Review Paper on Intrusion Detection using Data Mining Methods

Manish Arya[1] Dr. Sanjiv Sharma[2]
[1]Student [2]Assistant Professor
[1,2]Department of Computer Engineering
[1,2]Madhav Institute of Technology and Science Gwalior M.P

Abstract— Data mining is developed being accepted as a feasible means of analyzing enormous data sets, it is no longer possible to manually find useful information in this data when use in commercial and scientific data set approaches the terabytes and petabytes ranges. IDS is the system which identifies malicious activity on the network. As the Internet volume is increasing rapidly, security against the real time attacks and their fast detection issues gain attention of many researchers. Approaches of data mining can be successfully applied to IDS to tackle dynamic data problems and to increase performance of IDS. In this paper we study about intrusion detection based data mining, GA, soft set approach, data warehouse and also some previous paper study.

Key words: Data Mining, GA, IDS

I. INTRODUCTION

With the tremendous development in the computer networks use day by day, network as well as information security becomes the prime important factor. The basic aim of security is to develop protective software system which can provide three basic security goals that are privacy, integrity and authentication. Growth in the computer and network of computer use today's civilization seeks extremely secured and trusted communication. There is a methods range being utilized in IDS, but any systems are not completely perfect. We refer intrusion as any set of events that try to negotiate the integrity, confidentiality, or availability of a computer resource. The finding procedure out the asymmetrical activities on network and system is known as IDS. Intrusion is any activity which tries to violate these security goals [1]. The IDS plays a key role in identifying such malicious activities. The term IDS was first presented through Anderson in 1980.

II. DATA MINING AND INSTRUCTION DETECTION

A. Data Mining Technology

Data Mining means extracting or mining the knowledge from the mass data. To be specific, it means processing data so as to gain the implied, prior unknown, potential and useful knowledge, which can be expressed as patterns. The targets of digging include not only data source and file system, but also any data collections such as Web source [5]. Data mining is a latest offered intrusion recognition methodology. Its benefit lies in the fact that it can withdraw the needed and unknown information and regularities from the massive network information and host log data. It is a novel attempt to use data mining in achieving network security, both at home and abroad.

At the moment, the research on data mining algorithm is quite mature, and data mining itself is a general knowledge discovering technique. In the field of intrusion detection, we consider it as a data analysis process, which applies certain data mining algorithm to massive safe data in order to construct system of intrusion detection with self-adaptability and scalability. At current, an algorithm of data mining functional to intrusion recognition mostly has four basic patterns: association, clustering, sequence and classification [2].

B. Intrusion Detection System

IDS have become a primary study area in Computer-based security. It's a well-known skill for enlightening and is exploited to defend data consistency and system accessibility throughout an intrusion. When a person tries to access structure of knowledge in the particular system or does any unlawful action, the action is known as an intrusion that further has two forms such as interior and exterior. The exterior are those people which have not access authority the system information and still they try to obtain illegitimately with the aid of different saturation techniques. While interior is that who have a legal permission to access system, but try to do illegal activities. Software bugs exploitation and miss configurations of the system cause intrusion. Sniffing unsecured traffic, password cracking, or exploiting the particular protocols design blame are also some of the ways that cause intrusion.

C. Genetic Algorithm (GA)

GA is a find heuristic which gives a suitable solution to find and optimization difficulties. GAs includes a technique to get optimal combinatorial state using interest parameters set. Genetic programming also helps in simulating the population evolution process. Genetic algorithm evolves the inhabitants of fixed length by applying mutation operators and crossover along with a fitness function. The output concludes how likely individuals are to reproduce. A group of rules is developed each of which is calculated approximately to get fitness. Rules having higher fitness value create a novel generation. Numerous generations go through the same procedure to produce a solution that is acceptable. GA is an adaptive heuristic find method depends on natural election idea [3]. They are motivated thru Darwin's evolution theory– "survival of the fittest", which is a randomized search technique.

