Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.11 No.1, 2020

A Review on Various Methods of Intrusion Detection System

Diptee Agrawal
CSE, RITS, Bhopal, India

Chetan Agrawal
CSE, RITS, Bhopal, India

Abstract
Detection of Intrusion is an essential expertise business segment as well as a dynamic area of study and expansion caused by its requirement. Modern day intrusion detection systems still have these limitations of time sensitivity. The main requirement is to develop a system which is able of handling large volume of network data to detect attacks more accurately and proactively. Research conducted by on the KDDCUP99 dataset resulted in a various set of attributes for each of the four major attack types. Without reducing the number of features, detecting attack patterns within the data is more difficult for rule generation, forecasting, or classification. The goal of this research is to present a new method that Compare results of appropriately categorized and inaccurately categorized as proportions and the features chosen. Data mining is used to clean, classify and examine large amount of network data. Since a large volume of network traffic that requires processing, we use data mining techniques. Different Data Mining techniques such as clustering, classification and association rules are proving to be useful for analyzing network traffic. This paper presents the survey on data mining techniques applied on intrusion detection systems for the effective identification of both known and unknown patterns of attacks, thereby helping the users to develop secure information systems.
Keywords: IDS, Data Mining, Machine Learning, Clustering, Classification
DOI: 10.7176/CEIS/11-1-02
Publication date: January 31st 2020

1. INTRODUCTION

Now a day's usage of internet and world wide connectivity has been grown, well-proportionate with cyber attacks. Maintaining Cyber security is a severe universal fright. The chance of computerized attacks occurring is inevitable despite, the worst security safeguards which result in considerable troubles to persons, firms and companies. Intrusion Detection System (IDS) has turn out to be an indispensable part of system security to identify several attacks with an intension of shielding systems from extensive harms and recognizing risks of the intruded system. Therefore finding intrusions accurately becomes chief functionality of most Intrusion Detection Systems. So keeping information protected in every organization either private, public or government is necessary. Even though many IDSs exist but main difficulty, lies in poor detection rates and larger false positives. This issue is the focus of our research, increasing the capability of detecting intrusions rates and lessening the false alarms. The worst performance of existing techniques is because of raw dataset that bewilders the classifier and results with inaccurate detection due to unnecessary features. There are several methods in data mining used for the development of IDS systems, like Clustering, Classification, Association Rules, Genetic Algorithms, Decision Trees and Artificial Neural Networks.

Network security has a remarkable history, with some of the initial inspection technologies emerging in the 1980s. While that time, lots of techniques have been employed to attempt to create this task easier. The inspirations are fundamentally the similar, but the methods become more complicated over time. This chapter will converse the environment of designs behind assessment and intrusion detection, as well as fundamental machine learning models that can be utilized. A few examples of how these models have been applied will be discussed, as will the fundamental algorithms behind the construction of these models.

1.1 Overview of Ids

In day to day life the need for speed access of information through internet has increased. Hence the room for maintaining security in any organization either public or private system has become fundamental. Because of increase in network connections and systems, unauthorized access and interruption of the data is triggered. As a result, it is indispensable to create a virtual access path. In general intruders have capacity to find out defect in systems or networks and can spawn vulnerabilities. Even though the access control points exist in network, they fail in providing scrupulous security to the systems. To identify intruders, developing Intrusion Detection Systems (IDSs) is the best solution to protect systems and networks. Therefore the task of IDS is not only to detect intruders but also to monitor the raid of intruders. An accurate system of protecting data and resources from illicit access, damaging and denial of use is to be built. For every system, the security perspective is to be planned based on the expected performance. Mainly security is concerned with the following aspects in a computer system.

- ***Confidentiality*** **–** information is to be accessed only by permissible persons.
- ***Integrity*** **–** information must remain unaffected by destructive or malicious attempts.
- ***Availability*** **–** computer is responsible to function without downgrading of access and provide resources to legal users when they require it.

Specifically an intrusion is defined as a set of events which are unknown and unforeseen to the user, which compromises the protection of a computer system. It can be done from external side or internal side of the system. Earlier in 1980‟s James P Anderson has defined intrusion as the scope of illegal force to access information, defraud information, or making the computer system unsafe. Intrusion Detection System (IDS) was commercially promoted in the year 1990. From then a variety of layouts were introduced to adapt intrusion detection systems [1] [2]. It acts like a burglar alarm and detects any kind of violation and generates alarms like audible, visual and also messages like e-mail. On the whole, IDS is primarily exploited for stopping defective activities that may attack or misuse the system by identifying attacks through providing desirable support for defense management and also give constructive information regarding intrusion. But structure of IDS should possess low fake alarms while undertaking the discovery of attacks. IDSs have become shielding mechanisms everywhere in current networks. There is no thorough and proficient methodology offered in checking the strength of these systems.

