Journal Of Harmonized Research (JOHR)

Journal Of Harmonized Research in Engineering
8(2), 2020, 23-33

ISSN 2347 – 7393

Review Article

A LITERATURE REVIEW ON INTRUSION DETECTION SYSTEM USING KDDCUP'99 DATASET

Kusum Lata, Mr. Manoj Yadav and Mr. Kailash Patidar

Sri Satya Sai University of Technology and Medical Sciences, Sehore, India

Abstract: An Internet is extensively used technology for the data communication now-a-days but during the communication of data network can be influenced by rigorous type of security threats and security issues which may lead the loss of data or can corrupt the data. So, to preserve the integrity or confidentiality of personal data or information a system is developed is known an intrusion detection system which collect and examine the various areas from which network can get trap. For the detection of intrusion different methodologies has been developed. In this paper, literature of some of the techniques to identify and thwart from the intrusion and their detection techniques is discussing using KDDCUP99 dataset and also presents its merits and demerits.

Keywords: - IDS, Security Threats, Integrity, KDDCUP99

Introduction*:* As the use of internet technology grows for the data communication, the network can compromised from different attack or threats. The information or network safety is becoming significant issue for any organization to preserve data and information in their computer network beside different types of attack with the help of resourceful and robust Intrusion Detection System (IDS). IDS can be developed using various machine learning techniques. IDS act as a classifier which classifies the data as normal or attack. Classification is a process of putting different categories of data together. Classification is one of the very common applications of the data mining in which similar type of samples are grouped together in supervised manner. An intrusion detection system can be classified into two categories [1]: network based and host based. The network attack also is of two types such as anomaly and misuse attack. The network-based attacks are detected from the interconnection of computer systems. Since the system communicates with each other, the attack is sent from one computer system to

For Correspondence:
kusummudgil08@gmail.com.
Received on: April 2020
Accepted after revision: April 2020
Downloaded from: www.johronline.com

another computer system by the way of routers and switches. The host based attacks are detected only from a single computer system and is easy to prevent the attacks. These attacks mainly occur from some external devices which are connected. The web based attacks are possible when systems are connected over the internet and the attacks can be spread into different systems through the email, chatting, downloading the materials etc. Nowadays many computer systems are affected from web based dangerous attacks. IDS are widely used area for the research and progress. This happens because detection of attack from the computer and network instead of IT security becomes major issue now-a-days. IDS efficiently and effectively detect the malicious activities on the network but the majority of existing system faces variety of challenges such as low detection rate and high false alarm rate. These problems happen due to the superiority of attack and intended similarities to normal behavior. In this paper, for the detection intrusion use KDDCUP'99 dataset [2]. KDDCUP99 Data set is an intrusion related data with almost 50 lacks samples. Ten percent of this data is publically available in UCI repository site for the experimental purpose of the researcher's. This optimum size of data contains samples for all 22 classes. A higher sample size data will require more computational resources which are not possible with simple desktop computers. So relatively low sample size data of KDD99 (10% of KDD) is used in this research work as raw material for developing a model. This data set contains about 5 million records as TCP/IP connection with 41 features, some of which are qualitative while others are continuous. Twenty two samples are categorized into five broader categories along with normal as DoS, R2L, U2R and Prob. Figure 1 illustrate the process for intrusion detection system.

The KDD'99 dataset may also get affected by several type of attack such as user to roots, denial of service, remote to local and probe [4].

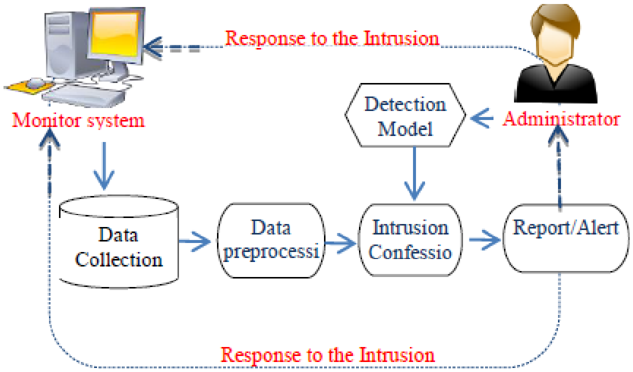


Fig.1 Intrusion detection system [18]

- *Denial of Service (dos):* Attacker tries to prevent legitimate users from using a service.
- *Remote to Local (r2l):* Attacker does not have an account on the victim machine, hence tries to gain access.
- *User to Root (u2r):* Attacker has local access to the victim machine and tries to gain super user privileges.
- *Probe:* Attacker tries to gain information about the target host.