D. Soft Set Approach

This is now a portion of soft computing. It is a new approach for data mining that deals with uncertain data. Experiments reveal which in few aspects the rough sets and soft sets are alike. Soft sets are mainly used to analyze enormous quantity of data and help in taking well informed decisions. Therefore soft sets can be exploiting in real global decision support schemes. Such systems can benefit management of an association to take decisions that can result in expected results or profits. This theory is an element of the soft data mining. Experiments as illustrate in literature indicate that soft set theory can be exploit in tandem with other some computing techniques in sequence to produce highly effective results. Soft sets can also be

exploiting to achieve reducts that will help in better performance. Soft sets are being exploited in several applications such as classification of medical data, taxonomy of musical instruments, evaluation of student marks, better grouping of data etc. In many applications soft sets can be exploiting to generate decision rules that can help to take well idea out decisions as the soft sets are capable of providing required business intelligence.

III. DATA MINING BASED INTRUSION DETECTION SYSTEM ARCHITECTURE

The overall system architecture is intended to support a data mining-based IDS with the properties described. The architecture is consists of sensors, detectors, a data warehouse, and a model generation component. This architecture is able to supporting not only data gathering, sharing, and analysis, but also data archiving and model generation and distribution. The system is designed to be free of the sensor data format and model representation. A piece of sensor data can contain an arbitrary number of features. Each feature can be continuous or discrete, numerical or symbolic.

A. Sensors

These observe raw data over monitored system and compute features for use in model evaluation. Sensors insulate the remaining IDS from the specific low level properties of the target system being monitored. This is completed by having the entire sensors implement a Basic Auditing Module (BAM) framework. In BAM, features are calculated from raw data and encrypted in XML.

B. Detectors

These receipts processed data from sensors and use a detection model to evaluate the data and determine if it is an attack. The detectors also send back the result to the data warehouse for further analysis and report. There can be several (or multiple layers of) detectors monitoring the same system. There may also be a "back-end" detector, which employs very sophisticated models for correlation or trend analysis, and several "front-end" detectors that perform quick and simple intrusion detection.

C. Data Warehouse

The data warehouse serves as a centralized storage for data and models. One benefit of a central repository for the data is that different components can manipulate the same piece of data asynchronously with the existence of a database, for example manually labeling and off-line training. The data warehouse also facilitates the integration of data from multiple sensors. By correlating data/results from various IDSs or data gathered over a long period of time, the detection of complicated and large scale attacks becomes possible.

D. Model Generator

The main purpose of the model generator is to facilitate the rapid development and distribution of new (or updated) intrusion detection models. In this, an attack perceived first as an anomaly may have its exemplary data processed by the model generator, which in turn, using the archived normal and intrusion data sets from the data warehouse, automatically generates a model which can observe the new intrusion and distributes it to the detectors. Especially useful are unsupervised anomaly algorithms of detection because they can operate on unlabeled data which can be directly collected by the sensors.

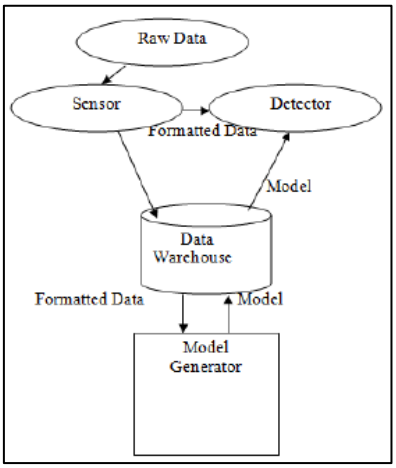

Fig. 1: The Architecture of Data Mining based IDS [4]

IV. APPLICATION OF DATA MINING IN INTRUSION DETECTION

In traditional intrusion detecting system, security experts firstly classify attacking actions and system weakness, choose statistical methods due to the detecting types, then manually enter the code and establish the corresponding detecting rules and modes. For complex network system, the limitation of experts' knowledge grows with the change of time and space, so it is not good to improve the effectiveness of detecting the intrusion detecting modes. Security experts usually concern about the known attacking features and system weakness and research on that, which causes the lack of adaptability of the detecting pattern to the unknown intrusion that the system is about to be facing. Meanwhile, the long upgrade cycle of security system, the high cost, these are not advantageous for improving the adaptability of intrusion detecting pattern.