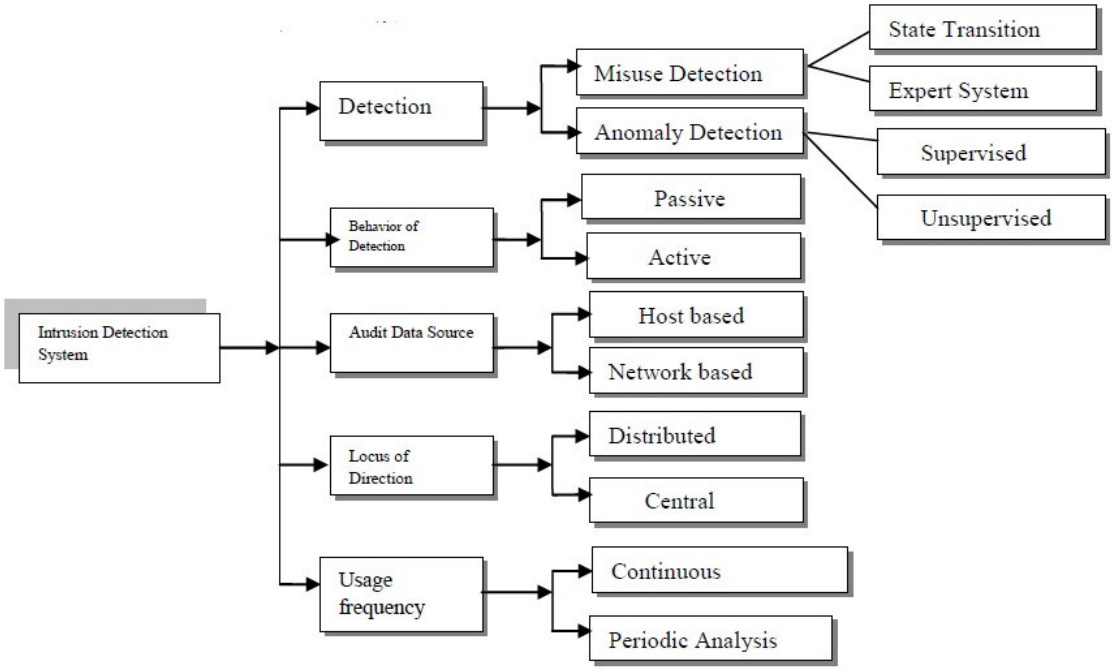

Figure. 1: Classification of IDS based on its characteristics

Because Intrusion Detection Systems performance is increased with usage of the Soft Computing methods to IDS, the computer security researchers are trying to apply. Soft computing is the collection of approaches that were set up to model and obtain guaranty solutions to real world problems, which are not modeled or very difficult to model mathematically. Soft computing is a general term for developing the enduring for imprecision, partial truth, uncertainty, and estimate of achieving flexibility and minor solution cost. A masterful and accurate tool for real time intrusion discovery is the target of main experimenters in IDSs. There is a variety of Artificial Intelligence (AI) concepts were exploited for transforming intrusion discovery procedure, therefore human involvement is decreased. And also in common, the procedures which deal with IDS are utilizing machine learning. Basically Soft Computing techniques that were used in IDS implementation are Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Bayesian Networks, Fuzzy logic, Particle Swarm Optimization and Genetic Algorithms (GA).

In detecting intrusions, IDS defends a computer network from illicit persons, possibly insiders. The attack recognition task is considered as the model of classification expert in distinguishing “harmful” connections referred as intrusions or attacks, and „sympathetic‟ connections referred as normal. There are various categories of IDSs are prevailing that are based on structure and detection method. In addition to these, there are other characteristics one can used to classify IDS as shown in the fig. 1.