The KDD 99 intrusion detection benchmark consists of three components, which are detailed in Table 1. In the International Knowledge Discovery and Data Mining Tools Competition, only "10% KDD" dataset is employed for the purpose of training [3]. This dataset contains 22 attack types and is a more concise version of the "Whole KDD" dataset. It contains more examples of attacks than normal connections and the attack types are not represented equally. Because of their nature, denial of service attacks account for the majority of the dataset. On the other hand the "Corrected KDD" dataset provides a dataset with different statistical distributions than either "10% KDD" or "Whole KDD" and contains 14 additional attacks.

Table 1 Fundamental features of KDD'99 intrusion detection dataset [4]

Dataset	DoS	Probe	u2r	r2l	Normal
"10%KDD"	391458	4107	52	1126	97277
"Corrected KDD"	229853	4166	70	16347	60593
"Whole KDD"	3883370	41102	52	1126	972780

This paper presents the literature of the proposed techniques for the detection of the some of the serious intrusion which provides destruction of the network.

The organization of the rest of the paper is as follows: next section contains literature of the previous works done after that various intrusion detection techniques are discussed with their advantages and demerits and finally conclusion of the presented paper in last section.

Related work: In this section discuss related work in current scenario intrusion detection technique using soft computing and data mining approach. In recent research trend soft computing and data mining play a vital role for intrusion detection. The role of data mining such as clustering classification and rule mining apply for detection of known and unknown type of attack, Instead of that soft computing implied in form of attribute and feature selection process in intrusion section system. Some work discuss here in current trend.

Heba F. Eid et al., [5] proposed intrusion detection system by using Principal Component Analysis (PCA) with Support Vector Machines (SVMs) as an approach to select the optimum feature subset. We verify the effectiveness and the feasibility of the proposed IDS system by several experiments on NSL-KDD dataset. A reduction process has been used to reduce the number of features in order to decrease the complexity of the system. The experimental results show that the proposed system is able to speed up the process of intrusion detection and to minimize the memory space and CPU time cost. *S. Revathi and A. Malathi [6]* proposed a new technique of combining swarm intelligence (Simplified Swarm Optimization) and data mining algorithm (Random Forest) for feature selection and reduction. SSO is used to find more appropriate set of attributes for classifying network intrusions, and Random Forest is used as a classifier. In the preprocessing step, we optimize the dimension of the dataset by the proposed SSO-RF approach and find an optimal set of features. SSO is an optimization method that has a strong global search capability and is used here for dimension optimization. The experimental result shows that the proposed approach performs better than the other approaches for the detection of all kinds of attacks present in the dataset. *Shafigh Parsazad et al. [7]* proposed a very simple and fast feature selection method to eliminate features with no helpful information on them. Result faster learning in process of redundant feature omission. We compared our proposed method with three most successful similarity-based feature selection algorithm including Correlation Coefficient, Least Square Regression Error and Maximal Information Compression Index. After that we used recommended features by each of these algorithms in two popular classifiers including: Bayes and KNN classifier to measure the quality of the recommendations. Experimental result shows that although the proposed method can’t outperform evaluated algorithms with high differences in accuracy, but in computational cost it has huge superiority over them. *Rajeswari et al., [8]* detection system that uses a combination of tree classifiers which uses Enhanced C4.5 which rely on labeled training data and an Enhanced Fast Heuristic Clustering Algorithm for mixed data (EFHCAM). The main advantage of this approach is that the system can be trained with unlabelled data and is capable of detecting previously “unseen” attacks. Verification tests have been carried out by using the 1999 KDD Cup data set. From this work, it is observed that significant improvement has been achieved from the viewpoint of both high intrusion detection rate and reasonably low false alarm rate. *Asim Das and S. Siva Sathya [9]* focused on the association rule mining in KDD intrusion dataset. Since the dataset constitutes different kinds of data like binary, discrete & continuous data, same technique cannot be applied to determine the association patterns. Hence, this paper uses varying techniques for each type of data. The proposed method is used to generate attack rules that will detect the attacks in network audit data using anomaly detection. Rules are formed depending upon various attack types. For binary data, A-priori approach is used to eliminate the non-frequent item set from the rules and for discrete and continuous value the proposed techniques are used. *LI Han [10]* used the unsupervised K-MEANS algorithm to model and detects anomaly activities. The aim is to improve the detection rate and decrease the false alarm rate. A K-MEANS algorithm based on information entropy (KMIE) is proposed to detect anomaly activities. KMIE can filter the outliers on the dataset to reduce the negative impact, and identify the initial cluster centers using entropy method. Then, KMIE can use these centers to iterative calculate and classify records into different clusters. This paper uses KDD CUP 1999 dataset to test the performance of KMIE algorithm. The results show that our method has a higher detection rate and a lower false alarm rate, it achieves expectant aim. *Devendra kailashiya and R.C. Jain [11]* presented the a method to improve accuracy Rate of intrusion detection using decision tree algorithm. Intrusion detection systems aim to identify attacks with a high detection rate and a low Error rate. In this paper we have supervised learning with preprocessing step for intrusion detection. We are using the stratified weighted sampling techniques to generate the samples from original dataset. These sampled applied on the proposed algorithm. The accuracy of proposed model is compared with existing results in order to verify the validity and accuracy of the proposed model. The results showed that the proposed approach gives better and robust representation of data. The experiments and evaluations of the proposed intrusion detection system are performed with the KDD Cup 99 dataset. The experimental results clearly show that the proposed system achieved higher Accuracy and Low Error in identifying whether the records are normal or attack one. *Yang Li and Li Guo [12]* proposed a novel supervised network intrusion detection method based on TCM-KNN (Transductive Confidence Machines for K-Nearest Neighbors) machine learning algorithm and active learning based training data selection method. It can effectively detect anomalies with high detection rate, low false positives under the circumstance of using much fewer selected data as well as selected features for training in comparison with the traditional supervised intrusion detection methods. A series of experimental results on the well-known KDD Cup 1999 data set demonstrate that the proposed method is more robust and effective than the state-of-the-art intrusion detection methods, as well as can be further optimized as discussed in this paper for real applications. *Tabash et al.[18]* the smart hybrid model was developed to explore any penetrations inside the network. The model divides into two basic stages. The first stage includes the Genetic Algorithm (GA) in selecting the characteristics with depends on a process of extracting, Discretize And dimensionality reduction through Proportional K-Interval Discretization (PKID) and Fisher Linear Discriminant Analysis (FLDA) on respectively. At the end of the first stage combining Naïve Bayes classifier (NB) and Decision Table (DT) using NSL-KDD data set divided into two separate groups for training and testing. The second stage completely depends on the first stage outputs (predicted class) and reclassified with multilayer perceptrons using Deep Learning4J (DL) and the use of algorithm Stochastic Gradient Descent (SGD). In order to improve the performance in terms of the accuracy in classification of penetrations, raising the average of discovering and reducing the false alarms. The comparison of the proposed model and conventional models show the superiority of the proposed model and the previous conventional hybrid models. The result of the proposed model is 99.9325 of classification accuracy, the rate of detection is 99.9738 and 0.00093 of false alarms.

