Experiments on Detection of Denial of Service Attacks using Bayesian Network Classifier

Mr. Vijay D. Katkar
Department of Information Technology,
Pimpri Chinchwad College of Engineering,
Pune, India.
katkarvijayd@gmail.com

Mr. Yogesh P. Jagdale
Department of Information Technology,
Pimpri Chinchwad College of Engineering,
Pune, India.
ypjagdale@gmail.com

Mr. Rohan P. Petare
Department of Information Technology,
Pimpri Chinchwad College of Engineering,
Pune, India.
petarerohan@gmail.com

Mr. Raviraj B. Mohite
Department of Information Technology,
Pimpri Chinchwad College of Engineering,
Pune, India.
ravibmohite@gmail.com

***Abstract*-Denial of Service (DoS) and Distributed Denial of Service (DDoS) attack exhausts the resources of server/service and makes it unavailable for legitimate users. It can result in huge loss of money. With increasing use of online services and attacks on these services, the necessity of Intrusion Detection System (IDS) for detection of DoS/DDoS attacks has also marked by organizations. Different techniques such as data mining, neural network, genetic algorithms, pattern recognition are being used to design IDS. This paper evaluates variation in performance of Bayesian Network classifier for intrusion detection when used in combination with different data pre-processing and feature selection methods. Experimental results prove that accuracy of Bayesian Network classifier is improved and performs better than other classifiers when used in combination with Feature Selection and data pre-processing methods.**

***Keywords*-Bayesian Network, Feature selection, Intrusion Detection System, Denial of Service Attack**

I. INTRODUCTION

Now-a-days Internet has become the Integral part of a human life, but along with its advantages there are some threats to it like DoS and DDoS attacks. According to [1] number of attacks on online system is increasing quickly. DoS is a class of attacks where an attacker makes some computing or memory resource too busy or too full to handle legitimate requests, thus denying legitimate users access to a system. There are different ways to launch DoS attacks: by abusing the computers legitimate features by targeting the implementation bugs or by exploiting the system's Misconfigurations.

In Distributed Denial of Service (DDoS) attack multiple systems which are infected by Trojans and are used to attack a single system. As a result IDS becomes necessary for every online system. The most popular way to detect intrusions is to audit data generated by operating systems and other devices present in network. This makes IDS a valuable real-time detection and prevention tool as well as a forensic analysis tool. There are two basic ways for detecting intrusion: Misuse-based detection & Anomaly-based detection.

The idea of misuse detection [2] is to represent attacks in the form of a pattern or a signature so that the same attack can be detected and prevented in future. These systems can detect many or all known attack patterns, but they are of little use for detecting naive attack methods.

Anomaly-based [3] detection establishes a normal activity profile for a system and flag all system states varying from the established profile as intrusion attempts. Anomaly detection is normaly accomplished with thresholds and statistics but can also be done with Soft computing, and inductive learning. Data mining [4] is used to implement IDS by researchers.

Rest of the paper is organized as follows. Section 2 briefly describes related work. Section 3 explains Naïve Bayesian Classifier in short. Section 4 and 5 explain in brief about feature selection and data pre-processing methods respectively. Section 6 presents experimental results and section 7 compares the experimental results with existing work. Section 8 concludes the paper.

II. RELATED WORK

Siva S. Sivatha Sindhu et al. [5] have proposed Decision Tree based Light Weight IDS (DT-LWIDS). The entire process is divided into 3 steps. Initially redundant records were removed from the training set to improve the accuracy of classification algorithm. Then feature selection process was applied to select a subset of features from available set; it reduce a load on classifier and make it light weight.

Finally classification is performed using combination of neural network and decision tree (i.e. neurotree).