1.2. Types of Ids Based On Structure

IDSs are mainly classified into following three categories based on framework. They are namely

- Network based IDS

- Host based IDS
- Application based IDS

1.2.1. Network based Intrusion Detection System

Network based IDS are best suited for alert generation of intrusion from outside the perimeter of the enterprise. NIDS are inserted at various points on Local Area Network (LAN) and observe packets traffic on the network information is assembled into packets and transmitted on LAN or Internet. NIDS are more worth when they are placed outside the firewalls, thereby alerting personals of incoming packets that might get avoided to the firewall. Some NIDS allow taking input of custom signatures taken from user security policy which permits limited detection security policy violation. This limitation is due to packets traffic information that does not work well today in switched and encrypted environments, where packets analysis is weak in detecting, attacking or originating from authorized network users. NIDS make use of raw network packets as the data source. The IDS typically use a network adapter in licentious mode that listens and analyses all traffic in real-time as it travels across the network.

1.2.2. Host Based IDS

HIDS monitors incoming and outgoing activity on a particular system in the network. Specifically, it monitors the dynamic behavior and the state of the computer system. The administrator will be notified once an intrusion has been detected. An NIDS is usually used alongside a HIDS in order to identify any activities that HIDS overlooked. Host-based IDS places monitoring sensors also known as agents on network resources nodes to monitor audit logs which are generated by network operating system or application program. Audit logs contain records for events and activities taking place at individual Network resources. It is done because these HIDS can detect attacks that cannot be seen by NIDS such as Intrusion and can be misused by trusted insider. Host based systems utilize signature rule base which is derived from site-specific security policy. Host based can overcome the problems associated with Network based IDS immediately after alarming the security personnel who can locate the source provided by site security policy. HIDS also verifies if any attack is unsuccessful, either because of immediate response to alarm or any other reason. But this is not available at packet level.

1.2.3. Application Protocol-based IDS (APIDS)

APIDS monitors activity in the specific protocols used in the computer system. Then, it looks for protocols and insists on the correct use in the systems. An APIDS monitors the state of the protocol and dynamic behavior. APIDS have a system or agent which is located between a process, or group of servers, monitoring and analyzing the application protocol between two connected devices. APIDS is usually placed between a web server and the database management system. It keeps on monitoring the SQL protocol specific to the middleware / business logic, as it interacts with the database. At the preliminary stage, APIDS finds, and enforces the correct use of the protocol.

2. DATA MINING METHODS

The term data mining is used to describe the process of extracting useful information from the large databases. Data mining analyses the observed sets to discover the unknown relation and sum up the results of data analysis to make the owner of data to understand [3]. Hence data mining problems are considered as a data analysis problem. Data mining framework automatically detect patterns in our data set and use these patterns to find a set of malicious binaries.ie, Data mining techniques can detect patterns in large amount of data, such as byte code and use these patterns to detect future instances in similar data.

In intrusion detection system, information comes from various sources like host data, network log data, alarm messages etc. Since the variety of different data sources is too complex, the complexity of the operating system also increases. Also network traffic is huge, so the data analysis is very hard. The data mining technology have the capability of extracting large databases; it is of great importance to use data mining techniques in intrusion detection. By applying data mining technology, intrusion detection system can widely verify the data to obtain a model, thus helps to obtain a comparison between the abnormal pattern and the normal behavior pattern. Manual analysis is not required for this method. One of the main advantages is that same data mining tool can be applied to different data sources. An important problem in intrusion detection is how effectively can separate the attack patterns and normal data patterns from a large number of network data and how effectively generate automatic intrusion rules after collected raw network data. To accomplish this various data mining techniques are used such as classification, clustering, association rule mining etc. Examples for Data Mining based Misuse detection model of IDS are JAM (Java Agents for Meta learning), MADAM ID (Mining Audit Data for Automated Models for Intrusion Detection), and Automated Discovery of Concise Predictive Rules for Intrusion Detection.

Examples for Data Mining based Anomaly detection model for IDS are MINDS and EBays. Examples for Data Mining based both Anomaly and Misuse detection model for IDS are IIDS (Intelligent Intrusion Detection System Architecture) and RIDS-100 (Rising Intrusion Detection System).