*Mugabo and Zhang [19]*proposed method, the SVM classifier is adopted to classify network data into normal and attack behaviors, and due to the irrelevant and redundant features found in KDD datasets, IG is used to select the relevant features and remove unnecessary features. The KDD'99 and NSL-KDD datasets are used to evaluate the effectiveness of the proposed method. Compared with other methods, the experimental results show that the proposed method can detect malicious attacks with high accuracy, true positive rate, low false positive rate and high training speed.

Intrusion detection system: The National Institute of Standards and technology classifies[2] Intrusion Detection as "The process of monitoring the events occurring in a computer system, or network and analyzing them for signs of intrusions, defined as attempts to compromise the confidentiality, Integrity, availability or to bypass the security mechanism of a computer or network". An IDS is a system that attempts to identify intrusions. Which we done to be unauthorized uses, abuses or misuses of computer systems by either authorized users or external perpetrators. Intrusion Detection provides the following:

1. Monitoring and analyzing both user and system activities
2. Analyzing system configurations and vulnerabilities
3. Accessing system and file integrity
4. Ability to recognize patterns typical of attacks
5. Analysis of abnormal activity pattern
6. Tracking user policy violation

Intrusion detection systems can be classified based data collection methods into two categories as Host based and Network-based. A network-based intrusion detection system (NIDS) is used to monitor and analyze data from network traffic to protect a system from network-based attacks. A Host-based intrusion detection system (HIDS) monitors and analyzes data from system's log files that runs on a particular system. Intrusion detection systems can also be classified based on intrusion detection techniques into three categories as misuse- detection, specification-based detection and anomaly-based detection.