As the experts' rules and statistical methods often require the support of software and hardware, it stops the system from reusing and developing in new environment, meanwhile it causes the difficulty of embedding new detecting modules. All of these are not good for improving scalability of intrusion detecting pattern. Therefore, it has become an important issue how to establish an effective, self-adaptable and scalable intrusion detecting pattern in intrusion detecting field. Considering intrusion detection as a data analysis procedure through applying the data mining predominance in its effective use of knowledge, this is a technique that can automatically create accurate and applicable intrusion patterns from massive audit data, which creates intrusion detecting system, can be applied to any computer environment. This approach has become a popular topic of research, in the field of inter discipline of network security and artificial intelligence. The analysis methods of association, sequence, clustering and classification in the mining of data has been proved possible [5].

Need of Data Mining within Intrusion Detection Data Mining is a procedure of extracting unseen, formerly unknown and valuable information from huge databases. It is a suitable manner of extracting designs and focuses on subjects associating to their feasibility, utility, scalability

and efficiency. Thus data mining scheme help to identify designs within data set and utilize these designs to perceive future interruptions in related data. The below are some detailed things which make utilize of data mining significant in intrusion detection system: i) Manage firewall rules for anomaly detection. ii) Analyze huge volumes of network data. iii) Same data mining tool may be applied to different data sources. iv) Performs data summarization and visualization. v) Differentiates data that may be utilized for deviation analysis. vi) Clusters the data into groups such that it possess high intra-class similarity and low inter-class similarity[6].

V. CHALLENGES OF INTRUSION DETECTION

There are many challenges to doing intrusion detection: • False positives: Administrators can be totally bogged down by false positives which are essentially warnings about things. • Learning curves: Intrusion detection can be a technically challenging environment that may require a substantial learning curve. • Large Logs: Logs of events are useless unless that are looked at via some mechanism. • Placement of IDS: Where do you place your IDS in order to effectively catch intrusion attempts [7].

VI. IDS US ING DATA MINING APPROACHES

A. IDS using Classification Analysis is

The objective is to assign objects (intrusions) to classes based on the values of the object's features. Classification algorithms may be utilized for both anomaly and misuse detections. In misuse detection, traffic data of network are gathered and labeled as "normal" or "intrusion". This labeled dataset is utilized as training data to study different classifiers. The SVM is very prominent algorithm in data mining field, but it has extensive training time problem. The SVM is very effective algorithm of classification in the data mining area. SVM uses high dimension space to discover a hyperplane to perform binary classification. SVM approach is a classification technique based on Statistical Learning Theory (SLT). It is based upon the idea of hyper plane classifier. The objective of SVM is to discover a linear optimal hyper plane. SVM will classify the intrusion when intrusion behaviors occur. A classification work involves training and testing sets that comprise of instances. Each instance contains one "target value" (class labels: Attack or Normal) and numerous "attributes" (features). The objective of SVM is to create a model that forecasts desired value of data instance within testing set which is given only attributes.

B. IDS US ING Clus Tering Analysis is

Clustering assign objects (intrusions) to groups (clusters) on the basis of distance measurements made on the objects. Classification and clustering are unsupervised learning procedure since no information is present on the labels of the training data. In anomaly detection, clustering and outlier analysis can be used to drive the ID model. Distance or match measure are very important in grouping clarifications in homogeneous clusters. It is significant to express a metric to define whether an event is deemed normal or anomalous using measures such as Jaccard similarity measure, Cosine similarity measure, Euclidian distance measure and longest common subsequence (LCS) measure. Jaccard similarity coefficient is an arithmetical measure of match between sample sets and may be explained as the commonality degree between two sets. Cosine similarity is mostly utilized in text databases and it computes the angle of difference in two vectors direction, regardless of their lengths. Euclidean distance is an extensively used distance measure for vector spaces, for two vectors X and Y in a dimensional Euclidean space; Euclidean distance can be explained as the square root of the total of differences of the corresponding dimensions of the vectors.