System Design

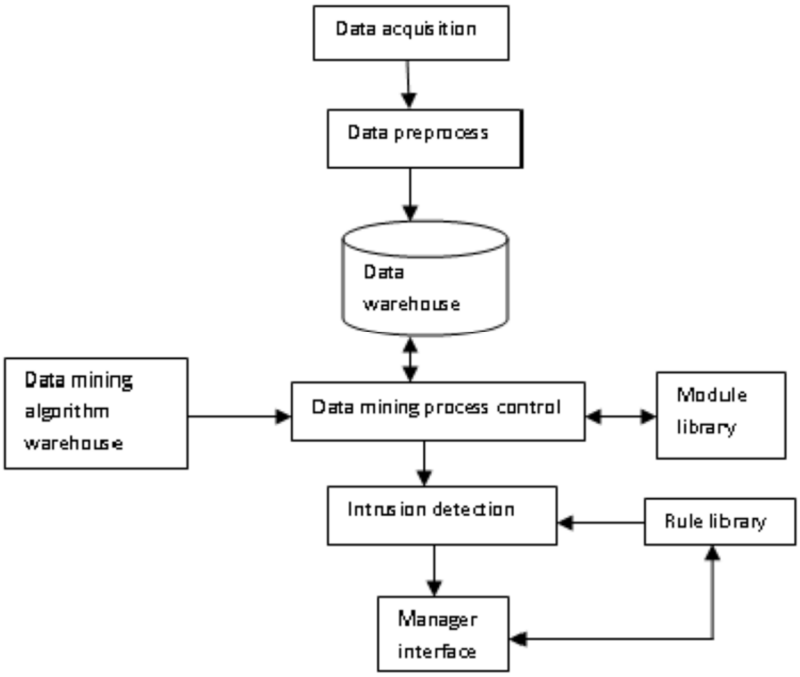

Figure 2: Data Mining based Intrusion Detection Pattern

Authors of [4] presented a system design based on Data Mining intrusion detection pattern which is illustrated in figure 2.The functions of each module is described as follows:

Data Acquisition- Examines the network data for attack patterns and captures every packet for processing.

Data Preprocess module- Certain techniques like data cleaning, data integration, data reduction techniques etc are used to convert a data packet into suitable mining forms.

Data Mining module- Data obtained from the data preprocess module is stored in the data warehouse. Also it mines the training data which the control module outputs. Data in the data warehouse becomes more and richer, since the data gathering process carries on continuously. So data warehouse contains all kinds of data and mine useful information. Data mining algorithm library consists of different data mining algorithms like sequence pattern analysis algorithm, connection rule algorithm, sorting algorithm, clustering algorithm etc. Data mining process control is responsible for choosing efficient and appropriate mining algorithm from the library and hence it is the key of the system model. Here system design considers training dataset, the data mining control module may extract features and the patterns for detection using the algorithms from the algorithms library. When the system begins running and the dataset is not trained, then the data mining control module may be trained by taking data from the data warehouse. Then data in the dataset can be categorized as normal data and attack data by choosing clustering algorithm. Finally, the output of the mining is transmitted to intrusion detection module.

Intrusion Detection Module- Intrusion detection module is a combination of rule library, Intrusion detection. Rule library is the repository of rules that is needed for intrusion detection and check for a matching with the output from the data mining modules.

Manager Interface Module- This module is responsible for making decision on the normal pattern and abnormal pattern. If the decision result is normal pattern, adds it to its close normal pattern in the library. If the result is abnormal pattern, add it to its close abnormal pattern in the library and carries out necessary preventive measures.

3. DATA MINING METHODS AND INTRUSION DETECTION

Data Mining is used in variety of applications that requires data analysis. Now a day's data mining techniques plays an important role in intrusion detection systems. Different data mining techniques like Classification, Clustering and Association rules are frequently used to acquire information about intrusions by observing network data. This section describes different data mining techniques that help in detecting intrusions.

Classification: Classification is a form of data analysis which takes each instance of a dataset and assigns it to a particular class. It extracts models defining important data classes. Such models are called classifiers [5]. A classification based IDS will classify all the network traffic into either normal or malicious. Data classification consists of two steps – learning and classification. A classifier is formed in the learning step and that model is used to predict the class labels for a given data in the classification step. Classification analysis requires that the end-user/analyst know ahead of time how classes are defined [6]. Each record in the dataset already has value for the

attribute used to define the classes. The objective of a classifier is not to explore the data to discover different classes, but to find how new records should be arranged into classes. Classification helps us to categorize the data records in a predetermined set. It can be used as attribute to label each record and for distinguishing elements belonging to the normal or malicious class [6]. Different types of classification techniques are decision tree induction, Bayesian networks-nearest neighbor classifier, genetic algorithm and fuzzy logic.