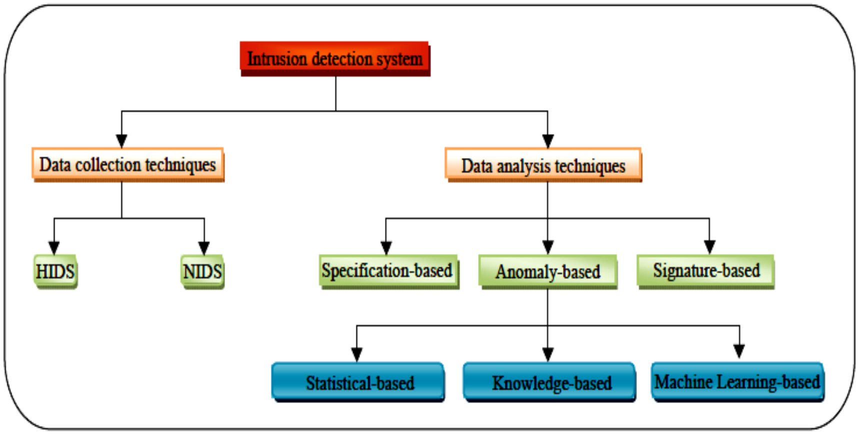


Fig.2 Classification of intrusion detection system

Signature based detection: A Signature based Intrusion Detection Systems references a stored collection of previous attack signatures such as specific patterns, known malicious instruction sequences, byte sequences in network traffic and known system vulnerabilities. Each intrusion gives a specific malicious signatures such as failed logins, failed attempt to run an application, failed file and folder access and nature of data packets. Signature based intrusion detection system uses these signatures to detect and prevent the same attacks in the future. The main advantage of signature-based intrusion detection system is that it is very easy to develop and understand if we know the behaviour of network traffic and system activity. For example, to exploit particular buffer-overflow vulnerability the signature based intrusion detection system uses a signature that looks for particular strings. On modern systems pattern matching can be more efficiently communicate via DNS, ICMP and SMTP it can enable the specific signatures and disable all other signatures. The main disadvantages of signature-based intrusion detection systems are the collection of signatures must be continually updated and maintained and signature-based intrusion detection systems may fail to identify unique attacks. Signature-based intrusion detection systems work well against attacks with fixed behavioral pattern, but it is hard to works with self-modifying behavioural characteristics. Intrusion detection is further difficult when the user uses advancing exploit technologies such as payload encoders, encrypted data channels and nop generators that permit malicious users. To works against these kinds of attacks the collection of signatures must be continually updated and maintained which decreases the efficiency of the signature based systems also reduces the performance of the system. To address this issue modern systems which uses signature based intrusion detection system uses many IDS engines with multi processors and multi Gigabit network cards. The efficiency of the system determine by the speed of creation of the new signatures between the developers and attackers.

Anomaly based detection

Anomaly based Intrusion Detection Systems (IDS) references a baseline or learned pattern of normal system activity to identify active intrusion attempts. Deviations from this baseline or pattern cause an alarm to be triggered. Events in an anomaly detection engine are caused by any behaviors that fall outside the predefined or accepted model of behavior. The major drawback of anomaly detection system is the difficultly of defining rules. All protocol being analysed must be well defined, implemented and tested for accuracy. The rule defined process for various protocols is also compounded by differences in vendor implementations. Defining rule in customized protocols needs great efforts. Detailed information of normal network behavior needs to be collected and maintained by system memory for detection to occur correctly. Once the behavior is defined and rules for the protocol have been well structured and built the system can scale more quickly and easily than the signature-based model and works well for anomaly detection. There is a chance for malicious behavior gets unnoticed if it is considered as a normal usage patterns. For example a directory traversal activity with server, which complies with network protocol, does not trigger any payload or bandwidth limitation flags or any other flags. However, anomaly detection has an advantage over signature-based systems to detect new automated worms, in that a new attack for which a signature does not exist can be detected if it falls out of the normal traffic patterns. When a new system is infected with a worm it usually starts scanning for other vulnerable systems at an abnormal rate flooding the network with malicious traffic, thus triggering a network bandwidth abnormality rule. If any abnormal behavior or intrusive activity occurs in the computer system which deviates from system normal behaviour then an alarm is generated. So this have follows a continuous monitoring process. The key advantage of anomaly detection is that it does not necessitate preceding information's or data of intrusion, so it can thus detect new intrusions. Based on behavior model processing type of the system anomaly based detection techniques can be classified in to three groups as statistical-based, knowledge-based and machine learning-based.