Ozgur Depren et al. [6] have proposed Intelligent Hybrid (Anomaly-Misuse detection) IDS (IH-IDS). Anomaly detection module is implemented using Self Organizing Maps and Misuse detection module is implemented using J.48 decision tree. Decision Support System is used to interpret output of both the modules and report the Intrusion detection event. Amuthan Prabakar Muniyandi et al. [7] have proposed an anomaly detection based IDS using combination of K-Means and C4.5 (KM + C4.5 IDS). The k-Means clustering is used initially to partition the training dataset into k clusters using Euclidean distance. Decision tree is built for each using C4.5 decision tree algorithm. It refines the decision boundaries by learning the subgroups within cluster created by k-means. Rules created by decision tree are used to detect intrusive events

Gang Wang et al. [8] have proposed FC-ANN IDS based on combination of Artificial Neural Network (ANN) and Fuzzy Clustering (FC). Fuzzy Clustering technique is used to divide training dataset into several homogenous subsets. This reduces the complexity of each training subset and increases the detection performance. Generated training subsets are used to train different ANN classifiers. Finally fuzzy aggregator is used to combine outputs of different classifiers for final prediction.

Shi-Jinn Horng et al. [9] have proposed design of IDS using combination of Support Vector Machine (SVM) and hierarchical clustering. Since SVM take very long time to train itself using large dataset; hierarchical clustering algorithm i.e. BIRCH is used to transform training dataset to a smaller sized dataset. This transformed dataset is divided into five groups (four types of attack and normal records). This reduced dataset is then used to train four different SVM classifiers. Outputs of all classifiers are merged to get final result.

Cheng Xiang et al. [10] have proposed multiple-level hybrid classifier for IDS (MLH- IDS) using decision trees and Bayesian clustering. Detection process is divided into four stages. Output of every stage is given as input to next stage. In first stage test data is classified into three categories (i.e. DoS, Probe, Others).Second stage categorizes records classified as ''Others" into two categories (i.e. Attack and Normal). Third stage further categorizes Attack class as U2Rand R2L. Finally stage further classifies every intrusive record into more specific attack types.

III. BAYESIAN NETWORK CLASSIFIER

Bayesian Network is also known as the 'Belief Network' represented by directed acyclic graph consisting of nodes and directed edges connecting the nodes. The Bayesian Network classifier is a probabilistic graphics model which uses discrete data structures to come to a conclusion and making good decision under uncertainty. It is one of the branches of graphical representation of distribution. Bayesian network model represents a factorization of the joint probability of all random variables.

Bayesian Network makes possible a factorization of the probability distribution of the n-dimensional random variable (x1,…xn) in the following way:

$$P(x1, \dots, xn) = \prod_{i=1}^{n} P(xi|\pi(xi))$$

Where xi represents the value of the random variable Xi, and π (xi) represents the value of the random variables which are parents of Xi. Thus, in order to specify the probability distribution of a Bayesian Network, one should give prior probabilities for all root nodes (nodes having no predecessors) and conditional probabilities for all other nodes, given all possible combinations of their direct predecessors. The value obtained from above formula along with the directed acyclic graph completes the Bayesian network.

IV. FEATURE SELECTION

Feature selection selects the subsets of original represented attributes for the maximum information in data set by making use of class information. Along with having an active research area in pattern recognition and statistics, feature selection is the unique approach which removes away the unnecessary & irrelevant features from dataset to prevent decrease in accuracy. The main principal of feature selection is to choose a subset of input variables by eliminating features with little information. In short its aim is to find a good subset of features that forms high quality of clusters for a given number of clusters. In this paper we had worked on the 3 Feature selection subsets with the Bayes Net classifier and combination of different pre-processors. Researchers have used various feature selection methods to obtain most relevant attributes like - CFS subset, Information gain and gain ratio [10, 11 and 12].

A. CFS Subset Evaluation

CFS subset evaluates and ranks feature subsets which prefer the set of attributes that are highly related with the class but with low interconnection. It first calculates a matrix of feature-class and feature-feature correlations from the training data and then searches the feature subset space using a best first. It also gives high scores to subsets that include features

that are highly correlated to the class attribute but have low correlation to each other. The following formula is used,

$$Merit_{sk} = \frac{Kr_{cf}}{\sqrt{k + k(k-1)r_{ff}}}$$

Where,
rcf = average of feature-classification correlations
rff = average of feature-feature correlations

B. Information gain

The IG evaluates attributes by measuring their information gain with respect to the class. It discretizes numeric attributes first using MDL based discretization method. It also measures amount of information in bits of the class prediction .It measures the expected reduction in entropy uncertainty associated with a random feature.

