IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 4, April 2016

Intrusion Detection System Built around Hybrid Technology: A Review

Premansu Sekhara Rath[1], Dr.Nalinikanta Barpanda[2], Subodh Panda[3]

Research Scholar, SSSUTMS, Bhopal, Madhya Pradesh, India[1,3]

Department of Electronics, GIET, Gunupur, Orissa, India[2]

Abstract: The use of computer network based services play an incredible, vital and major role in our daily life. Hence, network protection and security is getting more and more significance than ever. It is the intrusion attack which poses a serious security risk in network environment. Hence, to maintain a high level security to ensure safe and trusted communication of information between various parties is very important. This paper focuses on study of existing intrusion detection system by using hybrid technology that is data mining along with soft computing techniques. There are different approaches being employed in IDS but unluckily each of the technique so far is not entirely ideal. In this paper, I have gone through a few research papers regarding the basics of IDS, the methodologies, good fuzzy classifiers using genetic algorithms and data mining techniques which are the focus of the solution for the problem of IDS. Ultimately, a discussion of upcoming technologies and various methodologies which promise to improve the capability of computer system to detect intrusion is offered.

Keywords: Intrusion detection system, Genetic algorithm, Fuzzy Logic, Data mining, Attack , Clustering.

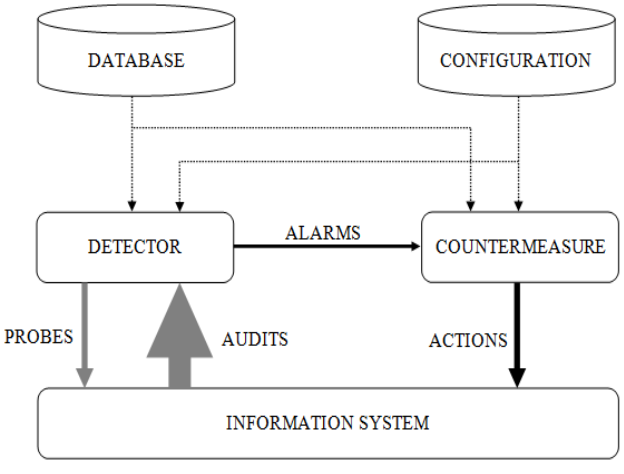
1. INTRODUCTION

Intrusion detection System (IDS) is a type of security management system for computers and networks [1]. An intrusion detection system (IDS) inspects all outbound and inbound network action and find out the doubtful patterns that may point to a network or system intrusion or attack from someone trying to crack into a system. Intrusive events to computer networks are expanding because of the liking of adopting the internet and local area networks [6] and new automated hacking tools and strategy. Computer systems are evolving to be more and more exposed to attack, due to its wide spread network connectivity. There are four major categories of networking attacks: Denial of Service, Probing, User to Root (U2R) and Remote to Local (R2L).Intrusion detection system is the area where data mining concentrate heavily. There are two fold reasons for this first an IDS is very common and very popular and extremely critical activity. Second, large volume of the data on the network is dealing so this is an ideal condition for the data mining to use it. The data mining technology has the enormous benefits in the data extracting attributes and the rule, so it is significant to use data mining methods in the intrusion detection [6]. A significant problem of IDS is how to efficiently divide the normal behavior and the abnormal behavior from a huge number of raw information's attributes and how to effectively generate automatic intrusion rules by following composed raw data of the network. To accomplish this, different data mining methods must be studied, like classification, correlation analysis of data mining methods and so on [6].

The ever rising new intrusion or attacks type poses severe difficulties for their detection. The human labeling of the accessible network audit information instances is generally tedious, expensive as well as time consuming.

This paper focuses on study of existing intrusion detection task by using data mining techniques and discussing on various issues in existing IDS based on data mining techniques.

Intrusion Detection Systems (IDS) are primarily focused on identifying probable incidents, monitoring information about them, tries to stop them, and reporting them to security administrators [7] in real-time environment, and those that exercise audit data with some delay (non-real-time). The latter approach would in turn delay the instance of detection. In addition, organizations apply IDSs for other reasons, such as classifying problems with security policies, documenting existing attacks and preventing individuals from violating security policies. IDSs have become a basic addition to the security infrastructure of almost every organization. A usual Intrusion Detection System is demonstrated in Figure 1.



IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 4, April 2016

NOTE: The arrow lines symbolize the amount of information flowing from one component to another Figure

1. Very Simple Intrusion Detection System

Intrusion Detection Systems are broadly classified into two types. They are host-based and network-based intrusion detection systems. Host-based IDS employs audit logs and system calls as its data source, whereas network-based IDS employs network traffic as its data source. A host- based IDS consists of an agent on a host which identifies different intrusions by analyzing audit logs, system calls, file system changes (binaries, password files, etc.), and other related host activities. In network-based IDS, sensors are placed at strategic position within the network system to capture all incoming traffic flows and analyze the contents of the individual packets for intrusive activities such as denial of service attacks, buffer overflow attacks, etc. Each approach has its own strengths and weaknesses. Some of the attacks can only be detected by host-based or only by network-based IDS.

The Intrusion Detection System (IDS) is also carried out by implementing Genetic Algorithm (GA) to efficiently identify various types of network intrusions. The genetic algorithm [2] is applied to achieve a set of classification rules from the support-confidence framework, and network audit data is employed as fitness function to judge the quality of each rule. The created rules are then used to classify or detect network intrusions in a real-time framework. Unlike most available GA-based approaches remained in the system, because of the easy demonstration of rules and the efficient fitness function, the proposed system is very simple to employ while presenting the flexibility to either generally detect network intrusions or precisely classify the types of attacks.

The normal and the abnormal intrusive activities in networked computers are tough to forecast as the boundaries cannot be well explained. This prediction process may generate false alarms [2] in many anomaly based intrusion detection systems. However, with the introduction of fuzzy logic, the false alarm rate in determining intrusive activities can be minimized; a set of fuzzy rules (non-crisp fuzzy classifiers) can be employed to identify the normal and abnormal behavior in computer networks and fuzzy inference logic can be applied over such rules to determine when an intrusion is in progress. The main problem with this process is to make good fuzzy classifiers to detect intrusions.

2. INTRUSION DETECTION STRATEGIES

The intrusion detection strategies concerns four primary issues. First is the dataset that is captured from network communications. The second is Genetic Algorithms (GA) which use mutation, recombination, and selection applied to a population of individuals in order to evolve iteratively better and better solutions and a way to generate fuzzy rules to characterize normal and abnormal behavior of network systems. The third is to generate alerts and reports for malicious traffic behavior, and the fourth is the maintenance of the ids for observation of placement of sensors, and qualified trained intrusion analysts so that the latest malicious traffic is being detected.

2.1. The Dataset:

To implement the algorithm and to evaluate the performance of the system, I propose the standard datasets employed in KDD Cup 1999 [27] "Computer Network Intrusion Detection" competition.

The KDD 99 intrusion detection datasets depends on the 1998 DARPA proposal, which offers designers of intrusion detection systems (IDS) with a standard on which to evaluate different methodologies [23]. Hence, a simulation is being prepared from a fabricated military network with three 'target' machines running various services and operating systems. They also applied three extra machines to spoof different IP addresses for generating network traffic.

A connection is a series of TCP packets beginning and ending at some well defined periods, between which data floods from a source IP address to a target IP address under some well defined protocol ([23], [24], [25]). It results in 41 features for each connection.

Finally, there remains a sniffer that accounts all network traffic by means of the TCP dump format [10]. The total simulated period is seven weeks. Normal connections are shaped to outline that expected in a military network and attacks are categorized into one of four types: User to Root; Remote to Local; Denial of Service; and Probe.

The KDD 99 intrusion detection benchmark consists of different components [27]:
kddcup.data; kddcup.data_10_percent;
kddcup.newtestdata_10_percent_unlabeled;
kddcup.testdata.unlabeled;
kddcup.testdata.unlabeled_10_percent; corrected.

I propose to use "kddcup.data_10_percent" as training dataset and "corrected" as testing dataset. In this case the training set consists of 494,021 records among which 97,280 are normal connection records, while the test set contains 311,029 records among which 60,593 are normal connection records. Table 1 shows the intrusion types distribution in the training and the testing datasets.

Table 1. Intrusion types distribution in datasets

Dataset	normal	prob	Dos	u2r	r2l	Total
Train	97280	4107	391458	52	1124	494021
Test ("corrected")	60593	4166	229853	228	16189	311029

2.2 Some Popular Data Mining Methods used in Various Researches:

K-Means: The K-Means algorithm is one of the most popular methods of clustering analysis that aims to partition 'n' data objects into 'k' clusters in which each data object belongs to the cluster with the nearest mean. It uses Euclidean metric as a similarity measure.