Intrusion detection techniques

Decision Tree: Decision trees are a popular structure for supervised learning. Its construction process is top down, divide and conquer, and also a greedy algorithm. The basic ID3 algorithm works well for limited number of records in data set and it cannot handle missing values and also when the data set size is increased the tree is not accustomed to the changes [13]. One of the greatest advantages of decision tree classification algorithm is that: It does not require user to know a lot of background knowledge in the learning process [14]. The ID3 algorithm uses the Information Gain to select a splitting attribute and then construct the tree. There are some other algorithms that consider the gini index and gain ratio to select the splitting criterion and attribute. For ID3 decision tree, concept used to quantify information is called entropy. Entropy is used to measure the amount of uncertainty in a set of data. When all data in a set belongs to a single class the entropy is zero that is there is no Uncertainty [14].

The following Step is done recursively [16]:

1) Computing the Information Gain for each Attribute.

2) The attribute with the highest information gain, is selected as a splitting attribute.

3) If the selected attribute is discrete (categorical), the node is branched with all possible values. If the attribute is continuous, a cut point with the highest Information gain is selected.

4) After splitting, consider whether or not these new nodes are leaves; otherwise, new nodes are the root of the sub tree.

5) Repeat Step 1 to 4.

Given the probabilities P1, P2, . . . , Pn

n , Where Σ Pi=1,

i=1

Entropy I is calculated as

n

I (P1, P2…Pn) = Σ (Pi log (1/Pi) (1)

i=1

Information Gain (D,G) = I (D)- Σp(Di)I(Di) (2)

Where D= Select attribute

Graph-Base Cluster Algorithm: Graph-based clustering algorithm [17] is a method commonly used in automatically partition for a data set in several clusters. It proceeds by setting a parameter of clustering precision to control the result of clustering. Records in dataset are packaged as a note. These notes are treated as vertex of a complete undirected graph, and the distance values between these notes as weight of the edge. The distance is calculated by Euclidean distance function. According these values of distance, we could construct a distance matrix I. And the threshold δ is computed by a parameter of cluster precision α.

$$\delta = dis_{min} + (dis_{max} - dis_{min}) \times \text{Cluster Precision} \quad (1)$$

dis_{min} and dis_{max} represent the minimum and maximal value of matrix I respectively. So an edge is cut down from this graph if its value of weight greater than threshold δ in Fig.3.

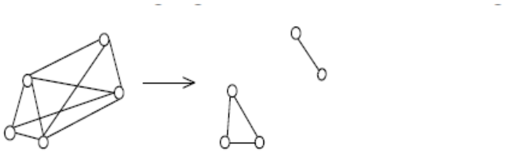
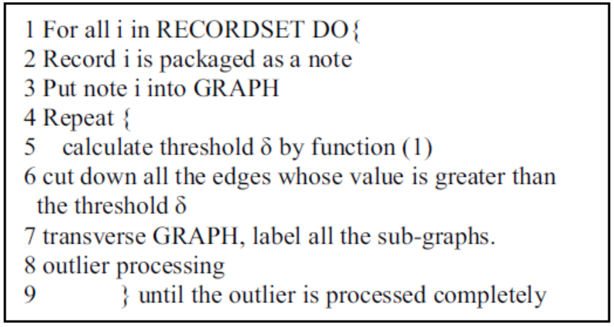


Fig.3 GB Cluster

Finally, transverse the whole graph, the notes would be classified into the same cluster if there is edge between them. Therefore, several sub-graphs are created. Each sub-graph represents a cluster. Finally, outlier is processing. The steps of this algorithm is showed in Fig.4



```
For all i in RECORDSET DO{
Record i is packaged as a note
Put note i into GRAPH
Repeat {
  calculate threshold δ by function (1)
cut down all the edges whose value is greater than
 the threshold δ
transverse GRAPH, label all the sub-graphs.
outlier processing
        } until the outlier is processed completely
```

Fig. 4 GB clustering algorithm

GB algorithm has been used for clustering for decades. However, it mainly has two shortcomings when it is applied for intrusion detection: the first one is that it distinguishes the normal and abnormal cluster just by a value of threshold. So the clustering accuracy is far from enough. Second, it doesn't offer a reasonable method to address outliers, but just simply throw it away. With this coarse granularity partition, it can't receive a satisfied detection rate. On the other hand, the ability to detect any shape of cluster is made it very suitable for the dataset with complex shape in real network.

