Network Intrusion Detection using Selected Data Mining Approaches: A Review

Munawara Saiyara Munia
Department of CSE
Ahsanullah University of Science and Technology
Dhaka-1208, Bangladesh

Samira Samrose
Department of CSE
Ahsanullah University of Science and Technology
Dhaka-1208, Bangladesh

Pranab Dey
Department of CSE
Ahsanullah University of Science and Technology
Dhaka-1208, Bangladesh

Afsana Salauddin Annesha
Department of CSE
Ahsanullah University of Science & Technology

Syeda Shabnam Hasan
Department of CSE
Ahsanullah University of Science & Technology

ABSTRACT
Due to the rapid progress in network technologies, easy availability of the internet and lower cost of mobile devices with wireless network connection facility, the number of internet users is increasing at an exponential rate now-a-days, so does the number of intrusion. Despite the implausible advancement in Information Technology, Intrusion Detection has remained as one of the biggest challenges encountered by network security specialists. Data mining can play a vital role in addressing this issue. In this paper, some selected data mining algorithms available for Network Intrusion Detection have been reviewed, such as- Support Vector Machine, K-Nearest Neighbor, Naïve Bayesian Classifier, Decision tree Algorithm (C4.5), Genetic Algorithm, Logistic Regression, Artificial Neural network, K-means clustering, EM algorithm, Fuzzy Logic and Hidden Markov Chain; along with addressing the advantages and disadvantages of each of them.

General Terms
Network Security, Data Mining Algorithms

Keywords
Intrusion Detection, Data mining, Neural Networks, Fuzzy Logic, Support Vector Machine, Network Security, Naïve Bayes classifier, Genetic Algorithm, K-Nearest Neighbor, Logistic Regression, K-means clustering, The EM algorithm, Decision trees, C4.5,Hidden Markov Chain.

1. INTRODUCTION
The universal acceptance of wireless technology has created the computer networks idea to be re-sharped. As a result, in the past few years, lots of networking architectures have been refined. But like other systems network architecture also have some vulnerable sections. Potential cyber threats like network intrusion generate the breaking of a moral protection. So, the main idea of developing intrusion detection system (IDS), firewall, authentication, encryption and other hardware, software based programs is to protect the precious information or data from the unauthorized users or intruders. But due to the increasing growth of attacking tools and tricks which support the network attack, effective method for intrusion detection has become the highest priority to protect the network.

2. INTRUSION DETECTION TERMINOLOGY
Intrusion detection system is established to prevent the intruders and ensure the safety of the system. So the study of intrusion detection terminology and classification is necessary.

2.1 Intrusion Detection System Classification
Based on data sources, Intrusion detection system can be classified as following:

Host-based detection: The network architecture of host-based intrusion detection system is agent-based. It works for individual host or device in a network. It observes the activity and raises an alarm to the human security officer or administrator to investigate if any change or modification is found on the system file.

Network-based detection: Network-based intrusion detection is used to observe data exchange between computers. It observes the traffic movement on the network from a strategic point or points. It analyzes the passing traffic on the whole subnet and checks if there is any similarity among the passed traffic and the database of known attacks.

2.2 Intrusion Detection Approaches
The approaches to detect intruder in system is done with many possible ways. But there are two major approaches to detect the type of intruder and their drawbacks are discussed in this section.

2.2.1 Anomaly-based or profile-based detection
In anomaly-based detection, user's usual behavior is stored is database, then comparing user's current behavior with those stored in the database. If any kind of divergence is found, it is said an intrusion has detected. It can detect an already known attack or an attack that was detected never before. Anomaly based intrusion detection is about discrimination of malicious and legitimate patterns of activities in variable characterizing system normality.

Draw Back of Anomaly-based Detection: Though it can detect new attacks, but at the same time it provides false positives results. In addition, selection of the correct set of system features to measure is ad hoc and based on experience, which is quite difficult to perform.

2.2.2 Misuse-based or signature-based detection:

In misuse detection approach, previously known intrusions are acknowledged and hand-coded. There is a human security analyst which takes appropriate decision based on their experience in identifying intrusions.The system searchs for pattern or significant user behavior in the network traffic. Any match with the signature is count as possible attack or threat. Misuse detection has are based on expert system. In general it can be said that misuse detection use pattern of well-known attack to identify intrusions.