The basic algorithm is:

1. Select k objects as initial centroids.
2. Assign each object to the closest centroid.
3. Recalculate the centroid of each cluster.
4. Repeat steps 2 and 3 until centroids do not change.

DOI 10.17148/IARJSET.2016.3432

Important properties of K-Means algorithm:
1. Efficient in processing large data sets.
2. Works only on numerical values.
3. Clusters have convex shapes.

ID3: ID3 (Iterative Dichotomiser 3) invented by Ross Quinlan used to generate a decision tree from a data set and also a precursor to the C4.5 algorithm. It uses the entropy of attributes.

The algorithm can be summarized as:
1. Take all attributes using the data set S and calculate their entropies.
2. By using the attribute which has minimum entropy split the data set S into subsets.
3. Make a decision tree node containing that attribute.
4. By using remaining attributes recurs on subsets.

Important properties of ID3 algorithm:
1. Usually produces small trees.
2. Only one attribute is tested at a time for making a decision.
3. Classifying continuous data may be computationally expensive.

Naive Bayes: Naive Bayes is one of the most efficient learning algorithms. It is based on a strong independence assumption with quite simple construction. It analyzes the correlation between independent variable and dependent variable to obtain a conditional probability for every correlation. By using Bayes theorem we write:

$$P(H|D) = P(D|H)\, P(H) / P(D)$$

Here D may be a data record and H is a hypothesis represents data record D. P (H|D) is the posterior probability of H conditioned on D and P (H) is the prior probability. Similarly, P (D|H) is the posterior probability of D conditioned on H.

Important properties of Naive Bayes algorithm:
1. It is very easy to construct and training is also easy and fast.
2. Highly scalable.

K-NN: K-NN (K-Nearest Neighbor) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. It is a type of instance based learning or lazy learning. It uses Euclidean distance as a distance metric.

The K-NN algorithm for determining the class of a new object C:
1. Calculate the distance between object C and all objects in the training data set.
2. Select K-nearest objects to C in the training data set.
3. Assign C to the most common class among its K-nearest neighbors.

Important properties of K-NN algorithm:
1. It is simple to implement and use.
2. Needs lot of space to store all objects.

2.3 Genetic algorithm

2.3.1. Genetic algorithm overview

A Genetic Algorithm (GA) is a programming technique that uses biological evolution as a problem solving strategy [22]. It is based on Darwinian's theory of evolution and survival of fittest to make effective a population of candidate result near a predefined fitness [15].

The proposed GA based intrusion detection system holds two modules where each acts in a dissimilar stage. In the training stage, a set of classification rules are produced from network audit data using the GA in an offline background. In the intrusion detection phase, the generated rules are employed to classify incoming network connections in the real-time environment. Once the rules are generated, the intrusion detection system becomes simple, experienced and efficient on GA applies an evolution and natural selection that employs a chromosome-like data structure and evolve the chromosomes by means of selection, recombination and mutation operators [15]. The process generally starts with randomly generated population of chromosomes, which signify all possible solution of a problem that are measured candidate solutions. From each chromosome different positions are set as bits, characters or numbers. These positions are regarded as genes. An evaluation function is employed to find the decency of each chromosome according to the required solution; this function is known as "Fitness Function". During the process of evaluation "Crossover" is applied to have natural reproduction and "Mutation" is applied to mutation of species [13]. For survival and combination the selection of chromosomes is partial towards the fittest chromosomes.

When I use GA for solving various problems three factors will have crucial impact on the use of the algorithm and also of the applications [2]. The factors are : i) the fitness function, ii) the representation of individuals, and iii) the genetic algorithm parameters. The determination of these factors often depends on implementation of the system.

2.3.2 Fuzzy logic

Zadeh explained that Fuzzy logic [29] is an extension of Boolean logic that is often used for computer-based complex decision making. While in classical Boolean logic an element can be either a full member or non-member of a Boolean (sometimes called "crisp") set, the membership of an element to a fuzzy set can be any value within the interval [0, 1], allowing also partial membership of an element in a set.

A fuzzy expert system consists of three different types of entities: fuzzy sets, fuzzy variables and fuzzy rules. The membership of a fuzzy variable in a fuzzy set is determined by a function that produces values within the interval [0, 1]. These functions are called membership functions. Fuzzy variables are divided into two groups: antecedent variables, that are assigned with the input data of the fuzzy expert system and consequent variables, that are assigned with the results computed by the system.