Machine learning based detection: Machine learning is a method of data analysis, in which the system learns and gathers knowledge from the tasks performed by the system. The system will improve the performance by using the knowledge learned from the previous results it means that machine learning provides the ability to a system to enhance the execution strategy [21]. The systems with machine learning techniques can be used for various applications but these types of systems are expensive. In many application context the machine learning technique uses methods similar to that of the statistical techniques and data mining techniques [22]. Machine learning technique can be classified into Neural networks, Fuzzy logic approach and Support vector machines. Neural Networks machine learning technique use neural network concepts acquire the ability to use the sequence of commands by the user to anticipate for the next command. The neural network model is well suitable for developing user behavior model since it does not require the explicit information on user behavior. A well trained neural network with back propagation and feed forward mechanism works efficiently as signature matching system [23].Multilayer Perceptron's, Radial Basis Function- Based neural networks are used for anomaly based intrusion detection systems. IDS using neural network consists of three phases [24].In first phase the audit log is analyzed to obtain sufficient training data. Next phase is to train the neural network for understand the each user behavior. In the final phase each user behavior is compared with trained data to detect malicious behavior of the user if there exist any such user activities an anomaly is alarmed.

Bayesian Inference: A Bayesian network consists of nodes and arcs representing variables and relations between the variables. Anomaly detection using Bayesian networks is a three-step process. The first step is to select the variables used to monitor the system. There is no restriction on the kind of the variables. The second-step is to evaluate the relationship between these variables to construct a Bayesian network, so this step is the learning phase of the algorithm and it is called the "profile" of the system. The third step to determine the "support" by using profile, it gives the current state of the variables describing the system. The support is the probability of occurrence of the states observed. If this probability is less than the thresh hold, an alarm can be raised. In Bayesian networks, computing the "support" can be made by using mathematical formulas and the probability distributions computed for the profile.

Fuzzy Logic: Fuzzy Logic means to the model of uncertainty of natural language. In this case the logic depends on linguistics by taking the minimum of set of events or maximum instead of stating OR, AND or NOT operation in the if-then-else condition. Basically, intrusion detection systems distinguish between two distinct types of behaviors, normal and abnormal. Fuzzy logic could create sets that have in-between values where the differences between the two sets are not well defined.

Support Vector Machine: SVM is developed on the principle of structural risk minimization. It is one of the learning machines that map the training patterns into the high-dimensional feature space through some nonlinear mapping. SVM has been successively applied to many applications in the multiclass classification [25, 26]. By computing the hyper plain of a given set of training samples, a support vector machine builds up a mechanism to predict which category a new sample falls into (Figure 1).

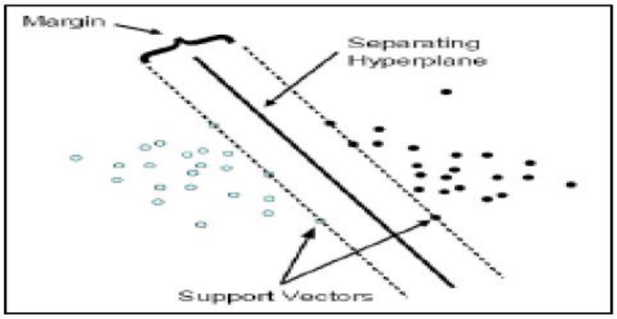


Fig.5 Separating Hyperplane with SVM

In an SVM, a data point is viewed as a vector in the d-dimensional feature space. Assume that all data points belong to either class A or class B. Each training data point can be labeled by based on (1):

$$Yi = \begin{matrix} -1\ xi \in class\ A \\ 1\ xi \in class\ B \end{matrix} \quad (1)$$

Therefore as it is revealed in Fig. 2, the training data set

can be designated as:

$$D = xi\ yi\ 1,2,3 \ldots \ldots \} \quad (2)$$

Data points with label 1 and −1 are referred to as positive and negative points, respectively. In the linear separable case, there are many hyper-planes which might separate the positive from the negative points. The algorithm merely looks for the major margin separating hyper-plane where the "margin" of a separating hyper-plane is defined to be the sum of the distances from the hyper-plane to the closest positive and negative points. In order to calculate the margin of a separating hyper-plane H, consider the hyper-planes H1 and H2 that include the closest positive training points and the closest negative training points to H, respectively:

$$\begin{aligned} &H: w.x - b = 0,\ x\ R^d \\ &H1: w.x - b = 1\ x\ R^d \\ &H2: w.x - b = -1\ x\ R^d \end{aligned} \tag{3}$$