As compared to the clustering technique, classification technique is less efficient in the field of intrusion detection. The main reason for this is the enormous amount of data needed to be collected to use classification. To classify the dataset into normal and abnormal, large amount of data is required to analyze its proximity. Classification method can be useful for both misuse detection and anomaly detection, but it is more commonly used for misuse detection. Authors of [6] presented a data classification for intrusion detection that can be achieved by the following steps:-

1. In order to study about the classification models of the normal and abnormal sequences of system calls, we want to supply it with a training data set, containing pre-labeled normal or abnormal sequences. Different techniques like linear discrimination, decision tree or rule based methods is used to scan the network traces. Then generate a collection of unique sequence of system calls and named it as normal list.
2. Next scan each of the intrusion traces. Find each sequence of system calls in the normal list. If an exact match be found then labeled them as normal. Otherwise it is labeled as abnormal.
3. Next ensure that the normal traces consist of all possible normal short sequence of system calls. An intrusion trace contains combination of normal and abnormal sequences of system calls since abnormal sequence only appear in some places.

Clustering Since the amount of available network data is too large, human labeling is time-consuming, and expensive. Clustering is the process of labeling data and assigning into groups.ie, Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of members from the same cluster are quite similar and members from the different clusters are different from each other. Hence clustering methods can be useful for classifying network data for detecting intrusions. Clustering algorithms can be classified into four groups: partitioning algorithm, hierarchical algorithm, density-based algorithm and grid based algorithm [7].

Clustering techniques can discovers complex intrusions over a different time period. Clustering is an unsupervised machine learning mechanism for discovering patterns in unlabeled data with many dimensions. Clustering is the collection of patterns based on similarity. Patterns within a cluster are equivalent to each other, but they are different with other clusters. Therefore patterns that are far from any of these clusters indicate that an unusual activity happened. That can be part of a new attack. Clustering can be applied on both Anomaly detection and Misuse detection.

Authors of [6] presents basic steps involved in identifying intrusion are follows:-

1. Find the largest cluster, which consists of maximum number of instances, and label it as normal.
2. Sort the remaining clusters in an ascending order of their distances to the largest cluster.
3. Select the first K1 clusters so that the number of data instances in these clusters sum up to ¼`N, and label them as normal, where ` is the percentage of normal instances.
4. Label all other clusters as malicious.
5. After clustering, heuristics are used to automatically label each cluster as either normal or malicious. The self labeled clusters are then used to detect attacks in a separate test dataset.

Association Rule

The association rule considers each attribute/value pair as an item. Collection of items referred as an item set in a single network request. The algorithm searches to find an item set from large number of dataset that frequently appears in network. The main aim of association rule is to derive multi-feature correlations from a database table [6]. Association rule mining determines association rules and/or correlation relationships among large set of data items. Association rule shows conditions for attribute values that occur frequently in the dataset. An example of association rule mining is Market Basket analysis. Association rules are obtained from the dataset and they are in the form of “if-then” statements. Apriori was the first scalable algorithm developed for association rule mining. Association rule mining in intrusion detection is very useful in many ways. Authors of [6] present basic steps for incorporating association rule for intrusion detection as follows:-

1. First network data need to be arranged into a database table where each row is an audit record and each column is a field of the audit records.
2. It is always shows that the intrusions and user activities shows frequent correlations among network data. Consistent behaviors’ in the network data can be captured in association rules.
3. Also rules based on network data can continuously merge the rules from a new run to the aggregate rule set of all previous runs.
4. Thus with the association rule, we get the capability to capture behavior in association rule for correctly detecting intrusions

4. LITERATURE SURVEY

Manjula C. Belavagi et al [8] states Intrusion detection model is a predictive model used to predict the network data traffic as normal or intrusion. Machine Learning algorithms are used to build accurate models for clustering, classification and prediction. In their paper, classification and predictive models for intrusion detection are built by using machine learning classification algorithms namely Logistic Regression, Gaussian Naive Bayes, Support Vector Machine and Random Forest. These algorithms are tested with NSL-KDD data set. Their Experimental results show that Random Forest Classifier out performs the other methods in identifying whether the data traffic is normal or an attack.