The formula for computing IG is as follows -

$$info(D) = -\sum_{i=1}^{\infty} p_i \log(p_i)$$

Where, p_i is the probability that an arbitrary tuple D belongs to a class c_i. Info(D) is also known as the entropy of the tuple D.

C. Gain ratio

The information gain select attributes having a large number of values. It is an extension of info gain that attempts to overcome this bias. It considers number and size of branches when choosing an attribute. It also corrects the information gain by taking the intrinsic information of a split into consideration ,Where Intrinsic information is an entropy of distribution of instances into branches (i.e. how much information do we need to tell).
The formula for computing gain ratio is:

$$Gain\_Ratio(Attribute) = \frac{Gain(Attribute)}{Intrinsic\_Info(Attribute)}$$

V. DATA PRE-PROCESSING

Given data may be incomplete, inconsistent & having some errors. For example it may contain impossible data (such as child = Yes, age = 36). Processing such data may lead to unreliable result. Data Preprocessing is proven method to resolving such issue. Data preprocessing consist of Data Cleaning, Data Integration, Data Transformation, Data Discretization etc. This paper studies the effects of following Data pre-processing methods on performance of 'Bayesian Network' classifier.

A. Discretize

This preprocessing method checks whether range of numeric attributes in given data set is in nominal values. If the range is within the nominal values it will skip the class feature otherwise it will discretize it.

B. Normalize

This preprocessing method normalizes all numeric attributes in given data sets apart from class attribute if set. By default the resulting values are in [0, 1].

C. Numeric to Binary

It converts all numeric attributes into binary attribute apart from the class attribute, if set. If value of numeric attribute is exact zero then will value of new attribute be zero, if missing then missing otherwise it will be one. The value of new attribute is nominal.

D. Nominal to Binary

It converts all nominal attributes into numeric binary attributes. An attribute with M values is converted into M binary attributes if the class is nominal

E. First Order

This preprocessing takes a range of N numeric attributes and replaces them with N-1 numeric attributes, the values of which are the difference between consecutive attribute values from the original instance.

F. PKI Discretize

Discretizes numeric attributes using equal frequency binning, where the number of bins is equal to the square root of the number of non-missing values.

VI. EXPERIMENTAL RESULTS

Intel core-i5 (2.67 GHz) machine having 6GB RAM is used to perform the experiments. The performance of Bayes Net classifier in combination with the feature selection and data pre-processing methods is evaluated using KDD 99 dataset [13]. Records belonging to the known Dos/DDos attacks and the normal behavior were extracted from training and the testing dataset provided by KDD99 to create the training and testing dataset for experimentation. All the performances were performed using the Weka tool (version 3.7.9) [14].Throughout these experiments the heap size allocated to the Weka tool was 3048 MB using the Java –Xmx command.

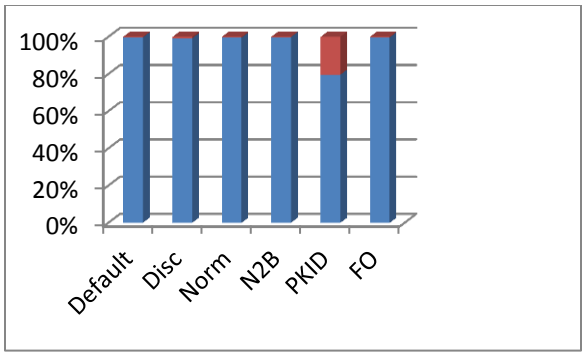

Fig.1.Without Attribute Selection

Legend:
Red: incorrectly classified
Blue: correctly classified

TABLE I: ABBREVIATIONS USED IN GRAPHS

Abbreviation	Description
Default	Default
Disc	Discretize
Norm	Normalize
N2B	Numeric To Binary
NO2B	Nominal To Binary
PKID	PKI Discretize
FO	First Order

It can be observed from the figure 1 that, Bayes net (Belief Network) classifier gives the accuracy of 99.8425 when combined with the numeric to binary pre-processing.