The process of a fuzzy system has three steps. These steps are Fuzzification, Rule Evaluation, and Defuzzification. In the fuzzification step, the input crisp values are

transformed into degrees of membership in the fuzzy sets. The degree of membership of each crisp value in each fuzzy set is determined by plugging the value into the membership function associated with the fuzzy set. In the rule evaluation step, each fuzzy rule is assigned with a strength value. The strength is determined by the degrees of memberships of the crisp input values in the fuzzy sets of antecedent part of the fuzzy rule. The defuzzification stage transposes the fuzzy outputs into crisp values.

It has been revealed by Baruah [6] that a fuzzy number [a, b, c] can be explained with reference to a membership function $\mu(x)$ remaining between 0 and 1, $a \leq x \leq c$. Further, he has extended this definition in the following way. Let $\mu1(x)$ and $\mu2(x)$ be two functions, $0 \leq \mu2(x) \leq \mu1(x) \leq 1$. He has concluded $\mu1(x)$ the fuzzy membership function, and $\mu2(x)$ a reference function, such that $(\mu1(x) – \mu2(x))$ is the fuzzy membership value for any x. Finally he has characterized such a fuzzy number by $\{x, \mu1(x), \mu2(x); x \in \Omega\}$.

The complement of μ_x is always counted from the ground level in Zadehian's theory [29], whereas it actually counted from the level if it is not as zero that is the surface value is not always zero. If other than zero, the problem arises and then we have to count the membership value from the surface for the complement of μ_x. Thus I could conclude the following statement –

Complement of $\mu_x = 1$ for the entire level
Membership value for the complement of $\mu_x = 1- \mu_x$

I have forwarded Baruah's definition of complement of an extended fuzzy set where the fuzzy reference function is not always taken as zero. The definition of complement of a fuzzy set recommend by Baruah [28] could be considered a particular case of what I am giving. I would use Baruah's definition of the complement of a normal fuzzy set in my proposed work.

In the two classes' classification problem, two classes are available where every object should be classified. These classes are called positive (abnormal) and negative (normal). The data set employed by the learning algorithms holds a set of objects where each object contains n+1 attributes. The first n attributes identifies the monitored parameters of the object characteristics and the last attribute identifies the class where the object belongs to the classification attribute.

A fuzzy classifier is a set of two rules for solving the two classes' classification problem, one for the normal class and other for the abnormal class, where the conditional part is described by means of only the monitored parameters and the conclusion part is viewed as an atomic expression for the classification attribute.

3. SURVEY WORK

Memon V I, Chandel G S [7] presented work is a grouping of three data mining methods to decrease false alarm rate in IDS that is called a hybrid IDS which has k-Means, K-nearest neighbor and Decision Table Majority method for anomaly detection. Presented hybrid IDS evaluated over the KDD-99 Data set; such type of data set is used worldwide for calculating the performance of various IDS. Initially clustering executed via k-Means over KDD99 data sets then executed two-classification method; KNN followed by DTM. The presented system can detect the intrusions and categorize them into four types: Remote to Local (R2L), Denial of Service (DoS), User to Root (U2R) and Probe.

Wankhade K, Patka S, Thool R [8] presents a hybrid data mining approach encompassing feature selection, filtering, clustering, divide and merge and clustering ensemble. An approach for evaluating the number of the cluster centroid and selecting the suitable early cluster centroid is presented.

Dhakar M, Tiwari A [9], in perspective to enhance performance, the work presents a model for IDS. This improved model, named as REP (Reduced Error Pruning) based IDS Model gives output with greater accuracy along with the augmented number of properly classified instances. It uses the two algorithms of classification approaches namely, K2 (BayesNet) and REP (Decision Tree). Here REP provides an effective classification along with the pruning of tree with quick decision learning capability.

Subramanian P.R and Robinson J.W [10] have discussed on network security through Intrusion Detection Systems (IDSs) with data mining approaches. This model uses binary classifier (C4.5) and multi boosting technique. Here binary classifier is used to classify bit by bit transmission of the packet and used for each type of attack to improve the accuracy and to reduce the variance and bias multi boosting technique is used.

Chandolikar N.S and Nandavadekar V.D [11] presented an approach for intrusion detection using J48 decision tree classifier and also compared with some other tree based algorithms in which J48 tree shows the best performance. To evaluate the performance of the algorithm correctly classified instances, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Root Relative Squared Error and Kappa statistics measures are used.