Where w is the normal to H and b is the distance from H to the origin. Obviously, H2, H1, and are parallel. In addition,

$$\begin{aligned} &w.xi - b \geq 1 \text{ for } yi = 1 \\ &w.xi - b \leq -1 \text{ for } yi = 1 \end{aligned} \tag{4}$$

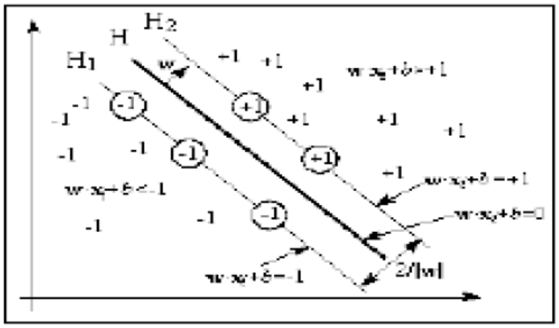


Fig.6 Data Points and Their Classes

Table 2: Advantages and disadvantages of SVM

ADVANTAGE	DISADVANTAGE
1. Speed of SVM, as the capability of detecting intrusion in real time.	1. SVM can only handle binary-class classification, whereas intrusion detection requires multi-class classification.
2. SVM can learn a larger set of pattern.	2. If the number of features is much greater than the number of samples, the method is likely to give poor performances.
3. SVM also has ability to update the training pattern dynamically.	3. The biggest limitation of the support vector approach lies in the choice of the kernel.

Principal Component Analysis: Typical datasets for intrusion detection are typically very large and multidimensional. With the growth of high speed networks and distributed network based data intensive applications storing, processing, transmitting, visualizing and understanding the data is becoming more complex and expensive. To tackle the problem of high dimensional datasets, researchers have developed a dimensionality reduction technique known as Principal Component Analysis (PCA). [27] In mathematical terms, PCA is a technique where n correlated random variables are transformed into d ≤ n uncorrelated variables. The uncorrelated variables are linear combinations of the original variables and can be used to express the data in a reduced form. Typically, the first principal component of the transformation is the linear combination of the original variables with the largest variance. In other words, the first principal component is the projection on the direction in which the variance of the projection is maximized. The second principal component is the linear combination of the original variables with the second largest variance and orthogonal to the first principal component, and so on. In many data sets, the first several principal components contribute most of the variance in the original data set, so that the rest can be disregarded with minimal loss of the variance for dimension reduction of the dataset. PCA has been widely used in the domain of image compression, pattern recognition and intrusion detection. They measured the distance of each observation from the center of the data for anomaly detection. The distance is computed based on the sum of squares of the standardized principal component scores.

Conclusion: Intrusion detection over the network is a serious issue which degrades the performance of the network and also corrupts the information travel over it. To detect / thwart the intruders or threats different technology and algorithm has been developed by various researchers and it is analyzed that some generates accurate results and reduces the false alarm rate. In this paper, we present review of the literature and also presented the types of intrusion detection system with their detection technique. The ensemble solution of network based and host based HIDS can also be used in various application domains. It is observed that anomaly detection domain has several promising research directions, numerous anomaly detection methods necessitates huge amount of test date set for detecting anomalies. Foremost directions in the direction of the researches in anomaly detection are to progress proficient anomaly detection systems which work with complex systems (e.g. airplane system, railways system) and interaction among several components in real time.

Reference

[1]. S. Devaraju and S. Ramakrishnan "Performance Comparison For Intrusion Detection System Using Neural Network With KDD Dataset" ICTACT Journal On Soft Computing, April 2014, Volume: 04, Issue: 03743 ISSN: 2229-6956(Online).

[2]. Pratibha Soni, Prabhakar Sharma "An Intrusion Detection System Based on KDD-99 Data using Data Mining Techniques and Feature Selection", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-4 Issue-3, July 2014.

[3]. S. Hettich, S.D. Bay, "The UCI KDD Archive" Irvine, CA: University of California, Department of Information and Computer Science, http://kdd.ics.uci.edu, 1999.

[4]. [4] H. Günes Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets".

[5]. Heba F. Eid, Ashraf Darwish, Aboul Ella Hassanien and Ajith Abraham "Principle Components Analysis and Support Vector Machine based Intrusion Detection System", in proceeding of IEEE.

[6]. S. Revathi and A. Malathi "Network Intrusion Detection Using Hybrid Simplified Swarm Optimization and Random Forest Algorithm on NSL-KDD Dataset" International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 3 Issue 2 February, 2014 Page No. 3873-3876.