SMH Bamakan et. al. [9] states many organizations recognize the necessities of utilizing sophisticated tools and systems to protect their computer networks and reduce the risk of compromising their information. Although many machine-learning-based data classification algorithm has been proposed in network intrusion detection problem, each of them has its own strengths and weaknesses. In this paper, we propose an effective intrusion detection framework by using a new adaptive, robust, precise optimization method, namely, time-varying chaos particle swarm optimization (TVCPSO) to simultaneously do parameter setting and feature selection for multiple criteria linear programming (MCLP) and support vector machine (SVM). In the proposed methods, a weighted objective function is provided, which takes into account trade-off between the maximizing the detection rate and minimizing the false alarm rate, along with considering the number of features. Furthermore, to make the particle swarm optimization algorithm faster in searching the optimum and avoid the search being trapped in local optimum, chaotic concept is adopted in PSO and time varying inertia weight and time varying acceleration coefficient is introduced. The performance of proposed methods has been evaluated by conducting experiments with the NSL-KDD dataset, which is derived and modified from well-known KDD cup 99 datasets. The empirical results show that the proposed method performs better in terms of having a high detection rate and a low false alarm rate when compared with the obtained results using all features.

Adel Sabry Eesa et. al. [10] presents a new feature-selection approach based on the cuttlefish optimization algorithm which is used for intrusion detection systems (IDSs). Because IDSs deal with a large amount of data, one of the crucial tasks of IDSs is to keep the best quality of features that represent the whole data and remove the redundant and irrelevant features. The proposed model uses the cuttlefish algorithm (CFA) as a search strategy to ascertain the optimal subset of features and the decision tree (DT) classifier as a judgment on the selected features that are produced by the CFA. The KDD Cup 99 dataset is used to evaluate the proposed model. In this study, they investigated the combination model of CFA and DT for feature selection for intrusion detection and evaluated its performance based on the benchmark KDD Cup 99 intrusion data. Firstly, we have modified the CFA to be used as a feature selection tool. Then, they used DT classifier as measurement on the generated features. Empirical results reveal that the produced features are performed the Detection Rate (DR) and Accuracy Rate (AR) (especially when the number of produced features was equal or less than 20 features. In general whenever the number of features is decreased, the AR and DR are increased. The results show that the feature subset obtained by using CFA gives a higher detection rate and accuracy rate with a lower false alarm rate, when compared with the obtained results using all features.

Bin Luo et. al. [11] presents, a four-angle-star based visualized feature generation approach, FASVFG, is proposed to evaluate the distance between samples in a 5-class classification problem. Based on the four angle star image, numerical features are generated for network visit data from KDDcup99, and an efficient intrusion detection system with fewer features is proposed. The FASVFG-based classifier achieves a high generalization accuracy of 94.3555% in validation experiment, and the average Mathews correlation coefficient reaches 0.8858. Compared with the previous IDSs, the key improvement of this new intrusion detection system stems from the novel feature generation approach with visualization strategy. As we mentioned before, visualization is an intuitive way for feature selection and feature reduction. The main contributions of this work consist of two parts.

First is the visual simulation for high dimensional data. Through the proposed FASVFG system, people are capable of mapping an unknown network visit into a geometric point in the four angle star and inferring whether it is a normal visit or a certain malicious attack. The closer the point is to one vertex, the more possible it could belong to an attack category. This approach transforms sophisticated network action into a visual vision and offers a meaningful and effective pre-knowledge for afterwards classifier.

Second is the meaningful attempt in feature reduction. As shown in experiments results, FASVFG is successfully to solve the 5-class classification problem, in which the classification accuracy of the IDS achieves 94.3555%, and average MCC value achieves 0.8858. Though the MCC value of FASVFG is slightly lower than that with full features, namely, 0.9, the size ratio of feature space is 16/43, that ensures a more efficient detection rate.

Salma Elhag et. al. [12] states Security policies of information systems and networks are designed for maintaining the integrity of both the confidentiality and availability of the data for their trusted users. However, a number of malicious users analyze the vulnerabilities of these systems in order to gain unauthorized access or to compromise the quality of service. For this reason, Intrusion Detection Systems have been designed in order to

monitor the system and trigger alerts whenever they found a suspicious event.

Optimal Intrusion Detection Systems are those that achieve a high attack detection rate together with a small number of false alarms. However, cyber attacks present many different characteristics which make them hard to be properly identified by simple statistical methods. According to this fact, Data Mining techniques, and especially those based in Computational Intelligence, have been used for implementing robust and accuracy Intrusion Detection Systems.