Barot V and Toshniwal D [12] presented a hybrid model that ensembles Naive Bayes (statistical) and Decision Table Majority (rule based) approaches. Naive Bayes predicts quickly because of less complex functioning of it and processes training data set only once to store statistics. Decision Table Majority (DTM) is a classifier that matches each of the attribute values all together. This model uses sequential reclassification approach for combining rule base classifier. Here correlation based feature selection (CFS) algorithm is used for attribute selection using BestFirst search. Author used KDDCUP'99 data set for their experiment.

Om H and Kundu A [13] presented a hybrid model that combines K-Means and two classifier methods: K-nearest neighbor and Naive Bayes. This model uses entropy based feature selection method for attribute selection. It applies K-Means clustering algorithm for clustering purpose (used number of clusters five) which is followed by K-nearest

neighbor (KNN) and Naïve Bayes classification algorithms for detecting intrusions. The model shows better approach than only K-Means and K-Means, KNN. Author also used the KDD99 cup data set for performing their experiment.

Thakur M R & Sanyal S [14] suggested a multi-dimensional method towards intrusions or attacks detection. Network system usage various parameters like destination and source IP addresses; destination and source ports; outgoing and incoming network traffic information rate and amount of CPU cycles per request are split into numerous dimensions. Observing raw bytes of information corresponding to the values of the network factors, an established function is inferred throughout the training phase for every measurement. This grown-up function takes a measurement value as an input and returns a value that represents the level of anomaly in the system usage relating to that dimension.

This mature function is referred as *entity Anomaly pointer*. *Entity Anomaly pointer* recorded for every of the measurement are then used to produced a Universal *Anomaly Pointer*, a function with n variables (n is the number of dimensions) that provides the U*niversal Anomaly Factor*, a pointer of anomaly over the system usage based on all the measurements measured together. The U*niversal Anomaly pointer* inferred throughout the training phase is then used to find out anomaly over the network traffic throughout the detection phase. Network traffic data encountered through the detection phase is fed back to the system to develop the maturity of the *Entity Anomaly Pointers* and hence the U*niversal Anomaly Pointer.*

Pathak V and Ananthanarayana V. S [15] have suggested a multi-threaded K-Means clustering approach. In this approach they have used six threads which run in parallel. Out of which five threads are used to cluster the data and the last sixth thread is used to take decision classify the data. Out of five threads, each is used to identify particular type of attack and normal or abnormal data. Author used KDD99 training data set for their experiment. Proposed approach i.e. multi-threaded K-Means gives better result in comparison to K-Means.

Wang P and Wang J Q [16] discussed about data mining which is popularly known as an important way to mine useful information from large volumes of data which is noisy, fuzzy, and random. In this, present the whole techniques of the IDS along with data mining method in details. Author mainly discussed about three data mining based approaches: Classification, Association and Sequence rules. Also discussed the system architecture of the IDS.

Muda Z, Yassin W, Sulaiman M.N and Udzir N.I [17] describe a hybrid learning approach that combines K-Means clustering method and Naive Bayes classification method. In the proposed approach, firstly K-Means (as a pre-classification) cluster all the data in to corresponding groups and then Naive Bayes classifier is used to classify the resultant clusters into attack classes as a final task. Because of this, the data that has been misclassified in the earlier stage (K-Means) may be classified correctly in the consequent classification task (Naive Bayes). Here author took number of clusters (K) = 3 and KDD Cup 99 benchmark data set for evaluating the performance of their approach.

Dewan Md. Farid, Nouria Harbi [18] offered a learning algorithm for adaptive network intrusion detection using Naive Bayesian classifier and ID3 algorithm which performs good detections and keeps less false positives and also eliminates redundant attributes in addition to contradictory examples from training data set that make complex detection model. Author also addresses some difficulties of data mining such as handling continuous attribute, missing attribute values and reducing noise in training data. This model used Knowledge Discovery Data Mining (KDD) CUP 99 dataset for experiment.

Bharti K K, Shukla S and Jain S [8], a number of techniques are available for IDS. Data mining methods are the proficient methods available for IDS. Data mining techniques may be supervised or unsupervised. Various researchers have applied various clustering algorithm for IDS, but all of these are trouble from class ascendancy, force assignment and No Class problem. This work proposed a model that is based on feature selection (as a first phase), K-Means clustering model generation (as a second phase) and classification (as a third phase). This model used CfsSubSetEval method along with BestFirst search for feature selection. In the final phase of the proposed model i.e. classification phase, author used J48 and Random Forest. To evaluate the performances of proposed model KDD Cup 1999 dataset is used. This model used precision and recall as a performance metric.