Drawbacks of Misuse-based detection: The intrusions which are saved in the database have to be manually coded by experts. Also the database needs to be updated when any new intrusion is detected.

2.2.3 Combination of anomaly and misused intrusion detection:

Anomaly and Misuse based detection systems have paucity that hampers their efficiency of detecting intrusion. The combined techniques refer to use the benefits of both approaches. External and internal attacks can be easily monitored by single IDS. While it is possible to achieve significant advantages by combining both methods rather than using any single one separately, the combined approach has some serious disadvantages. More dedicated system resources, increase in memory requirements are some of its disadvantages.

3. TYPES OF ATTACK

There are some basic types of attacks that are recognized. The types can be modified and applied to crash the system but the basic remains same.

DoS attack: A denial of service is a process to hide computer resources from its intended user. It generally resists a network site or service to perform its tasks efficiently or properly.

User to root: It allows the unauthorized access to local super user. Attackers can access normal user account on the host system and can use vulnerabilities to get root access of the system.

Probing: An attacker uses the services which are provided by that site and the machine map for exploits. It scans the network to detect vulnerabilities or gather information. It requires a very little technical support to perform.

Remote to user (U2R): This type of attack describes the unauthorized access from a remote machine into the super user account of the target system. A packet is send to the machine over the network by the attacker to get illegal access of data as a user.

Eavesdropping attack: It illegally captures the packets transmitted by other's computer and accesses the sensitive information such as session token, password, or any sort of private data.

Man-in-the-middle-attack: The attackers independently connect with the victims and make them to believe that their conversation is kept as private although the entire conversation is controlled by the attackers.

4. INTRUSION DETECTION SYSTEM

The major steps of the IDS archicture and the procedure of it will be described in this section.

4.1 Intrusion Detection System archicture

IDS can be recited as an indicator which processes data coming from the system to make it more secured. The main task of the indicator is to delete the undesirable information from the audit data. Then it presents an artificial view of security related action which is taken during normal system usage. Then a decision is taken to decide the action or the state whether carries the symptoms of being an intrusion or not. Countermeasure components can then take effective step to either resist the action from being performed or allow the state of the system to the protected state.

4.2 Working Phases of Intrusion Analysis

Intrusion analysis process can be divided into following four phases:

1. Processing phase: It is the first phase of IDS. Major step of this phase is to classify data. So, data are organized in some pattern. The classification depends on the analysis schemas being used. The phase determines the format of the data. The format might be canonical or structured database. This phase collects the behavior from IDS.

2. Analysis: It is the second phase of IDS. After preprocessing, data need to compare with knowledge base which results will be taken as an intrusion or threat, else it will be discharged to examine the next data.

3. Response: Response can be set to be performed automatically. In case of manual analysis situations it can be done manually.

4. Refinement: This phase does tuning task based on previously detected intrusions. It can reduce more false positive levels. It supports to have more security tools which ensure that the raised alert is valid or not.