In this paper, they consider the use of Genetic Fuzzy Systems within a pair wise learning framework for the development of such a system. The advantages of using this approach are twofold: first, the use of fuzzy sets, and especially linguistic labels, enables a smoother borderline between the concepts, and allows a higher interpretability of the rule set. Second, the divide-and-conquer learning scheme, in which we contrast all possible pair of classes with aims, improves the precision for the rare attack events, as it obtains a better separability between a ‘‘normal activity’’ and the different attack types. The goodness of our methodology is supported by means of a complete experimental study, in which we contrast the quality of our results versus the state-of-the-art of Genetic Fuzzy Systems for intrusion detection and the C4.5 decision tree.

Ning Cao et al [13] states with the arrival of big data era, data mining techniques have been widely used to build models for cyber security applications such as spam filtering, malware or virus detection, and intrusion detection. This project proposes a novel approach that uses randomness to improve robustness of data mining models used in cyber security applications against attacks that try to evade detection by adapting. Their approach addresses three problems. First, they build a diverse pool of mining models to improve robustness of a variety of mining algorithms. These methods are similar to ensemble learning but optimize the tradeoff between mining quality and robustness. These methods also require very little modification to existing algorithms. Second, they randomly select a subset of models at run time (when the model is used for detection) to further boost robustness. Third, they propose a theoretical framework that bounds the minimal number of features an attacker needs to modify given a set of selected models.

Yi Yi Aung et al [14] states the security of the computer system is of great importance, And in the last few years, there have seen an affected growth in the amount of intrusions that intrusion detection has become the dominant of current information security. Firewalls cannot provide complete protection. Applying on a firewall system alone is not enough to prevent a corporate network from all types of network attacks. Therefore more system should be complemented by intrusion detection system. Data mining skills can be used as an effective approach to detect intrusions in intrusion detection system. Data Mining and Knowledge Discovery is the computerized process of trenching and analysis of huge amounts of data, and then extract the meaning of the data. Data mining tools can assist to predict future behaviors and trends, so that organizations proactively, can make decisions based on knowledge. Data mining can answer organization questions that were too traditional time, to solve. Data mining takes its name from the valuable information in a large database. Recent studies display that cascading based approaches of several algorithms are much better performance than an individual algorithm. In this research, they use K-means and Random Forest algorithm to classify instances. This model was verified using KDD’99 data set. Experimental results show that hybrid methods can support suitable detection rates and lower model training time than using single algorithm.

5. PROBLEM STATEMENT

In ideal IDS should have an important characteristic, such as high accuracy detection rate, low false positive and negative rate, and low computational cost. Present IDS should cope with a new challenge such as.

- Fast expansion in network systems and digital devices, i.e. smart mobile phones, tablets, laptops.
- Increasing numbers of sophisticated network intrusions, such as systematic multistep detailed process taken by a hacker before conducting the attack.
- Occurrence of a new intrusions and software vulnerabilities.

It has been widely observed that the large amount of dataset available for text categorization and clustering often give unexpected and irrelevant results due to noisy, irrelevant or misleading features found in stored information. To overcome this problem some of the important and most relevant classifiers should be selected which serve as the basis for further analysis of the given data

The DARPA KDD Cup ‘‘99 dataset has been used by most of the researchers as a testing and training purpose for the development of efficient and secure System. As it is very huge with each record composed of 41 features, making of a rule set is very complicated. The main problem is that how we decided which classifiers are best on which factors? Some basic problems in classification which include:

- How do we identify best classifiers for each individual attacks.
- What are the strengths and weaknesses of existing classifiers applied to KDD data set?

6. OBJECTIVE AND RESEARCH SCOPE

Intrusion detection systems are nowadays recognized as fundamental tools for the security of computer systems.

IDSs aim at identifying violations of security policies and perform automatic counteractions to protect computer systems and information. As soon as IDSs are deployed, they may become target of attacks that may severely undermine or mislead their capabilities. To the best of our knowledge, this paper is the first survey on adversarial attacks against IDSs, a relevant topic especially for safety–critical environments. In this synopsis we provided the following contributions:

- We provided a general taxonomy of attack tactics against intrusion detection systems;
- We subdivided the IDS task into three different phases,

Moreover, throughout the paper we identified a number of challenging issues that should be addressed by future research activities on intrusion detection. We focus our attention on a few of them:

- Strengthening the measurement mechanisms by relying on both host and network sensors, and exploiting the concept of redundancy as performed by data reconciliation techniques in process control. In addition, in-VM monitoring showed to be a very promising way to strengthen measurements at the host level.
- Enhancement of the description of alerts in anomaly based systems through automatic attack inference mechanisms. This may definitely cope with the lack of informative output in anomaly based systems that may allow for the detection of variants of known, or never-before- seen intrusions. Moreover, exploiting contextual information about the systems being monitored (e.g., for performing alert verification) seems the natural way to deal with over stimulation attacks, as well as false alarms in general.
- Responses against intrusions based on cost-sensitive models, game theory and proactive techniques should be further investigated. Human expertise will always play a central role, but these methods can be helpful to automate the response process and make it effective against an adversary.
- IDS solutions are expected to increasingly implement machine learning mechanism s, to deal with the complexity of the intrusion detection task. Consequently, techniques based on adversarial machine learning are worth being further investigated.

CONCLUSION

Presently so many techniques, method and tools are uses to detect Intrusion in computer network and continues research is being carried out to make them even better to recognize intrusion. But simultaneously new attacks arrived which will difficult to Handel because they continues changes their behaviors. In this paper, we describe different data mining technique applied for detecting intrusions. This paper provides the details of two types of intrusion detection and general working principle of IDS. Misuse detection techniques are not sufficient for identifying unknown attacks. For detecting unknown intrusions, we need to go for anomaly detection. Also this paper presents the main concepts of data mining process and the system design for data mining based intrusion detection pattern. Different Data mining techniques like classification, clustering and association rule are very helpful in analyzing the network data. Since large amount of network traffic needs to be collected for intrusion detection, clustering is more suitable than classification in the domain of intrusion detection. Data mining technology helps to understand normal behavior inside the data and use this knowledge for detecting unknown intrusions.

References

[1.] Teresa F. Lunt., “A survey of intrusion detection techniques”, Computers and Security, Elsevier Advanced Technology Publications, 12(4):405-418, 1993.
[2.] Emilie Lundin, Erland Jonsson” Survey of Intrusion Detection Research” , Technical Report 02-04, Department of Computer Engineering, Chalmers University of Technology, 2002
[3.] Wang, Xiao-bin, Guang-yuan Yang, Yi-chao Li, and Dan Liu. "Review on the application of artificial intelligence in antivirus detection system i." In *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*, pp. 506-509. IEEE, 2008.
[4.] Min, L. I. "Application of Data Mining Techniques in Intrusion Detection." *An Yang Institute of Technology* (2005).
[5.] Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
[6.] Lu, C-T., Arnold P. Boedihardjo, and Prajwal Manalwar. "Exploiting efficient data mining techniques to enhance intrusion detection systems." In *Information Reuse and Integration, Conf, 2005. IRI-2005 IEEE International Conference on.*, pp. 512-517. IEEE, 2005.
[7.] Jianliang, Meng, Shang Haikun, and Bian Ling. "The application on intrusion detection based on k-means cluster algorithm." In *Information Technology and Applications, 2009. IFITA'09. International Forum on*, vol. 1, pp. 150-152. IEEE, 2009.
[8.] Belavagi, Manjula C., and Balachandra Muniyal. "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection." *Procedia Computer Science* 89 (2016): 117-123.
[9.] Bamakan, Seyed Mojtaba Hosseini, Huadong Wang, Tian Yingjie, and Yong Shi. "An effective intrusion detection framework based on MCLP/SVM optimized by time-varying chaos particle swarm optimization."

Neuro computing 199 (2016): 90-102.
[10.] Eesa, Adel Sabry, Zeynep Orman, and Adnan Mohsin Abdulazeez Brifcani. "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems." Expert Systems with Applications 42, no. 5 (2015): 2670-2679.
[11.] Luo, Bin, and Jingbo Xia. "A novel intrusion detection system based on feature generation with visualization strategy." Expert Systems with Applications 41, no. 9 (2014): 4139-4147.
[12.] Elhag, Salma, Alberto Fernández, Abdullah Bawakid, Saleh Alshomrani, and Francisco Herrera. "On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems." Expert Systems with Applications 42, no. 1 (2015): 193-202.
[13.] Cao, Ning, and Yingying Wang. "A Novel Approach to Improve Robustness of Data Mining Models Used in Cyber Security Applications." In *Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), 2017 IEEE International Conference on*, vol. 2, pp. 297-300. IEEE, 2017.
[14.] Aung, Yi Yi, and Myat Myat Min. "An analysis of random forest algorithm based network intrusion detection system." In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017 18th IEEE/ACIS International Conference on*, pp. 127-132. IEEE, 2017.