Panda M, Patra M R [6] evaluated the performance of different rule based classifiers like: NNge (Non-Nested Generalized Exemplars), RIDOR (Ripple-Down Rules), JRip (Extended Repeated Incremental Pruning) and Decision Table using ensemble methods in order to make a proficient network IDS, by combining AdaBoost with different base learners. This model used KDD Cup 99 data set for performing experiment.

In this section, I describe the important and relevant research works of different authors that I have come across during the literature survey of my proposed work. I illustrate each attack manner and point to the impact of this attack and its intrusive activities. From an intruder's point of view, I analyze each of the attack's modes, intention, benefits and suitable conditions and try to find out the solution how to improve the attack by introducing the concept of fuzzy logic-based technique and genetic algorithm.

The normal and abnormal behaviors [1] in networked computers are hard to forecast, as the limits cannot be explained clearly. This prediction method usually generates fake alarms in many anomaly based intrusion detection systems. In [1] the authors introduced the concept of fuzzy logic to reduce the fake alarm rate in determining intrusive behavior. The set of fuzzy rules is

applied to identify the normal and abnormal behavior in a computer network. The authors proposed a technique to generate fuzzy rules that are able to detect malicious activities and some specific intrusions. This system presented a novel approach for the presentation of generated fuzzy rules in classifying different types of intrusions.

The advantage of their proposed mechanism is that the fuzzy rules are able to detect the malicious activities. But they failed to implement the real time network traffic, more attributes for the classification rules. In determining the fuzzy rules, they used the concept of fuzzy membership function and reference function, but they said that the membership function and reference function are same. In reality, these two concepts are totally different concepts. I have forwarded the extended definition of fuzzy set of Baruah [28].

In [29] Zadeh initiated the idea of fuzzy set theory and it was mainly intended mathematically to signify uncertainty and vagueness with formalized logical tools for dealing with the vagueness connected in many real world problems. The membership value to a fuzzy set of an element describes a function called membership function where the universe of discourse is the domain and the interval lies in the range [0, 1]. The value 0 means that the element is not a member of the fuzzy set; the value 1 means that the element is fully a member of the fuzzy set. The values that remain between 0 and 1 distinguish fuzzy members, which confined to the fuzzy set merely partially. But the author gave an explanation that the fuzzy membership value and fuzzy membership function for the complement of a fuzzy set are same concepts and the surface value is always counted from the ground level.

Baruah [28] have forwarded an extended definition of fuzzy set which enables us to define the complement of a fuzzy set. My proposed system agrees with them as this new definition satisfies all the properties regarding the complement of a fuzzy set.

In [2] Gong, Zulkernine, Abolmaesumi gave an implementation of genetic based approach to Network Intrusion Detection using genetic algorithm and showed software implementation to detect the malicious activities. The approach derived a set of classification rules from network audit data and utilizes a support-confidence framework to judge the quality of each rule. The generated rules are then used in intrusion detection system to detect and to classify network intrusions efficiently in a real-time environment.

But, some limitations of their implemented method are observed. First, the generated rules were partial to the training dataset. Second, though the support-confidence framework is simple to implement and provides improved accuracy to final rules, it requires the whole training datasets to be loaded into memory before any computation. For large training datasets, it is neither efficient nor feasible. But their performance of detection rate was poor and they failed to reduce the false positive rate in Intrusion Detection System.

4. DISCUSSION

This section discusses the strengths and weakness of various existing methods:

An intrusion detection system based on genetic algorithm approach has proposed by author [19]. This approach showed a reasonable detection rate but can be improved by using better equations or heuristic in the detection process.

A hybrid learning approach [18] that combines Naive Bayesian classifier and ID3 algorithm overcomes the problem of moderate detection rate and false positives but needs improvement for false positives in remote to user (R2L) attack.

A Y-means clustering algorithm [20] has improved detection rate and low false alarm rate. But it cannot solve the real time anomaly detection, because it cannot revise the data set dynamically during the process.

A clustering based algorithm uses SOM and K-Means [21] overcomes the drawback of traditional SOM which cannot give the precise clustering results, and it also overcomes the drawback of traditional K-Means that depends on the initial value and it is also hard to locate a suitable center of the cluster.

A parallel clustering ensemble algorithm [22] forms the clusters more rapidly to mass data. It also achieves high detection rate but false alarm rate. Because it is an ensemble approach it requires extra memory.