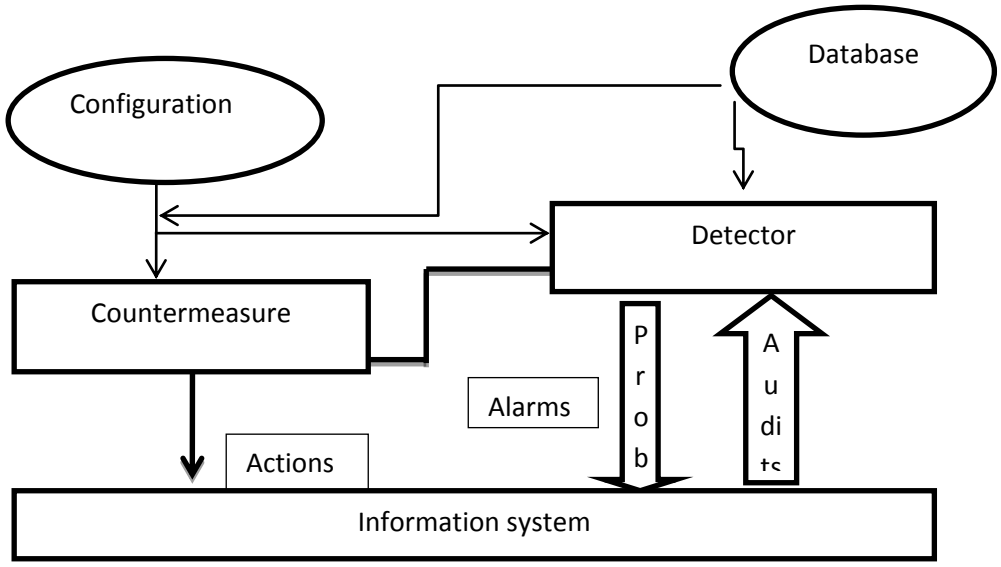

Fig 1: Simple IDS

5. IDS ALGORITHAMS

Intrusion detection system follows some algorithms to detect the abnormality of the system. Some of the most used and effective algorithms are discussed here.

5.1 Fuzzy Logic

Fuzzy logic is proposed as a model of uncertainty of natural language. The fuzziness of fuzzy logic is used to clear the unexpected partition of normality and abnormality. [4] In the field of intrusion detection fuzzy logic has demonstrated potentially applicable. [5] This approach is certainly be declared as the 'degree of truth' rather than 'true or false'.

Fuzzy logic is obtained from fuzzy set theory and for classification it applies rule based system which has different conceptual components. [6] Different rules are used for different pattern classification problems. It is very suitable for applying on the intrusion detection. Fuzzy association rules are applies to detect abnormality and to mine the network audit data model. [6]

Giving warning for every intrusion steps is bothersome. On the basis of individual conditions warning should be raised at the degree of intrusion. [4]Comparing the audit data and mined normal data warning can be raised.

Fuzzy membership function defines the fuzzy set for the possible values. A membership function μ(x) with fuzzy number [a,b,c] is between 0 and 1,$a \le x \le c$. Extensionally, fuzzy number is $\{x, \mu1(x),\mu2(x); x \epsilon \Omega\}$ where μ1,μ2 is fuzzy membership and reference function respectively and $0 \le \mu2(x) \le \mu1(x) \le 1$. Membership function is $(\mu1(x)-\mu2(x))$ [7, 8].

Advantages:

- Permits partial membership of a component of a set.
- Link recognition between predecessor and variable is very flexible.

Disadvantages:

- Providing frequent warning.

5.2 Support Vector Machine

Support vector machine is a sort of supervised and statistical learning technique. [9, 10] Support vector machine can be implemented in classification, regression and tagging. [10] It is basically a binary classification model and linear classifier model. This technique is established on structural risk minimization principle to detect the lowest probability of error [11].

SVM do not need any feature reduction to ignore over fitting. The number of parameter depends on the margin that separated the data point. [9] So it can be defined in the feature space with maximal margin.

The main support vector machine is based on three elements. Those are model, strategy and algorithm. [10] Support vector machine has both linear and non-liner realizations. Liner separable SMV shows the highest margin between two supporting hyper-planes. And linear SMV shows the highest soft margin. And using kernel method support vector machine can be described as non-linear terms. It shows the turning of input space $x \epsilon R^n$ to a Hilbert space $x \epsilon H$. And also can be seen as the highest soft margin for non-linear SVM. [10]

Determining the hyper plane to differentiate the data point support vector machine is advance to the squared optimization problem. [11, 12]Actually linear input space can never handle all problems. So Kernel method is introduced to perform this type of problems. During the training process, a function to the SVMs can allow the surface to the function. Total number of free parameters which used in SVMs depends on the margin that differentiates data points. [6]

Support vector machine input function, [13]

$S = \{(x_1, y_1), (x_2, y_2),\ldots\ldots, (x_l, y_l)\}$ where $x_i \epsilon R^n$ and $i=1,2,\ldots,l$ is a set of input and classifies into target class, $y_i \epsilon \{\pm 1\}$.

For optimal hyperplane,

$\text{Min } \emptyset(w,\varepsilon_i) = \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{l} \varepsilon_i$ S.T. $y_i(w.\emptyset(x_i) + b) \ge 1 - \varepsilon_i$

Here, $\varepsilon_i \ge 0$.

C is a regularization parameter. [9]

Deriving the equation or using basic model genetic algorithm, genetic algorithm with decision tree, fuzzy support vector machine can be applied to support vector machine.

Advantages:

- Efficiently produce unique solution within a very short time. Can handle more abnormality.
- By introducing Kernel method SVM Recognizes large scale of patterns and update it automatically.