C. IDS US ING ANN

Artificial Neural Network (ANN) is relatively crude electronic models based on the neural structure of the brain. The brain basically learns from his experience. This is natural proof which has some difficulties that are beyond the range and scope of current computers is certainly solved by small energy efficient packages. The brain modeling of a technical way to develop machine solutions is the new arrival approach to computing also provides a more graceful degradation during system overload than its more habitual counter-parts. A neural network is an interrelated group of artificial neurons that employs arithmetic or computational model for information processing based on a connection approach to computation. It cannot comprises domain information at the starting, but it may be managed to create results by mapping instance pairs of input data into instance output vectors, and estimating its weights so that it maps each input example vector into the corresponding out-put example vector approx. [8] .

VII. TYPES OF IDS

There are several types of IDS; they are considered on the base of various analyses and monitoring approach. Another technique of categorizing IDS is to group them by data source. Some IDS study data sources produced by the Operating system or application software for symbols of intrusion. Other examines the network packet taken from Internet link to discover attackers. Protected IDS systems are Host and Network based system. Host based system watch a specific host machine. Network based system watch the packet traversing over network link. People require utilizing the IDS to recognize attacks in host and network based system. Network Based IDS observers the packet which traverses over LAN segment and examines the network activity to detect the attacks. Network based Intrusion detection system can monitor the network traffic affecting numerous host which are linked to the network segment, so that it may defend those hosts. Network-based IDS frequently comprise of hosts or a group of single-drive sensors located at many points in LAN. Mostly Sensors are design to work in ―secrecy manner, for the reason of creating it very hard for attackers/intruders to define their existence and place. It is generally used at a border between networks, for example in VPN servers, remote access and wireless network servers. Below are the benefits of using networking based IDS: 1) it can be made undetectable to several attackers to offer safety against attack. 2) Some can monitor a huge network. 3) These are generally passive devices which listen on a network line without intersecting

with the common network operation. Therefore, it is generally simple to adjust in a present network to comprise network-based IDSs with least effort. Disadvantages of using network based IDS are: 1) these are unable to examine encrypted data because mostly organization utilizes VPN. 2) It do not apply to small network segment for example switch based network. Monitoring range of switches is not worldwide, this bounds the network based IDS monitoring range to one host. 3) Some have difficulty in dealing with network based attacks that include the packet fragmentation. This anomalously made packets result the IDS to be unbalanced and crash. A host-based IDS monitors actions related with a specific host and intended at gathering information about movement on a host system or in a computer system. In host based IDS isolated sensors will be required for computer system. Sensor monitors the activity occur in the system. Sensors gather the information from system logs, logs generated by operating system processes, application activity, file access and modification. These log file can be simple text file or operation on a system.

A. Advantages of using Host based IDS are given below

- It can identify attacks which cannot be seen by network based IDS because they monitor local events of a host.
- It function on operating system inspects trails, which can aid to identify attacks include in software integrity breaches.
- It remains genuine by switched networks.

B. Disadvantages of using Host based IDS are

- It may be unseen by certain DoS attacks.
- These are not right for identifying the attacks, those targets an entire network.
- Host based IDS are difficult to manage, as for every individual system; information is configured and managed [9].

VIII. CLASSIFICATION TECHNIQUES

It is a data mining function which allocates items in a group to target groups or classes. Its aim is to correctly forecast the target class for every situation in the data. [10][11]

- Association rule mining
- Bayesian Classification
- Decision tree classification
- Nearest Neighbor
- Neural Networks (Back Propagation)
- SVMs

A. Association Rule Mining

It is the most wide and well explored data mining technique. It aims to extract unusual connection, common patterns or unplanned structures among item sets in the transaction databases or further data sources. These rules are extensively used in many areas for example telecommunication, inventory control, market and risk management etc. many association mining schemes and algorithms will be shortly presented and matched later. Association rule mining is to discover association rules which fulfill the predefined least support and confidence from particular database. The problem is usually decomposed into two sub problems. One is to discover those item sets whose existences surpass a predefined threshold in the database; those item sets are named frequent or huge item sets. The second difficulty is to create association rules from those huge item sets with the limits of least confidence.

