International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue IX, September 2017- Available at www.ijraset.com

Intrusion Detection System And Mlsvm Classifier: A Survey

Tanvi upadhyay[1], Shushil chaturvedi[2]
[1]*Dept. of (CSE/IT)), S.R.C.E.M College, Gwalior (India)*
[2]*PROF. Dept. of. (CSE/IT), S.R.C.E.M College, Gwalior (India)*

Abstract: ***In today's scenario to maintain the security of network system is important. We need a secure and safe network system towards intruders attack. Intrusion detection system is used for identifying the various types of attack in a network. IDS are available in various types network based, host based and hybrid based on the technology detected by them in market. Firewall and IDS advance the security of network components. Intrusion detection system uses a security rule to detect unusual activity. These rules are decided by the administrator based on the requirements of the organization. SVM is a technique which is worn in data mining to mine data which is predicted. MLSVM Classifier for misuse detection is constructed using the kddcup 99 data which first classifies the data into normal and attacks and then into various classes of attacks. MLSVM Classifier for masquerader detection is constructed which classifies the data into masquerader and non-masquerader first and then into various categories of the masqueraders.***

Keywords: Intrusion Detection System, Multilevel support vector machine, k-means clustering,etc.

I. INTRODUCTION

IDS are grouping of hardware, software so as to monitors a system or the network of the systems against any type of activity which is harmful. This is used for detecting misuse of the network. IDS are producing the indication for the network as much like a alarm, IDS detects the occurrence of an attack in the network and increase an alert. Intrusion Detection System provides following main process: monitor, detect and produce an alert IDS are some time measured as the functionality of firewall. There is a bit variation between two of them. A firewall may be as a barrier that protects the information flow and avoid intrusions where as IDS detects if the network is under attack or if the security compulsory by the firewall has been infracted. Firewall and IDS advance the security of network components. Intrusion detection system uses a security rule to detect unusual activity. These rules are decided by the administrator based on the requirements of the organization. Any movement that break this security policy will be measured a security threat and will be reported to the higher administration by email or as SNMP traps. These policies are required to be updated frequently to keep up with the threats and requirements of the safety incidents that happen on a network, the huge majority come from the network. Such attacks consist of otherwise allowed users who are dissatisfied employees. The residue come from the exterior, in the type of denial of service attacks or attempts to penetrate through a network infrastructure. IDS stay on the only proactive means of caught and responding to threats that branch from both inside and outside a commercial network. There are following types of Intrusion Detection Systems [1]:

A. Hot Based IDs
B. Network Based IDs
C. Stack Based IDs
D. Signature Based IDs
E. Anomaly Based IDs

II. INTRUSION DETECTION TECHNIQUES

Here, we discuss in brief the most general techniques that used to discover the intrusions: Artificial Neural Networks (ANNs) this Artificial neural networks give the flexible pattern of recognition capabilities. Inside ANNs, unique type of training is specified to the system so that it can identify different patterns of the arbitrary that are provided to it as the input data. When system completely recognizes these all patterns it is after that asked to match these patterns with production produced. By matching a variety of input and the output arbitrary patterns, it is found that intrusion has taken place or not. State Transition Tables in State Transition Table, sequence of actions performed by an intruder is described in the form of a state transition diagram and behavior of the system is observed. When it matches with identifiable compromised state and penetrated sate, an intrusion is detected. Genetic Algorithms (GAs) the function of Genetic Algorithms (GAs) is to imitate or mimic the natural reproduction system in nature. After

going through recombination and a variety of changes which are random , only fittest person will be reproduced in the following generations. In 1995, the application of GAs which take place in the research of IDS . It involves growing a signature that indicates the intrusion. Learning Classifier System (LCS) is the related technique, in which binary rules that recognize intrusion patterns are evolved. Bayesian Network In Bayesian Network, graphical models being introduced. These graphical type of models are being defined by set of rules of transition, which are represented as the probabilistic interdependencies. In this model, a conditional type of probability table and the state of random variables are described in each node. A restrictive probability table use to determines the probabilities of node in state, specified a state of its parent. This technique can hold data which is incomplete . Fuzzy Logic, this is planned to handle the indistinct and vague data. To indicate an intrusion, a relationship between input and output variables is defined by creating different set of rules. It uses membership functions to examine the intensity of truthfulness. All the above approaches are further summarized in Table [2]