A modified dynamic K-Means algorithm called MDKM [23] has fine detection rate but its false alarm rate is moderate.

A hybrid learning approach [17] that uses K-Means clustering and Naive Bayes classification overcomes the problem of moderate detection rate and high false alarm rate of existing methods. A hybrid intrusion detection system [13] that combines K-Means and two classifiers: k-nearest neighbor and naive bayes overcome the shortcoming of high false alarm rate in existing method.

Intrusion detection system that combines K-Means, fuzzy neural network and SVM classifiers [24] and by utilizing data mining methods such as neuro-fuzzy and radial basis support vector machine (SVM) for helping IDS to achieve a better detection rate. Back-propagation neural network based IDS [25] requires a very large amount of data and takes time to ensure the results accuracy.

Boosted decision tree approach [10] for intrusion detection system is an ensemble approach and its detection rate is fine but has moderate false alarm rate. Because it combines a number of decision trees, it becomes complex and needs more time and space.

5. CONCLUSION

In this paper, I have described an overview of some of the current and past intrusion detection technologies which are being utilized for the detection of intrusive activities against computer systems or networks. The different detection challenges that affect the decision policy of the

IDS employed in an organization are clearly outlined. To use the new definition of the complement of fuzzy sets which can classify efficient rule sets have been proposed. This would help in reducing the false alarm rate occurred in intrusion detection system. Since the study of intrusion detection began to gain momentum in the security community roughly ten years ago, a number of diverse ideas have emerged for confronting this problem. Intrusion detection systems vary in the sources they use to obtain data and in the specific techniques they employ to analyze this data. Most systems today classify data either by misuse detection or anomaly detection: each approach has its relative merits and is accompanied by a set of limitations. It is likely not realistic to expect that an intrusion detection system be capable of correctly classifying every event that occurs on a given system. Perfect detection, like perfect security, is simply not an attainable goal given the complexity and rapid evolution of modern systems. An IDS can, still endeavor to hoist the bar for intruder or attackers by dipping the efficacy of big classes of intrusion or attacks and rising the work issue required to get a system compromise. A good intrusion detection system promises to allow greater confidence in the results of and to improve the coverage of intrusion detection, making this a critical component of any comprehensive security architecture.

REFERENCES

[1]. J. Gomez & D. Dasgupta, (2002) "Evolving Fuzzy Classifiers for Intrusion Detection", IEEE Proceedings of the IEEE Workshop on Information Assurance, United States Military Academy, West Point, NY.

[2] R. H. Gong, M. Zulkernine & P. Abolmaesumi, (2005) "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection", Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Networks.

[3]. T. Xia, G. Qu, S. Hariri & M. Yousif, (2005) "An Efficient Network Intrusion Detection Method Based on Information Theory and Genetic Algorithm", Proceedings of the 24th IEEE International Performance Computing and Communications Conference, Phoenix, AZ, USA.

[4] K. Jungwon, J. B. Peter, A. Uwe, G. Julie, T. Gianni and T. Jamie, "Immune System Approaches to Intrusion Detection – A Review", Natural Computing: an international journal, vol. 6, Issue 4, (2007) December.

[5] E. J. Derrick, R. W. Tibbs and L. L. Reynolds, "Investigating New Approaches to Data Collection, Management and Analysis for Network Intrusion Detection", ACMSE, Winston-Salem, N. Carolina, USA, (2007) March 23-24, pp. 283-287.

[6] M. Panda and M. R. Patra, "Ensembling Rule Based Classifiers for Detecting Network Intrusions", IEEE International Conference on Advances in Recent Technologies in Communication and Computing, (2009), pp. 19-22.

[7] V. I. Memon and G. S. Chandel, "A Design and Implementation of New Hybrid System for Anomaly Intrusion Detection System to Improve Efficiency", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, vol. 4, Issue 5, (Version 1), (2014) May, pp. 01-07.

[8] K. Wankhade, S. Patka and R. Thool, "An efficient approach for Intrusion Detection using data mining methods", International Conference on Advances in Computing, Communications and Informatics (ICACCI), Print ISBN:978-1-4799-2432-5 INSPEC Accession no. 13861274, (2013) August 22-25, pp. 1615-1618.

[9] M. Dhakar and A. Tiwari, "A New Model for Intrusion Detection based on Reduced Error Pruning Technique" International Journal of Computer Network and Information Security, (2013), pp. 51-57.