B. Bayesian Classification

A simple Bayes classifier specified as the absence or presence of a specific feature is separate to the any other feature, assumed the class variable. This classifier reflects every feature to contribute freely to the possibility that this fruit is an orange, irrespective of the absence or presence of the other features. These can be trained capably in a supervised learning setting. It only needs a small quantity of training data to evaluate the parameters (variances and means of the variables) necessary for classification. Because independent variables are assumed, only the changes of variables for every class require to be resolute and not the whole covariance matrix.

C. Decision Tree Classification

This approach is most useful in classification problems. It is a flow chart like tree structure. Trees are built in a top-down recursive division and conquer manner. In this classification method utilized in many algorithms to categorize the data sets, are: [12].

- ID3 (Iterative Dichotomiser)
- C4.5 (a Successor of ID3)
- Classification and Regression Trees (CART)

The algorithm uses a top-down method that begins with a training tuples set and their related class labels.

1) Advantages

Rules can be generated which are simple to interpret and understand. It is scalable for huge database because the tree size is free of the database size. Every tuple within the database should be filtered by tree, and time is proportional to the height of the tree.

2) Disadvantages

It is does not handle continuous data. [31]Handling missing data is difficult because correct branches in tree could not be taken the labels.

D. Nearest Neighbor

These are depending on learning by analogy. The training samples are defined by n-dimensional arithmetic attributes. Each shows a point in an n-dimensional space. Like this, total samples are collected in an n-dimensional pattern space. When specified an indefinite sample, a k-nearest neighbor classifier finds the pattern space for the k training samples that are closest to the unknown sample. "Closeness" is specified in case of Euclidean distance. The indefinite sample is allocated the general class among its k nearby neighbors. When k=1, the indefinite sample is allocated the class of the training sample which is nearby to it in pattern space. Adjacent neighbor classifiers are instance-based or idle learners in that they collect entire training samples and do not create a classifier until a new (unlabeled) sample requires to be classified.

E. Neural Networks

Neural network represent a brain image or symbol for Information processing [29]. These have been presented to be very capable systems in numerous estimating and business classification applications because of their capacity

to "learn" from the data, their nonparametric behavior and their capacity to simplify. Neural computing introduces as pattern identification method for machine learning. The resultant model is often called a neural network or artificial neural network (ANN). These networks have been utilized in several business applications for pattern identification, prediction, and classification.

F. Support Vector Machine

SVM is very general and useful method for data classification [13]. It is used for classify the both linear and nonlinear data. The aim is to create a model which forecasts the objective value of data occurrence within testing set where only attributes are specified. The classification aim is to divide the two classes by a function formulate from present data and thus to generate a classifier which will effort well on other hidden data. Maximal margin classifier is a simplest form of SVM classification. It is utilized to solve the classification problem, specifically the case of a binary classification with linear divisible training data. The goal of this classifier is to discover the hyperplane with major margin, i.e., the maximal hyperplane, in practical difficulties, training data are not constantly linear separable. To control the nonlinearly separable conditions some slack variables have been presented to SVM in order to endure some training faults, with the influence of the noise in training data thereby decreased. This classifier with slack variables is referred to as a soft-margin classifier.

IX. LITERATURE SURVEY

W. Feng et al. [14] Retrieve network data for intrusion recognition through uniting SVMs and Ant Colony Networks attained good performance in faster running time and detection accuracy rate by joining two present machine learning approaches (CSOACN and SVM). Their proposed work is depending on five main interactive modules. The major aids of this work include the changes to the unsupervised learning CSOACN and supervised learning SVM so that they may be used composed interactively and capably. It also includes the improved SVM and CSOACN to reduce the training data set due to permitting new data points to be added to the training set.

The proposed work by Karim Al-Saedi et al. [15] in their paper Research Proposal: System of An Intrusion Recognition, Alert Decline and also assessment Framework depend on the Data Mining is IDS ARADMF which holds three different systems: Traffic data retrieval and also system of collection mechanism, reduction IDS alert procedures system IDS ready framework process. The movement information recovery and gathering component frameworks add to a system to spare IDS alarms, remove the standard elements as interruption location trade configuration and spare them in DB file (CSV-type). It includes the IDMEF that works as locating alerts and field reduction is utilized as data standardization to create alert format as standard as possible.