Disadvantages:

- SVM is a procedure for powerful 2-class but poor example of multi-class classification.
- Hard to understand functions.

5.3 Markov Chain

Markov model is sub divided into two types: Markov chain and hidden Markov model. [14, 15] Hidden Markov model is a state full model which indicates that it is a finite set of states. Each state is associated with a probability distribution. States transition with each other is generated by a set of probabilities called state transition probabilities. The states are not observable to the external users but the outcome is observable to the user. Hence, the model is defined as hidden Markov model. [16]

By discriminating between normal and abnormal conditions HMM based classifiers are very effective in detecting network based intrusion detection system. As the Markov Model applied in the hidden layer is a first order Markov model which explains the dependency of a particular layer with the previous layer. In a particular state, Markov model concludes the visible corresponding states. [16, 17, 18]

Hidden Markov model is a random, stochastic and probabilistic model. [14, 9] It is also a basic probabilistic model which is visible through another set of random or stochastic process. Set of random or stochastic process means the process which provides the order of visible symbols. [17] HMM can be discrete if the output is finite and continuous in case of being continuous. [18] This method is highly applicable in user activity based system.

The appearance of high accuracy in detecting intrusions the hidden Markov model proved itself to be a favorable model. But during the establishment of this model appears worst in quality because is of the long training time. Many approaches can be done like speech recognition, synthesis and sequencing, bioinformatics, compute vision, communication and control, crypt analysis.

Advantages:

- Performance is in high level.
- More security guaranteed model then others.

Disadvantages:

- Never stores the previous training phase.
- Do not have memory realization.

5.4 K-Means Clustering Approach

[19] K-Means Clustering, which is a hard partitioned clustering algorithm, is the simplest and most popular clustering technique due to its simplicity and speed. Its' job is to allocate 'm' data objects into 'k' clusters and uses Euclidean distance metric for measuring the similarity.

According to the algorithm, the number of clusters, k is defined at first. For example, if k=3, there will be three clusters for the training data. Then, k objects are chosen as preliminary cluster centroids. To achieve this, K data objects are selected arbitrarily from dataset. Then each data item is assigned to a cluster by computing the distance between the data item and each cluster centroid and allocating it to the nearest cluster. In this way, the intra-cluster similarity remains high and inter-cluster similarity goes low [21]. Then the averages of all clusters is recalculated and renewed and the afresh computed mean is assigned as the new centroid. This process is replicated until the criterion function is converged.

Accuracy of k-means clustering depends upon the value of k. The data may also contain outliers. But if the number of outlier is comparatively smaller, the K-means clustering process is not hindered. By calculating the GSD (Generalized Squared Distance) and inspecting them for remarkably outsized values, outliers can be determined in intrusion detection [20]. The algorithm is basically specialized to train datasets which are not labeled yet. These data may contain both normal and anomalous traffic. K-means clustering performs the best when the data set is separated well enough from each other [21].

Advantages:

- Easy implementation, high accuracy and faster execution when it is applied on small dataset.
- Works only on numerical values.

Disadvantages:

- Determining the appropriate number of clusters is challenging area for researchers. Sensitive to outlier and noise.

5.5 Artificial Neural Network

Artificial Neural Network is a model simulation of the neurons in the human brain to imitate human brain functionality. It is a collection of several numbers of highly interconnected nodes (neurons) which has a weighted connection to several other nodes in adjacent layers. The output of each neuron is provided as the input to all of the neurons in the following layers and uses the weights together with a simple function [22] to compute output values. The system is adaptive in nature. During the learning phase, it can make changes in its configuration to adapt with the external or internal data flowing over the network. It plays a very remarkable role in both anomaly and misuse intrusion detection.

In anomaly detection, it learns the usual characteristics of system users by learning the typical sequence of commands executed by each user to identify statistically noteworthy deviations from user's recognized behavior. Explicit user model is not required. In misuse intrusion detection, data is received from network traffic and the information is analyzed for occurrences of misuse by the neural network. Neural network can be used as a filter for the incoming data to check for malicious events and send them to an available expert system for further actions. Or it can also be used as a standalone misuse detection system.

Implementing a neural network based Intrusion Detection System consists of the following phases [23]:

- Training data is collected by attaining the audit log for every user for a definite period of time.
- Next the user is identified by training the neural network, depending on the command distribution vector.