[10] P. R. Subramanian and J. W. Robinson, "Alert over the attacks of data packet and detect the intruders", Computing, Electronics and Electrical Technologies (ICCEET), IEEE International Conference on ISBN: 978-1-4673-0211-1, (2012) March 21-22, pp. 1028-1031.

[11] N. S. Chandolikar and V. D. Nandavadekar, "Efficient algorithm for intrusion attack classification by analyzing KDD Cup 99", Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on ISSN :2151-7681, (2012) September 20-22, pp. 1 - 5.

[12] V. Barot and D. Toshniwal, "A New Data Mining Based Hybrid Network Intrusion Detection Model" IEEE International Conference on Print ISBN: 978-1-4673-2148-8, (2012) July 18-20.

[13] H. Om and A. Kundu, "A hybrid system for reducing the false alarm rate of anomaly intrusion detection system", Recent Advances in Information Technology (RAIT), IEEE International Conference on Print ISBN:978-1-4577-0694-3, (2012) March 15-17, pp. 131-136.

[14] M. R. Thakur and S. Sanyal, "A Multi-Dimensional approach towards Intrusion Detection System" International Journal of Computer Applications, vol. 48, no. 5, (2012) June, pp. 34-41.

[15] V. Pathak and V. S. Ananthanarayana, "A novel Multi-Threaded K-Means clustering approach for intrusion detection" Software Engineering and Service Science (ICSESS), IEEE 3rd International Conference on Print ISBN: 978-1-4673-2007-8, (2012) June 22-24, pp. 757-760.

[16] P. Wang and J. Q. Wang, "Intrusion Detection System with the Data Mining Technologies" IEEE 3rd International Conference on Print ISBN: 978-1-61284-485-5, (2011) May.

[17] Z. Muda, W. Yassin, M. N. Sulaiman and N. I. Udzir, "Intrusion Detection based on K-Means Clustering and Naive Bayes Classification", 7th IEEE International Conference on IT in Asia (CITA)(2011).

[18] D. Md. Farid, N. Harbi and M. Z. Rahman, "Combining naive bayes and decision tree for adaptive intrusion detection", International Journal of Network Security & Its Applications (IJNSA), vol. 2, no. 2, (2010) April.

[19] M. S. Hoque, Md. A. Mukit and Md. A. N. Bikas, "An implementation of intrusion detection system using genetic algorithm", International Journal of Network Security & Its Applications (IJNSA), vol. 4, no. 2, (2012) March.

[20] Y. Guan and A. A. Ghorbani and N. Belacel, "Y-Means: A Clustering Method For Intrusion Detection", In Proceedings of Canadian Conference on Electrical and Computer Engineering, Montreal, Quebec, Canada, IEEE, (2003) May 4-7, pp. 1083-1086.

[21] W. Huai-Bin, Y. Hong-Liang, X. Zhi-Jian and Y. Zheng, "A clustering algorithm use SOM and K-Means in Intrusion Detection", International Conference on E-Business and E-Government, IEEE, (2010), pp. 1281-1284.

[22] H. Gao, D. Zhu and X. Wang, "A Parallel Clustering Ensemble Algorithm for Intrusion Detection System", Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, IEEE, (2010), pp. 450-453.

[23] L. Han, "Using a Dynamic K-means Algorithm to Detect Anomaly Activities", Seventh International Conference on Computational Intelligence and Security, IEEE, (2011), pp. 1049-1052.

[24] A. M Chandrashekhar and K. Raghuveer, "Intrusion Detection Technique by using K-means, Fuzzy Neural Network and SVM classifiers", International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, INDIA, (2013) January 4-6.

[25] S. T. F. Al-Janabi and H. A. Saeed, "A Neural Network Based Anomaly Intrusion Detection System", Developments in E-Systems Engineering, IEEE, (2011), pp. 221-226.

[26] C. F. Tsai and C. Y Lin, "A triangle area-based nearest neighbors approach to intrusion detection" Pattern Recognition, vol. 43, no. 1, (2010), pp. 222-229.

[27] KDD Cup (1999): Data; http://www.kdd.org/kddcup/index.php?section=1999&method=data

[28] Hemanta K. Baruah, (2011) "Towards Forming A Field Of Fuzzy Sets", International Journal of Energy, Information and Communications (IJEIC), Vol. 2, Issue 1, February, pp. 16-20.

[29] Zadeh L A, (1965) "Fuzzy Sets", Information and Control, Vol.8, pp. 338-353.

[30] KDD-CUP (99) Task Description; http://kdd.ics.uci.edu/databases/kddcup99/task.html