[7]. Shafigh Parsazad, Ehsan Saboori and Amin Allahyar "Fast Feature Reduction in Intrusion Detection Datasets" MIPRO 2012, May 21-25, 2012, Opatija, Croatia.

[8]. L.Prema Rajeswari, A. Kannan "An Intrusion Detection System Based on Multiple Level Hybrid Classifier using Enhanced C4.5" IEEE-International Conference on Signal processing, Communications and Networking Madras Institute of Technology, Anna University Chennai India, Jan 4-6, 2008. pp75-79.

[9]. Asim Das and S. Siva Sathya "Association Rule Mining for KDD Intrusion Detection Data Set" International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Volume-2, Issue-3, 2012.

[10]. LI Han "Research of K-MEANS Algorithm based on Information Entropy in Anomaly Detection" 2012 Fourth International Conference on Multimedia Information Networking and Security.

[11]. Devendra kailashiya and R.C. Jain "Improve Intrusion Detection Using Decision Tree with Sampling" International Journal of Computer Technology & Applications,Vol 3 (3), 1209-1216, ISSN:2229-6093

[12]. Yang Li and Li Guo "An active learning based TCM-KNN algorithm for supervised network intrusion detection" computers & security 2 6 (2007) 4 5 9 – 4 6 7 in proceeding of Elsevier.

[13]. Huang Ming, Niu Wenying, Liang Xu "An improved decision tree classification algorithm based on ID3 and the application in score analysis".

[14]. Juan Wang, Qiren Yang, Dasen Ren "An Intrusion Detection algorithm based on decision tree technology" 2009 Asia pacific conference on information processing.

[15]. T. Jyothirmayi, Suresh Reddy "An algorithm of Better decision tree" International journal of computer science and Engineering.

[16]. Mohammadreza Ektefa, Sara Memar, Fatimah sidi and Lilly Suriani Affendey "Intrusion Detection using Data Mining Technique"IEEE 2010.

[17]. Zhou Mingqiang, Huang Hui, Wang Qian "A Graph-based Clustering Algorithm for Anomaly Intrusion Detection" The 7th

International Conference on Computer Science & Education (ICCSE 2012) July 14-17, 2012. Melbourne, Australia.

[18]. Mohammed Tabash, Mohamed Abd Allah, and Bella Tawfik "Intrusion Detection Model Using Naive Bayes and Deep Learning Technique", The International Arab Journal of Information Technology, Vol. 17, No. 2, March 2020.

[19]. Emmanuel Mugabo and Qiu-Yu Zhang, "Intrusion Detection Method Based on Support Vector Machine and Information Gain for Mobile Cloud Computing", International Journal of Network Security, Vol.22, No.2, PP.231-241, Mar. 2020 (DOI: 10.6633/IJNS.202003 22(2).06).

[20]. Shijoe Jose, D.Malathi, Bharath Reddy, Dorathi Jayaseeli, "A Survey on Anomaly Based Host Intrusion Detection System", National Conference on Mathematical Techniques and its Applications (NCMTA 18) IOP Publishing IOP Conf. Series: Journal of Physics: Conf. Series 1000 (2018) 012049 doi :10.1088/1742-6596/1000/1/012049.

[21]. Animesh P.; Jung-Min P 2007 An overview of anomaly detection techniques: Existing solutions and latest technological *trends Elsevier, Science Direct, Computer Networks*, vol. 51, pp. 3448,3470

[22]. Garcia-Teodoro, Pedro, J. Diaz-Verdejo, Gabriel M and Enrique V 2009Anomaly-based network intrusion detection: Techniques, systems and challenges*computers & security*vol 28 pp 18, 28.

[23]. Sreenath.M, 2014 A Comprehensive Review on Intrusion Detection Systems, *CiiT International Journal of Networking and Communication Engineering* vol 6.

[24]. Biermann, Elmarie; Elsabe C., Lucas V 2001 A comparison of Intrusion Detection systems", Elsevier, Computers & Securityvol. 20, pp. 676, 683.

[25]. Vahid Golmah ―An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM‖ International Journal of Database Theory and Application Vol.7, No.2 (2014), pp.59-70

[26]. S.-W. Lin, K.-C. Ying, C.-Y. Lee and Z.-J. Lee, "An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection", Applied Soft Computing, vol. 12, (2012), pp. 3285-3290.

[27]. Adathakula Sree Deepthi, 2Dr. K.Venkata Rao, "Anomaly Detection Using Principal Component Analysis", International Journal of Computer Science And Technology, Vol. 5, Issue 4, Oct - Dec 2014 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print).