- Last the job of neural network is to check (based on command distribution vector) if the network behavior shows deviation from the real user behavior. If so, an anomaly is detected.

Advantages:

- Remarkable tolerance to noisy data.
- Multiple training algorithms available.

Disadvantages:

- Process is black box.
- Long training time.

5.6 Logistic Regression

Logistic [24] is a function based classifier which uses ridge

Estimator and well suited to two-class classification problems. This technique employs regression but with a binary response. That is, an outcome can have only two values → Intruder or Authorized User (I or A).Probability estimation is based on one or more predictors. Logistic regression calculates the relationship between the dependent variable (which is categorical in nature) and one or more independent variables (or predictors) by appraising their probabilities using a logit function [25], which is as follows:

$$log\left(\frac{Pr(Y=1|X=x)}{Pr(Y=0|X=x)}\right) = \beta^T X = \sum_{i=1}^{p} \beta i xi$$

It generates a logistic curve with values ranging between 0 and 1. The curve is not constructed using probability. Rather it is based on the natural log values of the odds of the target. The odds [26] are described as the probability that a particular consequence is a case, which is divided by the probability that the consequence is a non-case.

Advantages:

- Relative simplicity of implementation
- Allows thresholding of the response.

Disadvantages:

- Limited Outcome Variables.
- Independent Observations Required.

5.7 Naïve Bayes classifier

[27] Naïve Bayes is one of the most proficient learning algorithms with a very simple construction model. It plays a vital role in intrusion detection, combining with statistical approaches, to produce interdependencies between the variables, based on the implications of a probabilistic graphic model (PGM). This graphical model does the learning job. A probabilistic graphical model is a graph, with nodes representing random variables (which may be discrete or continuous), and the edges representing conditional interdependency assumptions. Hence it represents a joint probability distribution. A conditional probability table is maintained for every variable, by the classifier, for which higher computational effort is required. Naive Bayesian classifiers use the Bayes theorem to classify the test data which is given below:

$$P(A_i/B)=(P(B/A_i) \,.\, P(A_i))/P(B)$$

Where, P (Ai) is the Prior probability of class Ai, P(B) is the Prior probability of predictor, P(B/Ai) is the probability of predictor, given class, P (Ai/B) is the posterior probability of class, given predictor.

Naïve Bayes classifier is most likely used in anomaly based intrusion detection. It works as follows: [28]

Step1: Let T be a training set of tuples, represented by n dimensional vector, B=(b1,b2,….,bn), associated with their class labels.

Step 2: Let us assume that m classes are available - A1,A2….Am.

Given tuple B, naïve Bayes classifier forecasts that B belongs to the class Ai if and only if :

$$P(Ai/B) > P(Aj/B); \text{ for } 1 \le j \le m, j \ne i$$

Thus P(Ai/B) is maximized.

Step 3: P(B) is persistent for all the classes, only P(B/Ai). P(Ai) needs maximization. If the prior probabilities of the classes are not identified, then it is usually anticipated that the probability of the classes are all equal, that is, P(A1) = P(A2) = …… = P(Am), and which therefore maximizes P(B|Ai). Otherwise, P(B|Ai)P(Ai) is maximized.

Step 4: For a data set with many attributes, the computational expense will become very large to compute P(B|Ai). For this, the naïve assumption of class conditional independence is introduced. This presumes that if the class label of the tuple is given, the values of the attributes are conditionally independent of one another. Thus,

$$P\left(\frac{B}{Ai}\right) = \prod_{k-1}^{n} P\left(\frac{Bk}{Ai}\right) = P\left(\frac{b1}{Ai}\right) * P\left(\frac{b2}{Ai}\right) * \dots\dots\dots * P\left(\frac{bn}{Ai}\right)$$

Advantages:

- Easy construction, no complicated iterative parameter, fast and easy training, robust against noise.
- Naïve Bayesian classifier simplifies the computations. Highly scalable.

Disadvantages:

- Lack of available probability data [29].
- In naïve Bayes approach, it is assumed that the data attributes are conditionally autonomous, which is not right all the time [29].