Table 1: Techniques of IDS

Techniques	Functions
Artificial Neural Networks (ANNs)	System is trained by inserting related input/output data. This training is used afterwards to recognize arbitrary patterns, given as an input to the system.
State Transitions Tables	Intrusion take place or not is being detected by comparing behavior of system with the intruder's state transition figure.
Genetic Algorithms (GAs)	Mimic natural reproduction system inside nature where following positive changes, only the fittest users inside a generation will be reproduced in the subsequent generation.
Bayesian Network	Graphical models are introduced and deal with incomplete data.
Fuzzy Logic	Handles vagueness and impression

A. Signature based method

This is the traditional method for intrusion detection. It requires extensive knowledge of signature of previously known attacks. In this process monitored events are matched with the signature to detect intrusion. Data is mined from various dissimilar audit database and also comparing these features with to set of attack signature given by human specialist for the intrusion detection

1) Most Popular approaches

a) Anomaly Detection use to detect the irregular behaviors of the host or the network. It stores the features of user's usual behaviors hooked on database, and then it compares user's present behavior with database. The deviation of traffic which is monitored from normal profile is being measured.

b) Misuse Detection this works by penetrating for traces or the patterns of recognized attacks. There are two steps to be followed: Step one: Define abnormal system behavior: Step Two: Define any other behavior, as normal behavior. Deviations from such rules will point out an attack on network

2) In recent times emerged methods from the Machine learning

a) Supervised Learning Based techniques: Can become aware of attack which are known and pattern recognition have been utilized that to detect the intrusions.

b) Unsupervised Learning Based techniques : Can identify the intrusions that not been learned earlier .
Classification of IDS Based on Sources of audit information IDS divided into different types

c) Host Based IDS: Audit data held on individual computer that serve as hosts. Intrusion detection takes place on a single host system.

d) Network Based IDS: Network traffic considered as audit data source. Used to provide normal computing services and detect attacks from network.

e) Distributed IDS: Gather audit data from various hosts that connected by the network. Used to detect the attacks concerning multiple hosts

f) Hybrid intrusion Detection: It is a grouping of both host-based and network-based IDS. It gives flexibility and to increases level of security.

3) Major Problems with Current IDS

a) Data overload: Data source/Audit data which needs to be analyzed for intrusion detection must be discrete volume/size for efficient and effective analyze. Data overload is major problems of the Current IDS

b) False Positives: When IDS treats normal attack as malicious then it considered as false positives.

c) False Negatives: When the IDS does not produce an alert or alarm when intrusion are taking place in reality then it is well thought-out as False Negatives [2].

III. MULTI LEVEL SUPPORT VECTOR MACHINE

MLSVM Classifier for real time data is constructed using real time SNMP-MIB data which has 7 attributes. From the collected 22 attributes, 7 attributes are selected using the Correlation Based Feature Selection (CFS) Mechanism. The first level SVM classifier is constructed for the real time data which separates the Normal and Attack classes. The second level SVM classifier is constructed which classifies the attack into various classes of Network, Transport and Application Layer attacks. MLSVM Classifier for misuse detection is constructed using the kddcup 99 data which first classifies the data into normal and attacks and then into various classes of attacks. MLSVM Classifier for masquerader detection is constructed from the Shonlau's truncated command sequences which classifies the data into masquerader and non-masquerader first and then into various categories of the masqueraders [3].

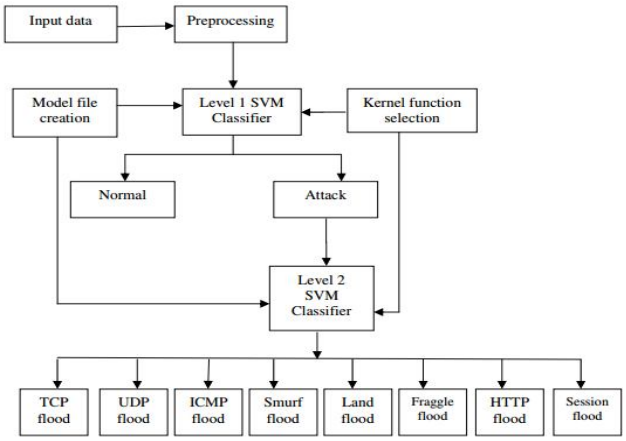

Fig.1 MLSVM Classifier

Table: 2 show the comparison of some intrusion detection using support vector machine technique. From the table, it is found that Ahmad et al. Achieved a better detection rate 99.6% and false alarm 0.4% compared to others. Also, S. Ganapathy et al. attained 98.51% detection rate which is slightly down when compared to Ahmad et al. approach [4].