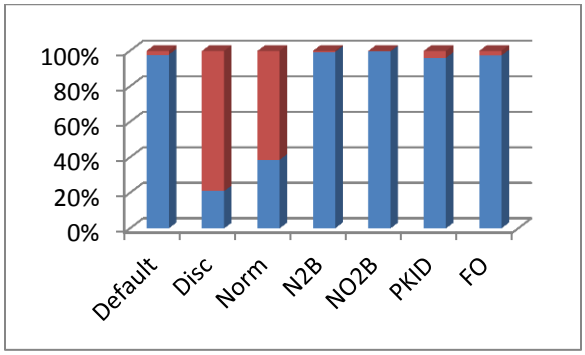

Fig.2. Information Gain Attribute Selection

It can be observed from the figure 2 that, Bayes Net classifier provides the accuracy of 99.7446% when coupled with Nominal to Binary Pre-Processing. Features selected by Information Gain attribute selection method when applied on training set is in table II:

TABLE II: FEATURES SELECTED

Feature selection	Selected features
Information gain + nominal to Binary	5,23,36,24,3,2,33,35,34,30,4,29,6,40,42

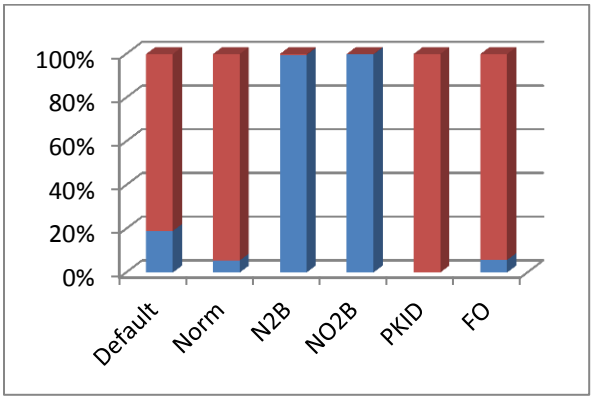

Fig.3. Gain Ratio

It can be observed from the figure 3 that, Bayes net in combination with the Gain Ratio feature selection and Nominal To Binary Pre-Processing gives the maximum accuracy of 99.7837% when applied on the KDD99 cup training set. Features selected by Gain ratio attribute selection method when applied on training set is in table III.

TABLE III: FEATURES SELECTED

Feature selection	Selected features
Gain ratio + nominal to Binary	7,3,4,2,13,10,29,12,36,28,42

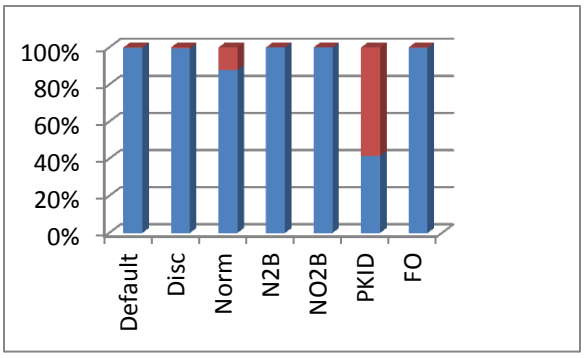

Fig.4. Correlation Feature Selection (CFS Subset Evaluation)

It can be observed from figure 4 that, Bayesian Network classifier provides the accuracy of 99.9348% when combined with Numeric to Binary Pre-Processing along with the feature selection Correlational Feature Selection (CFS) methods when applied on the training set are shown is in table IV:

TABLE IV: FEATURES SELECTED

Feature selection	Selected features
Correlational Feature Selection + Numeric to Binary	4,5,6,7,8,23,30,36,42

VII. COMPARISON

Table V shows comparison between BN+N2B classifier and solutions proposed by other researchers. FC-ANN and KM + C4.5 do not used entire KDD dataset for training and testing their proposed mechanism. Instead they use randomly selected records from KDD dataset for training and testing. Such random selection of records may results in different detection accuracy every time system is evaluated. DT-LWIDS with multi class classifier approach has used 10-Fold cross validation method to evaluate their performance instead of using testing dataset provided by KDD. It can also result in different detection accuracy every time system is evaluated.