5.8 k-Nearest Neighbor

[30] The k-Nearest Neighbor algorithm (k-NN algorithm) is the simplest, nonparametric classification techniques. It is a simple algorithm that stores all available cases (training examples) and classifies new cases (test examples) based on a similarity measure in the feature space. It uses Euclidean distance as a distance metric. The training examples are associated with a class labels. In the training phase, only the feature vectors (which defined the training examples) and class labels of the training data are stored. Similarity based search is used in this algorithm to discover the optimal hypothesis function.

An object is categorized by the majority vote of its neighbors. For classification, it computes the approximate distances between the new case (an unlabeled vector) and different

points on the input vectors and then the test data is allotted to the class most common amongst its k nearest neighbors. If k=1, then the object, without any doubt, is assigned to the class of its nearest neighbor. Large prediction time is needed I the value of K is large [31]. It is kind of a lazy learning in which the function is only estimated locally. All calculation is also postponed until the classification is done.

Advantages:

- Simple implementation, use of local information, highly adaptive behavior.
- Very comfortable in parallel implementations.

Disadvantages:

- Requires large and efficient storage for implementation of parallel hardware. Hence expensive.
- All the training samples are evaluated for classification and the performance is entirely dependent on the training set, which results in a very complex calculation.

5.9 EM Clustering

[32] The EM (Expectation-Maximization) algorithm is an extension of K-Means in which the job is to assign an object to the cluster to which it is similar. This clustering is done by calculating the mean of cluster. It is a refinement algorithm that is recursive in nature and can be beneficial for finding the estimation of parameter. It is particularly used when the label of an object is missing.

Basic approach to clustering is extended by the EM algorithm in the following two ways:

- Rather than assigning objects to clusters for maximizing the variances in means, the algorithm calculates the probabilities of cluster memberships. Then the clustering algorithm maximizes the overall probability, given the clusters. [33]
- Contrasting with the general application of k-means algorithm, EM algorithm can work on both continuous and definite variables.

The EM algorithm mainly deals with the missing labels by alternating between two steps:

1. Expectation step (E step): Fix model, estimate missing labels.

2. Maximization step (M step): Fix missing labels (or a distribution over the missing labels) and find new parameters for the model to maximize the expected log-probability of the data.

Advantages:

- The ability to simultaneously optimize a large number of variables
- The ability to find good estimates for any missing information in your data at the same time

Disadvantages:

- Slow convergence

5.10 Genetic Algorithm

[1, 2, 3] Genetic algorithm is mainly based on the principles of evolution and natural selection. After converting a specific problem into a model, it evolves the chromosomes using selection, recombination, and mutation operators. To feign the natural reproduction and mutation of species, two basic operators, crossover and mutation, are used during evaluation. It can be used to evolve some general rules for network traffic data. To segregate normal network connections from abnormal connections, these rules are used. These abnormal or anomalous connections refer to events which possess a pretty good possibility of intrusions. The rules that are stored in the rule base are usually of the following form:

if { condition } then { act }

which means if some specific conditions (like source IP address, destination IP address etc.) matches with any current network connection, that will be considered as malicious and necessary actions should be taken according to the rule base which may include stopping the establishment of connection, denial of the request for any service by that network etc.

The eventual reason of using GA is to generate rules that can detect anomaly. So as to fully feat the apprehensive level, all the fields of a network connection need to be examined properly.

Advantages:

- Provide multiple solutions for the problems.

Disadvantages:

- Slow convergence

5.11 C4.5 Decision Tree

[34, 35, 36] Decision trees are one of the most popular learning method which is supervised in nature. It partitions the data with same properties into groups and these groups are kept as analogous as possible. It takes a set of classified data as input, executes the algorithm on that and provides a tree as output where each leaf can be expressed as a decision and each intermediate node epitomizes a test.

C4.5 Algorithm: Entropy is used to measure the amount of uncertainty. The entropy will be zero if all of the data in the set resides in a single class. C4.5 algorithm works in the following three steps:

Step 1: Constructing a decision tree

Create a root node N:

- If T be a member of the same class C, then N will be returned as a leaf node, and it will be marked as class C.
- If the remnants sample of T is less than a specified value or attribute-list is null, then N will be returned as a leaf node, and will be marked as a category that appears most repeatedly.
- For each attribute, calculate its information gain ratio in the attribute list.