Table 2: Comparison of intrusion detection using support vector machine

S.No.	Authors	Method/Algorithms	Detection rate (%)	False alarm rate
1	Carlo A. Catania et al.	Support vector machine	87.02%	-
2	Ahmad et al.	Support vector machine	99.6%	0.4%
3	S. Ganapathy et al.	Support vector machine	98.51%	-

IV. K-MEANS BASED INTRUSION DETECTION

In this section, we will discuss the dissimilar papers about the k-means algorithm. In 2003 to 2004 a number of papers obtainable to represent K-means algo based on the intrusion detection. Some papers be discussed below. In the year 2003, a K-means based clustering algorithm, named Y-means, for incursion detection has been offered. Y-means surmount 2 failings of the K means: quantity of clusters dependency and decadence An appropriate no of clusters were being divided normally. This was one of the benefits of the Y-means algorithm for intrusion detection. The log data which is unprocessed of the information systems can openly be applied as the data training with out being labeled actually was an additional benefit. To improve learning capacity and decrease the strength computationally of competitive learning of multilayered neural network. Through back propagation learning way the recommended model is needed multilayered network of structural design. The acquired results showed that the suggested technique executes specially in terms of both precision and computation time when pertained to KDD99 dataset match up to a normal learning schema that utilized the full dataset. To decrease the amount of examples to be offered to the neural network, the K-means algorithm was initially used to the training dataset by automatically choosing a most favorable set of samples [5].

V. COMPARATIVE STUDY ON THE APPLIED DATA MINING BASED ON INTRUSION DETECTION APPROACHES [3]

Table 3 Comparative study on applied data mining based intrusion detection techniques

YEAR	PAPER NAME	TECHNIQUES	MERITS	DEMERITS
Dec 2012	A survey on intrusion detection using data mining techniques	1.Association rule of dependency mining 2.Classification &clustering	Used in transaction data applied for KDD task unsupervised technique	……
	A review on data mining based intrusion detection techniques	1.Novel IDS 2.K-means clustering Algorithm 3.Data dependency weighted sequence mining 4.Hybrid IDS KDD, Anomaly detection	Used to detect DOS attack detect black hole attack used to filter out extra rules generated by this approach combines the filter and wrapper models for selecting relevant features investigate more efficient methods against intrusions	Not reliable doesn't provide sufficient mining method architecture needs to be enhance for cryptographic mechanism. Required to survey more recent techniques
	A survey on data mining techniques of intrusion detection	1.Feature choice 2.Machine education 3.Hybrid approach	Used on finite data set improve automatically through experience	
	A survey on intrusion detection System in data mining	1.Data mining, Feature Selection, Multiboosting 2.K-means clustering distributed IDS	Find high detection rates for U2R and R2L and also to detect attacks False alarm rate has been decreased also clustering helps in to identify the data which is attacked.	
	A survey Network IDS based on the data mining approach	1.Support vector machine 2.Genetic Algorithm(GA) 3.KNN 4.Neural Network 5.Bayesian Method	High accuracy solves optimization problem simple and highly adaptive behavior implicitly detect the complex nonlinear relationship between dependent and independent variables simplifies the computations exhibit when applied to large database	More time space complexity no global optimum high storage requirement it requires long training time the assumption made in class conditional independence lack of available probability data

VI. LITERATURE SURVEY

Yanjie Zhao et al. [2016] The paper mainly does the following work, by analyzing the intrusion deeply, extract the properties which can reflect intrusion characteristics effectively; merge exploitation detection, the anomaly detection and the human interference, set up rule which is library based on C.45 decision tree algo and to formulate the finest pattern matching to develop the speed of detection; the hosts are being clustered to be the IP group which is based on visit number by k means clustering algo, the audit type of data are being divided into parts beneath the IP group's direction, and classifiers are made up by the separated audit data respectively, then detected Data relate different rules according to its individual IP group, thus reduce the false positives. The experiments proved that the method is effective to detect intrusion such as scanning and Deny of Service [7].