Basant Agarwal and Namita Mittal [16] proposed Hybrid Method for Recognition of Anomaly Network Traffic applying Data Mining techniques are hybrid method that exploits the advantages of both the methods i.e. entropy based and SVM based respectively. The hybrid anomaly recognition scheme studies the network traffic performance from the normalized various network features entropy values. Entropy based methods have the better representing advantage the network traffic properties and SVM is good for classification. The normalized entropies are sent to SVM model for study the performance of the network.

Crosbie [17] represents the Genetic programming and various agent methods can be exploiting for network intrusions detecting. A group of agents discovers the grid performances and displays one parameter of the grid review data and genetic programming. The advantage of this method is, many small agents that are independent can be used, but the communication between the agents is a drawback. This system identifies the attacks using a collection of rules generated by genetic algorithm, and then exploits rules for DoS, U2R, R2l, and probe attacks.

J. Gómez and E. León [18] proposed fuzzy and genetic algorithm to categorize activities of intrusion on the network. They used KDDCup99 dataset as input data that consists of 42 features. The fuzzy rule is modified using evolutionary technique and genetic algorithm. The algorithm can categorize the data into DoS, R2L, Probe, U2R, and Normal. This algorithm has detection rate of 98.28 %.

Chetia and Das [19] extended Biswas's manner for assessment of reply scripts of students. They supposed five gratification stages in sequence to assess the recital of students. They contain satisfactory, good, very good ,excellent and insufficient. They have developed an algorithm which takings scholar's data like input and construct a soft set matrix earlier assessing the recital of scholar's.

Parameterization reduction is also probable in soft sets and associated applications as offered by Chen et al. [20]. They additional said which the method followed thru Maji was improper and also demanded which the reduct is not similar for rough set theory and soft set theory. Their idea for reduction of attributes in soft sets was depend on the optimal election idea which addresses the issues of sub-optimal results.

Levent Koc, .et.al [21] presented an intrusion detection model depending on a multinomial classifier that is used to classify network events as normal or attack events, such as DoS, probe, R2L and U2R. This model is depending on a new data mining approach named Hidden Naive Bayes (HNB). The HNB model is applied to many datasets and presents promising outcomes matched with the traditional Naive Bayes and its advanced methods (Jiang, Zhang, & Cai, 2009). The research explores the traditional Naive Bayes and foremost structurally advanced Naive Bayes methods with the new HNB approach. In the study, they augmented the Naive Bayes and advanced Naive Bayes schemes with foremost discretization and feature selection approaches to enhance the accuracy and decrease the resource requirements of intrusion detection problem. Based on the outcomes of the proposed study, HNB based classifier model stands out as simple and practical intrusion detection system with better predictive accuracy and cost.

This study proposed by Ming-Yang Sua et al. [22] is a real-time NIDS with incremental mining for fuzzy association rules. While accepting fuzzy association guidelines to study network traffic, particularly for a NIDS, the time expense, including online data collection and mining procedures, are vital. Incremental fuzzy-rule mining

is appropriate to see real-time demands because it can generate the new rule set, whereas a new data record is collected online. This paper initially suggests an online incremental mining algorithm for creating fuzzy association rules, and then shows real-time intrusion detection system based on the algorithm. By reliably matching the two rule sets, one retrieved from network packets and the other from training attack-free packets, the proposed system can extract a result every 2 seconds. Therefore, matched with traditional static mining methods, the system can significantly increase the efficiency.

X. CONCLUSION

Data mining is a multi-disciplinary procedure, borrowing thoughts from AI and machine learning, statistics, signal theory and image processing, mathematical optimization, pattern recognition. IDS is the main study area in Computer-based security. It is a well-known skill for enlightening and is used to protect data consistency and system accessibility throughout an intrusion. In this paper we study about intrusion detection based data mining, GA, soft set approach, data warehouse and also some previous paper study.

