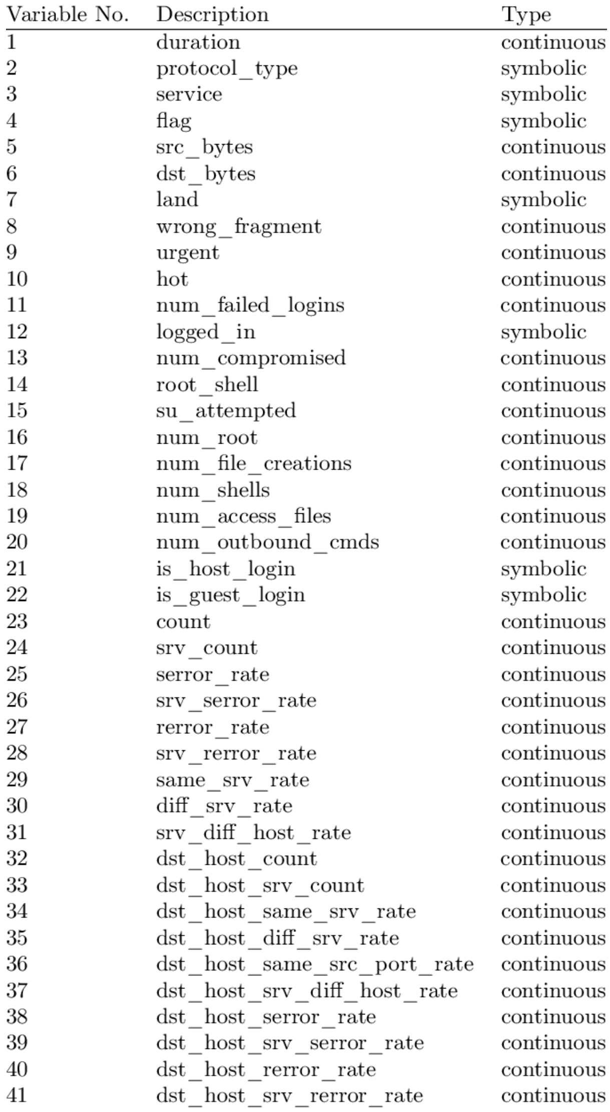
1. The pre-processing of KDDCUP99 takes more time.
2. The rate of false alarm generation is high.
3. Entropy based malware classification technique suffered by high false rate.
4. Outlier boundary detection value is high.

IV KDDCUP

This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 the Fifth International Conference on Knowledge Discovery and Data Mining [14].

Four main class of attack	**categories of attack**
Denial of Service (DoS)	back, land, neptune, pod, smurt, teardrop
Remote to User (R2L)	ftp_write, guess_passwd, imap, multihop, phf,spy, warezclient, warezmaster
User to Root (U2R)	buffer_overflow, perl, loadmodule, rootkit
Probing(Information Gathering)	ipsweep, nmap, portsweep, satan

Table 1: **Classification of attack types [9].**



Variable No.	Description	Type
1	duration	continuous
2	protocol_type	symbolic
3	service	symbolic
4	flag	symbolic
5	src_bytes	continuous
6	dst_bytes	continuous
7	land	symbolic
8	wrong_fragment	continuous
9	urgent	continuous
10	hot	continuous
11	num_failed_logins	continuous
12	logged_in	symbolic
13	num_compromised	continuous
14	root_shell	continuous
15	su_attempted	continuous
16	num_root	continuous
17	num_file_creations	continuous
18	num_shells	continuous
19	num_access_files	continuous
20	num_outbound_cmds	continuous
21	is_host_login	symbolic
22	is_guest_login	symbolic
23	count	continuous
24	srv_count	continuous
25	serror_rate	continuous
26	srv_serror_rate	continuous
27	rerror_rate	continuous
28	srv_rerror_rate	continuous
29	same_srv_rate	continuous
30	diff_srv_rate	continuous
31	srv_diff_host_rate	continuous
32	dst_host_count	continuous
33	dst_host_srv_count	continuous
34	dst_host_same_srv_rate	continuous
35	dst_host_diff_srv_rate	continuous
36	dst_host_same_src_port_rate	continuous
37	dst_host_srv_diff_host_rate	continuous
38	dst_host_serror_rate	continuous
39	dst_host_srv_serror_rate	continuous
40	dst_host_rerror_rate	continuous
41	dst_host_srv_rerror_rate	continuous

Table 2: **KDDCUP Features [14].**

V CONCLUSIONS AND FUTURE WORK

Intrusion detection system main task to differentiate between normal and abnormal activity of the system, each and every activity performed by the system is determined to normal or not. If the rate of abnormal activity is high then we have to produce alert signal for the system. In this paper we study various research paper for the intrusion detection system using various techniques and algorithms, in future we will develop a secure intrusion detection system mechanism whose create a difference between the normal and abnormal system activity and generate a more efficiently alert signal for the our system.

REFERENCES:-

[1] Gaby Abou Haidar, Charbel Boustany, "High Perception Intrusion Detection Systems Using Neural Networks", Ninth International Conference on Complex, Intelligent, and Software Intensive Systems, IEEE 2015. Pp 497-501.

[2] D.P.Gaikwad, Ravindra C. Thool, "Intrusion Detection System Using Bagging Ensemble Method of Machine Learning", International Conference on Computing Communication Control and Automation, IEEE 2015. Pp 291-295.

[3] Preeti Singh, Amrish Tiwari, "An Efficient Approach for Intrusion Detection in Reduced Features of KDD99 using ID3 and classification with KNNGA", Second International Conference on Advances in Computing and Communication Engineering, IEEE 2015. Pp 445-452.

[4] James Brown, Mohd Anwar, Gerry Dozier, "An Evolutionary General Regression Neural Network Classifier for Intrusion Detection", IEEE, 2016. Pp 1-5.

[5] Duygu Sinanc Terzi, Ramazan Terzi, Seref Sagiroglu, "Big Data Analytics for Network Anomaly Detection from Netflow Data", IEEE, 2017. Pp 592-598.

[6] Elike Hodo, Xavier Bellekens, Ephraim Iorkyase, Andrew Hamilton, Christos Tachtatzis, Robert Atkinson, "Machine Learning Approach for Detection of nonTor Traffic", ARES 2017. Pp 1-6.

[7] Syed Ali Raza Shah, Biju Issac, " Performance Comparison of Intrusion Detection Systems and Application of Machine Learning to Snort System", 2017. Pp 1-25.

[8] Enamul Kabir, Jiankun Hu, Hua Wang, Guangping Zhuo, "A Novel Statistical Technique for Intrusion Detection Systems", Pp 1-39.

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

[10] Wrushal K. Kirnapure, Arvind R. Bhagat Patil, " Survey on Classification, Detection and Prevention of Network Attacks using Rule based Approach", International Journal of Computer Applications, 2017. Pp 11-17.

[11] L. Khalvati, M. Keshtgary, N. Rikhtegar, "Intrusion Detection based on a Novel Hybrid Learning Approach", Journal of AI and Data Mining, 2018, Pp 157-162.

[12] Pratham Harshit Rajmahanty, S. Ganapathy, " Role of Decision Trees in Intrusion Detection Systems: A Survey", International Journal of Advances in Computer and Electronics Engineering, 2017. Pp 9-13.

[13] Sheetal Panjeta, Er. Kanika Aggarwal, "Review paper on Different Techniques in Combination with IDS", International Journal of Engineering Science and Computing, May 2017. Pp 11623-11630.

[14] http://kdd.ics.uci.edu/databases/kddcup99/