Hossein Gharaee et al. [2016] this paper has proposed an anomaly based IDS using Genetic algorithm and Support Vector Machine (SVM) with a new feature selection of method. The new model has used a feature selection method based on the Genetic with an innovation in fitness function decrease the dimension of the data, increase true positive detection and simultaneously decrease false positive detection. In addition, the calculation time for training will also have a remarkable reduction. Results illustrate that the method which is proposed can reach high accuracy and low false positive rate (FPR) simultaneously, though it had earlier been achieved in earlier studies separately. This study proposes a method which can accomplish more stable features in comparison with other techniques. The proposed model experiment and test on KDD CUP 99 and UNSW-NB15 datasets. Numeric Results and comparison to other models have been presented [8].

Hamid Reza Ghorbanii et. al. [2015] the proposed technique classifies the dissimilar security needs, based on the CIA triad model, that into groups of individuals with same safety necessities and then select suitable policy. By grouping related users requirements or security requirements and tuning each type of IDS accordingly, proposed technique has been capable to get better IDS efficiency. Results of our simulations show that approach which is proposed will decrease the total detection time by 21% in average while preserving adequate detection coverage. Improving efficiency of IDS imply that it also processes a bigger volume of data due to reduction in time, better use of resources and also loads balancing between the groups [9].

Hachmi Fatma et. al. [2015] here , a two stage method is based on data mining and the optimization is proposed having as the input the result of several IDSs. In the 1st stage, for every IDS set of basic alerts is clustered to generate set of the meta alerts. Then, we eradicate false positives from sets of the meta alerts via the difficulty of binary optimization. In the 2nd stage, we throw away the meta-alerts generated by all of the IDSs and only individuals missed by one, two or the most them are absent. This set is known as the set of possible false negatives. In fact, at this stage meta alerts combination is executed to keep away from the idleness between the meta-alerts collected from the numerous IDSs. At last, an algo of binary classification is proposed to categorize the potential false negatives either as the real attacks or not. Experimental type of results illustrate that our process which is proposed is outperforms simultaneous methods by considerably dropping the rate of the false positives and the false negatives [10].

Chordia Anita S.Sunil Gupta et al. [2015] here we 1st apply the combination on KDD'99 dataset then it can be classified into the 4 category as U2R, the R2L, DoS and the Probe. The significant objective here is to decrease the false positive rate of IDS and attempt to improve its efficiency [11].

Shengyi Pan et al. [2015] This presents a systematic and the automated approach to construct a hybrid IDS that use to learns sequential state based condition for the power system type of scenarios together with disturbances, the normal control operations, and the cyber attacks. A technique of the data mining known for the common path mining that is used to automatically learn the patterns and also learn patterns accurately for scenarios from the grouping of synchrophasor measurement of data , and also the audit logs of power system. As a theory proof, an IDS prototype was implemented and also validated. The IDS model precisely classifies trouble, usual managing operations, and the cyber attacks for distance guard scheme for a 2 line 3 bus power transmission system [12].

Jamal Esmaily et al. [2015] here , a technique is based on grouping of Decision Tree (DT) algo and the Multi-Layer Perceptron (MLP) ANN be proposed which is capable to recognize the attacks with the elevated accuracy and consistency. The development of the internet type of attacks is a main difficulty for today's networks of the computer . Hence, security technique implementation is being done to avert such type of attacks is vital for any type of computer network. With the support of Machine Learning and the approaches of Data Mining, IDS which stands for Intrusion Detection Systems are able to analyze the attacks and system anomalies efficiently more . Though, most of the methods which is being studied here in this field, together with Rule-based expert systems, are not able to successfully identify the attacks which have different patterns from expected ones. By using Artificial Neural Networks which is also called(ANNs), it is likely to identify the attacks and classify the data, even when the dataset is nonlinear, limited, or incomplete [13].

Manjiri V. Kotpalliwar, et al. [2015] it has proposed the usage of SVM (Support Vector Machine) for categorization of the attack from huge amount of raw intrusion detection datasets on standard personal computers. SVM is a technique which is worn in data mining to mine data which is predicted. We have make use of KDDCUP'99 IDS type of database for the categorization [14].