REFERENCE

[1] Muamer N. Mohammada, Norrozila Sulaimana, Osama Abdulkarim Muhsinb, "A Novel Intrusion Detection System by using Intelligent Data Mining in Weka Environment", Procedia Computer Science 3, 2011, pp. 1237 – 1242.
[2] Zhen-Ya Zhang, Hong-Mei Cheng, et al. From data mining to Opportunity / symptoms found. Computer Science. 2007.
[3] Anup Goyal, Chetan Kumar, "GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System".
[4] Sahilpreet Singh,Meenakshi Bansal" A Survey on Intrusion Detection System in Data Mining" Volume No. 2, Issue No. 6, June 2013
[5] Lokendra Singh Parihar, Akhilesh Tiwari "Survey on Intrusion Detection Using Data Mining Methods" IJSART - Volume 2 Issue 1 –JANUARY 2016
[6] Ms.Radhika S.Landge, Mr.Avinash P.Wadhe" Review of Various Intrusion Detection Techniques based on Data mining approach" cations (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 3, May-Jun 2013, pp.430.435
[7] Sneha Kumari, Maneesh Shrivastava "A Study Paper on IDS Attack Classification Using Various Data Mining Techniques" Volume-2 Number-3 Issue-5 September-2012
[8] D. Shona, A.Shobana "A Survey on Intrusion Detection using Data Mining Technique " Vol. 3, Issue 12, December 2015
[9] Sonam Chourse , Prof. Vineet Richhariya "Survey Paper on Intrusion Detection using Data Mining Techniques" (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 8, August 2014).
[10] Zhen-Ya Zhang, Hong-Mei Cheng, et al. From data mining to Opportunity/symptoms found. Computer Science. 2007.
[11] Dai Yingxia, lian Yi-feng, Wang hang. Security and intrusiondetection systems [M]. Beijing: Tsinghua University Press, 2002.
[12] Jiawei Han, Micheline Kambar, Jian Pei, "Data Mining Concepts and Techniques" Elsevier Second Edition.
[13] Pedro G. Espejo, Sebastian Ventura, and Francisco Herrera, "A Survey on the Application of Genetic Programming to Classification", Page: 121-144 VOL. 40, NO. 2 MARCH 2010.
[14] Feng Wenying, Zhang Qinglei, Hu Gomgzhu, Huang Jimmy Xiangi, Mining network data for intrusion detection through combining SVMs with Ant Colony Networks‖, Elsevier, Future Generation Computer Systems 37(2014), pp. 127-140.
[15] Al-Saedi Karim, r ManickamSelvakuma, Ramadass Sureswaran, Al-Salihy Wafaa and ALmomani Ammar, ―Research Proposal: An Intrusion Detection system Alert Reduction and assessment Framework Based on Data Mining, Journal of Computer Science, 2013, ISSN 1549-3636, 9(4): 421-426.
[16] AgarwalBasan, Mittal Namita, ―Hybrid Approach for Detection of Anomaly Network Traffic using Data Mining Techniques‖, Elsevier, Procedia Technology 6(2012)996-1003
[17] Crosbie, Mark, and Gene Spafford. 1995. "Applying Genetic Programming to Intrusion Detection". Proceeding of 1995 AAAI Fall Symposium on Genetic Programming, Cambridge, Massachusetts.
[18] J. Gomez and E. León, "A fuzzy set/rule distance for evolving fuzzy anomaly detectors," IEEE International Conference on Fuzzy Systems ART. No. 1682017, pp. 2286-2292.
[19] Samsiah Abdul Razak and Daud Mohamad, "A Soft Set based Group Decision Making Method with Criteria Weight", World Academy of Science, Engineering and Technology 58 2011.
[20] B. Chetia and P. K. Das, "Application of Vague Soft Sets in students' evaluation". Advances in Applied Science Research, 2011, 2 (6):418-423.
[21] Koc Levent, Mazzuchi Thomas A., Sarkani Shahram, ―A network intrusion detection system based on a Hidden Naive Bayes multiclass classifier‖, Elsevier, Expert Systems with Applications 39 (2012) 13492–13500.
[22] Su Ming-Yang, Yu Gwo-Jong, Lin Chun-Yuen, ―A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach‖, Elsevier, Computers & Security 28 (2009)301–309.