The Gain Ratio can be defined as-

$$Gain(A) = Info(T) - Splitinfo_A(T)$$

The information gain is defined based on the splitting criterion. The Split Information can be defined as-

$$Splitinfo_A(T) = \sum_{j-1}^{v} \frac{|T_j|}{|T|} X\ INfo(T)$$

The split information is potential information generated by splitting the training data set T; it gives positive results to most of the applications.

Step 2: Then every possible path from root to leaf node gives a new rule which can be expressed as the conjunction of the attributes and their values. This way, a decision tree can be transferred into if-then statement very easily.

Step 3: Determine the behavior of the new networks by determining whether it is an intrusion or not according to classification rules.

Advantages

- Both discrete and continuous attributes can be handled by C4.5. Indication of missing values is also possible by using "?" sign.

Disadvantages

- Over fitting generally occurs if the algorithm model works on data with infrequent features. Susceptible to noise.

6. CONCLUSION

Data mining techniques are extensively used recently for intrusion detection now-a-days, to decrease the great encumbrance of analyzing huge volumes of network traffic. Achieving high detection rate and reducing false alarm rates are the significant challenges in designing an intrusion detection system. With an intention to address these issues, this paper provides an empirical study of a detailed overview of Intrusion Detection System as well as brief explanations of some selected data mining approaches available for network intrusion detection. These techniques work very well for IDS but not even a single technique can identify all types of attacks. Also as data mining is still in a developing state, more study and research needs to be done and more effort is needed to improve the performance of these techniques or invention of more new hybrid techniques to identify all types of attacks.

7. REFERENCES

[1] Pohlheim, Hartmut. 30 Oct. 2003. "Genetic and Evolutionary Algorithms: Principles, Methods and Algorithms." Genetic and Evolutionary Algorithm Toolbox. Hartmut Pohlheim.

[2] Whitley, Darrell. 1994. "A Genetic Algorithm Tutorial." Statistics and Computing 4: 65-85.

[3] Crosbie, Mark, and Gene Spafford. 1995. "Applying Genetic Programming to Intrusion Detection." In Proceedings of 1995 AAAI Fall Symposium on Genetic Programming, pp. 1-8. Cambridge, Massachusetts.

[4] J. T. Yao, S. L. Zhao and L. V. Saxton, "A study on Fuzzy Intrusion Detection," Proceedings of Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security, 2005, pp. 23-30.

[5] Shanmugam, Bharanidharan, and Norbik Bashah Idris. *Hybrid intrusion detection systems (HIDS) using Fuzzy logic*. INTECH Open Access Publisher, 2011.

[6] Subaira, A. S., and P. Anitha. "Efficient classification mechanism for network intrusion detection system based on data mining techniques: A survey." *Intelligent Systems and Control (ISCO), 2014 IEEE 8th International Conference on*. IEEE, 2014.

[7] Hassan, Mostaque Md Morshedur. "Network Intrusion Detection System Using Genetic Algorithm and Fuzzy Logic." *International Journal of Innovative Research in Computer and Communication Engineering* 1, no. 7 (2013).

[8] Thu, 02 May 2013 15:54:11 +0200

[9] Current Studies On Intrusion Detection System, Genetic Algorithm And Fuzzy Logic CoRR abs/1304.3535 2013Survey on Intrusion Detection System using Support Vector Machine International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 12, December 2014)

[10] Wang, GuiPing, ShuYu Chen, and Jun Liu. "Anomaly-based Intrusion Detection using Multiclass-SVM with Parameters Optimized by PSO." *International Journal of Security and Its Applications* 9.6 (2015): 227-242.

[11] Patel, Reema, Amit Thakkar, and Amit Ganatra. "A survey and comparative analysis of data mining techniques for network intrusion detection systems." *International Journal of Soft Computing and Engineering (IJSCE) ISSN* (2012): 2231-2307.

[12] Mukkamala, S., Janoski, G., & Sung, A. (2002, May). Intrusion detection: support vector machines and neural networks. In *proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO* (pp. 1702-1707).

[13] Mulay, Snehal A., P. R. Devale, and G. V. Garje. "Intrusion Detection System Using Support Vector Machine and Decision Tree." *International Journal of Computer Applications IJCA* 3, no. 3 (2010): 40-43.

[14] Modelling Intrusion Detection System using Hidden Markov Model: A Review Preeti Saini,Ms. Sunila Godara Volume 4, Issue 6, June 2014 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering Research Paper Available online at: www.ijarcsse.com

[15] Rabiner, Lawrence, and Biing-Hwang Juang. "An introduction to hidden Markov models." ASSP Magazine, IEEE 3.1 (1986): 4-16.