VII. CONCLUSION

An intrusion detection system (IDS) monitors networked gadgets and appears for anomalous or malicious behavior inside the patterns of interest within the audit circulation. A comprehensive IDS requires a widespread amount of human know-how and time for development. Intrusion Detection System provides following main process: monitor, detect and produce an alert IDS are some time measured as the functionality of firewall. There is a bit variation between two of them. A firewall may be as a barrier that protects the information flow and avoid intrusions where as IDS detects if the network is under attack or if the security compulsory by the firewall has been infracted. SVM is a technique which is worn in data mining to mine data which is predicted. We have make use of KDDCUP'99 IDS type of database for the categorization.

REFERENCES

[1] Arushi Bhadouria, Er. Pawan Patidar , Dr. M. K. Rawat "Survey Paper on "Knowledge Based Improved Intrusion Detection System by Means of Information Gain" International Journal of Research in Engineering Technology and Management ISSN 2347 – 7539.

[2] Hussain Ahmad Madni Uppal , Memoona Javed and M.J. Arshad "An Overview of Intrusion Detection System (IDS) along with its Commonly Used Techniques and Classifications" International Journal of Computer Science and Telecommunications Volume 5, Issue 2, February 2014.

[3] Anthony Raj.A "A Study on Data Mining Based Intrusion Detection System" international Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 1 (March 2014).

[4] Yogita B. Bhavsar , Kalyani C.Waghmare "Intrusion Detection System Using Data Mining Technique: Support Vector Machine" International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 3, March 2013).

[5] Sravan Kumar Jonnalagadda, Ravi Prakash Reddy I "A Literature Survey and Comprehensive Study of Intrusion Detection" International Journal of Computer Applications (0975 – 8887) Volume 81 – No 16, November 2013.

[6] B.SENTHILNAYAKI1, J.KAVYA1 AND DR.K.VENKATALAKSHMI "A SURVEY OF INTRUSION DETECTION SYSTEM USING CLASSIFICATION" International Journal of Technology and Engineering System (IJTES) Vol 7. No.3 2015 Pp. 213-218

[7] Yanjie Zhao "Network Intrusion Detection System Model Based on Data Mining" 978-1-5090-2239-7/16/$31.00 copyright 2016 IEEE SNPD 2016, May 30-June 1, 2016, Shanghai, China.

[8] Hossein Gharaee, Hamid Hosseinvand "A New Feature Selection IDS based on Genetic Algorithm and SVM" 2016 8th International Symposium on Telecommunications (lST'2016).

[9] Hamid Reza Ghorbani, Roya Salek Shahrezaie "Toward a Policy-based Distributed Intruison Detection System in Cloud Computing Using Data Mining Approaches" Second International Congress on Technology, Communication and Knowledge (ICTCK 2015) November, 11-12, 2015 - Mashhad Branch, Islamic Azad University, Mashhad, Iran, 978-1-4673-9762-9/15/$31.00 ©2015 IEEE.

[10] Hachmi Fatma, Mohamed Limam "A two-stage process based on data mining and optimization to identify false positives and false negatives generated by intrusion detection systems" 2015 11th International Conference on Computational Intelligence and Security, 978-1-4673-8660-9/15 $31.00 © 2015 IEEE DOI 10.1109/CIS.2015.82.

[11] Chordia Anita S., Sunil Gupta "An Effective Model for anomaly IDS to Improve the Efficiency" 978-1-4673-7910-6/15/$31.00 ©20 15 IEEE.

[12] Shengyi Pan, Thomas Morris, and Uttam Adhikari, "Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems" 1949-3053 c 2015 IEEE.

[13] Jamal Esmaily, Reza Moradinezhad , Jamal Ghasemi "Intrusion Detection System Based on Multi-Layer Perceptron Neural Networks and Decision Tree" IKT2015 7th International Conference on Information and Knowledge Technology, 978-1-4673-7485-9/15/$31.00 ©2015 IEEE.

[14] Manjiri V. Kotpalliwar, Rakhi Wajgi" Classification of Attacks Using Support Vector Machine (SVM) on KDDCUP'99 IDS Database" 2015 Fifth International Conference on Communication Systems and Network Technologies, 978-1-4799-1797-6/15 $31.00 © 2015 IEEE DOI 10.1109/CSNT.2015.185.