FC-ANN uses Fuzzy Logic for data preprocessing, and DT-LWIDS uses Remove redundant records data pre-processing methods which require more computational power in terms of CPU utilization and memory as compared to Numeric to Binary conversion.

FC-ANN, SVM with Hierarchical Clustering, MLH- IDS and KM + C4.5 uses multiple classifiers, which require additional computational resources as compared to BN+N2B.

VIII. CONCLUSION

Bayesian Network classifier performed significantly better when combined with Numeric to Binary data pre-preprocessing. It can also be observed that, instead of going for set of multi-classifiers, one can achieve better performance using Bayesian Network classifier along with Numeric to Binary data pre-processing.

TABLE V: COMPARISON OF PROPOSED MECHANISM WITH EXISTING SOLUTIONS

Method	Training and Testing Dataset	No of Classifiers used	Data Preprocessing Method	Detection Accuracy for DDoS records
FC-ANN	Randomly Selected	Dynamically decided	Fuzzy Logic	99.91%
KM + C4.5	Randomly Selected	2	none	98.20%
DT-LWIDS	10-fold cross validation	1	Remove redundant records	98.87%
SVM with Hierarchical Clustering	All records belonging to DOS/DDoS attacks and Normal connection present in KDD 99	4	Division (Divide each attribute value by its max value)	99.53%
MLH- IDS	Randomly selected for training & complete dataset for testing	4	None	99.19%
BN + N2B	All records belonging to DOS/DDoS attacks and Normal connection present in KDD 99	1	Numeric To Binary	99.9348%

REFERENCES

[1] <http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_2011_21239364.en-us.pdf>(accessed January 2013)

[2] Ming-Yang Sua, Gwo-Jong Yub, Chun-Yuen Lina, A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach, *COMPUTERS & SECURITY* vol. 28 (2009) pp. 301-309

[3] Ming-Yang Su, Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers, *Expert Systems with Applications* vol. 38 (2011) pp. 3492–3498

[4] Ming-Yang Su, "Discovery and prevention of attack episodes by frequent episodes mining and finite state machines", Journal of Network and Computer Applications vol. 33, 2010,pp. 156–167

[5] Siva S. Sivatha Sindhu, S. Geetha, A. Kannan, Decision tree based light weight intrusion detection using a wrapper approach, Expert Systems with Applications vol. 39, 2012, pp. 129–141

[6] Ozgur Depren, Murat Topallar, Emin Anarim, M. Kemal Ciliz, An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks, Expert Systems with Applications vol. 29, 2005, pp. 713–722

[7] Amuthan Prabakar Muniyandia, R. Rajeswarib, R. Rajaramc, "Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision

Tree algorithm", International Conference on Communication Technology and System Design, Procedia Engineering vol. 30, 2012,pp. 174 – 182

[8] Gang Wang, Jinxing Hao, Jian Ma, Lihua Huang, A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering, *Expert Systems with Applications* vol. 37 (2010) pp. 6225–6232

[9] Shi-Jinn Horng, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai, Citra Dwi Perkasa, A novel intrusion detection system based on hierarchical clustering and support vector machines, *Expert Systems with Applications* vol. 38 (2011) pp. 306–313

[10] Cheng Xiang, Png Chin Yong, Lim Swee Meng, "Design of multiple-level hybrid classifier for intrusion detection system using Bayesian clustering and decision trees", Pattern Recognition Letters vol. 29, 2008, pp. 918–924

[11] Mrutyunjaya Panda, Ajith Abrahamb, Manas Ranjan Patra, "A Hybrid Intelligent Approach for Network Intrusion Detection", International Conference on Communication Technology and System Design, Procedia Engineering vol. 30, 2012, pp. 1 – 9

[12] Fatemeh Amiri, Mohammad Mahdi Rezaei Yousefi, Caro Lucas, Azadeh Shakery, Nasser Yazdani "Mutual information-based feature selection for intrusion detection systems", Journal of Network and Computer Applications vol. 34, 2011, pp. 1184–1199

[13] KDD,Kdd cup 1999 dataset, <*http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html*>, 1999 (accessed January 2013).

[14] <*http://www.cs.waikato.ac.nz/ml/weka/*> (accessed August 2013)