[16] Devarakonda, Nagaraju, et al. "Intrusion Detection System using Bayesian Network and Hidden Markov Model." Procedia Technology 4 (2012): 506-514.

[17] Khosronejad, Mahsa, et al. "Developing a hybrid method of Hidden Markov Models and C5. 0 as a Intrusion Detection System." *International Journal of Database Theory and Application* 6.5 (2013): 165-174.

[18] Khreich, Wael, Eric Granger, Ali Miri, and Robert Sabourin. "A survey of techniques for incremental learning of HMM parameters." Information Sciences 197 (2012): 105-130.

[19] Hari Om, AritraKundu, "A Hybrid System for Reducing the False Alarm Rate of Anomaly Intrusion Detection System", 1st Int'l Conf. on Recent Advances in Information Technology RAIT-2012, 978-1-4577-0697-4/12/$26.00 ©2012 IEEE

[20] Shakiba Khademolqorani, Ali Zeinal Hamadani "An Adjusted Decision Support System through Data Mining and Multiple Criteria Decision Making" The 2nd International Conference on Integrated Information Elsevier 2013

[21] Kanungo T., Mount D. M. 2002 An Efficient k-means Clustering Algorithm: Analysis and Implementation,

IEEE Transactions on Pattern Analysis and Machine Intelligence Vol: 24 , Issue: 7

[22] Parveen Kumar and Nitin Gupta, "A Hybrid Intrusion Detection System Using Genetic–Neural Network", International Journal of Engineering Research and application (IJERA) ISSN: 2248-9622 ,National Conference on Advances in Engineering and Technology (2014)

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

[24] C. Gates, J. J. McNutt, J. B. Kadane, and M. I. Kellner,"Scan Detection on Very Large Networks Using LogisticRegression Modeling," *11th IEEE Symposium on Computersand Communications,ISCC '06*, 2006, pp. 402-408.

[25] http://www.cs.unm.edu/~terran/downloads/classes/cs529-s11/fp_presentations/wadsworth.pdf

[26] https://en.wikipedia.org/wiki/Logistic_regression#Basics

[27] (COPY OF 141) Subaira.A.S, Mrs.Anitha.P, "Efficient Classification Mechanism for Network Intrusion Detection System Based on Data Mining Techniques: a Survey", *2014 IEEE 8th Proceedings International Conference on Intelligent Systems and Control (ISCO),* 978 -1- 4799 - 3837 - 7/14/$31.00© 2014 IEEE

[28] Manish Kumar Nagle, Dr. Setu Kumar Chaturvedi, "Feature Extraction Based Classification Technique for Intrusion Detection System", International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, Volume 8, Issue 2 (August 2013), PP. 23-38

[29] Ajayi Adebowale, Idowu S.A, Anyaehie Amarachi A. , "Comparative Study of Selected Data Mining Algorithms Used For Intrusion Detection", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-3, July 2013

[30] M.A. Maloof, Machine Learning and Data Mining for Computer Security, Springer- Verlag, 2006.

[31] Hind Tribak , Blanca L. Delgado-Marquez, P.Rojas, O.Valenzuela, H. Pomares and I. Rojas, " Statistical Analysis of Different Artificial Intelligent Techniques applied to Intrusion Detection System", IEEE, 2012

[32] Dharminder Kumar, Suman, " Performance Analysis of Various Data Mining Algorithms: A Review", *International Journal of Computer Applications (0975 – 8887) Volume 32– No.6, October 2011*

[33] Han J., Kamber M., "Data Mining: Concepts and Techniques, 2nd edition", Morgan Kaufmann, 2006.

[34] Johan Baltié, DataMining : ID3 et C4.5, Promotion 2002, Spécialisation S.C.I.A. Ecole pour l'informatique et techniques avancées.

[35] Rong Cao,Lizhen Xu,Improved C4.5 Decision tree algorithm for the analysis of sales.Southeast University Nanjing211189,china,2009.

[36] Gaurav L. Agrawal1, Prof. Hitesh Gupta2, Optimization of C4.5 Decision Tree Algorithm for Data Mining Application, IJTAE,ISSN 2250-2459, Volume 3, Issue 3, March 2013