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

Mr. Vijay D. Katkar, Mr. Siddhant Vijay Kulkarni,
Department of Information Technology, Pimpri Chinchwad College of Engineering, Pune, India.
(kulkarni.siddhant1@gmail.com)

Abstract—**Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks can result in huge loss of data and make resources unavailable for legitimate users. With continuous growth of Internet users and traffic, the importance of Intrusion Detection System (IDS) for detection of DoS/DDoS network attacks has also grown. Different techniques such as data mining and pattern recognition are being used to design IDS. Naïve Bayesian is a widely used classifier for design of IDS. This paper evaluates variation in performance of Naïve Bayesian classifier for intrusion detection when used in combination with different data pre-processing and feature selection methods. Experimental results prove that accuracy of Naïve Bayesian classifier is improved and performs better than other classifiers when used in combination with Feature Selection and data pre-processing methods.**

Keywords—**Naïve Bayesian, Feature selection, Intrusion Detection System, Denial of Service Attack**

I. INTRODUCTION

The internet is a globally public network being used on a daily basis by millions of users. The user base of the internet keeps on growing as more and more people are getting attracted towards the business opportunities provided by the internet. But every coin has two sides and so does the Internet. On one hand the internet provides huge number of potential opportunities for users and on the other it brings a huge risk factor. There exist both harmful and harmless users on the Internet. When some useful information is made available to regular/harmless users, it also becomes available to the harmful users. These users may use this information with various malicious intents.

According to [1] number of attacks on online systems is increasing continuously. DDoS attack is the most prevalent threat which either exploits vulnerability in computing and communication resources or floods them in order to make the system unavailable for legitimate users. This results in massive loss of data, resources and money. As a result IDS is a necessity of every online system.

There are two basic approaches of detecting intrusive activities: Anomaly detection and Misuse detection. Misuse detection [2] (also known as Signature based) looks for patters or signatures of known attacks. If match is found then it generates the alarm. Signature database of known attacks is specified a priori. On the other hand, Anomaly detection [3] attempts to estimate the 'Normal' behavior of the system to be protected and generate an alarm when the deviation between a current behavior of system and normal behavior exceeds a predefined threshold.

Denning [4] proposed first Intrusion detection model and described statistical techniques like threshold, standard deviation, multivariate model for Anomaly detection system. After that many researchers have used various approaches for Intrusion detection which include statistical, machine learning, data mining techniques. Data mining and machine learning methods that involve single classifier [5, 6] and ensemble classifiers [7, 8] have been widely used by researchers.

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

II. RELATED WORK

P. Arun Raj Kumar and S. Selvakumar [9] have proposed ensemble of adaptive and hybrid neuro-fuzzy systems (NFBoost algorithm) to detect known and novel DDOS attacks. Ensemble of classifiers is used to reduce total error and increase detection accuracy. Adaptive Neuro-Fuzzy Inference System (ANFIS) is used for designing base classifier. NFBoost algorithm uses two main classifiers to detect Class 1 (Normal) and Class 2 (Attack). Each classifier consists of many sub-classifiers. Weighted mean approach is used to aggregate outputs of sub-classifiers. Finally NFBoost algorithm combines output of ensemble of classifier and Neyman Pearson cost minimization strategy to get final classification decision.

Gang Wang et al. [10] have proposed FC-ANN IDS based on combination of Artificial Neural

978-1-4673-6126-2/13/$31.00 ©2013 IEEE

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.

Levent Koc et al. [11] have performed experiments with KDD99 dataset, Naïve Bayesian (NB) and its six variants; Tree-Augmented Naïve Bayesian (TAN), Averaged One-Dependence Estimators (AODE), Weightily AODE (WAODE), Decision Tree (NBTree), DTNB, Hidden Naïve Bayesian (HNBNB). Their experimental results show that HNB along with Proportion K-Interval discretization method gives high accuracy for DDoS attack detection.

Shi-Jinn Horng et al. [12] 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.

Srinivas Mukkamala et al. [13] have proposed ensemble of five classifiers; ANN (back propagation, scale conjugate gradient, one step secant), support vector machine and multivariate regression splines to design IDS. Each classifier is trained using complete training dataset (DARPA) to detect all types of attacks. Output of all classifiers is merged using majority vote approach. But it does not specify how to get final predication if votes for two or more classes are equal.

Xuan Dau Hoang et al. [14] have presented a fuzzy-based scheme for the integration of HMM Anomaly Intrusion Detection (AID) engine and normal sequence detection engine for program AID using collected system calls. This approach uses fuzzy sets to represent space of sequence parameters and creates set of fuzzy rules which combine multiple sequence parameters and determine the sequence status through a fuzzy reasoning process.

III. FEATURE SELECTION

Feature selection in itself has been a very active research area in data mining. The basic idea behind feature selection is to choose a sub set of variables which provide the maximum information in data set and eliminate all variables that provide minimum or non predictable information. Feature selection can be utilized to improve the performance of classifiers. More often than not it is the case that finding a correct subset of predictive features is in itself an important problem. For example, a mechanic can make a decision about the cause of a car not working by checking only a few major features of a car.

Even for a small network, the number of features that an Intrusion Detection System needs to analyze is usually huge. Many researchers work on and propose new classifiers to improve the detection rate of an Intrusion Detection System. But improving effectiveness of a classifier is a very tough task. Feature Selection methods help in optimizing the existing classifiers by providing a subset of relevant data.

Feature Selection works on a central assumption that the provided data contains redundant and irrelevant features. Feature Selection helps reduce the computational complexity, remove information redundancy and improve the generalization.

The algorithms used for Feature Selection can be categorized as Wrapper and Filter methods. As a part of selection process, Wrapper methods try to optimize a set of predefined criteria's with respect to feature set. Filter methods rely on characteristics of training data for selecting features that are highly dependent on the output but are independent of each other. In this paper we evaluate Naïve Bayes classifier in combination with Correlation Feature Selection (Best First and Greedy Stepwise) and Information Gain based Attribute Selection.

A. Correlation Feature Selection (CFS Subset Evaluation)

This filter method of feature selection evaluates the worth of a subset of features by considering the individual predictability of each feature and the degree of redundancy between them. It prefers to select the features that are highly related to the class and at the same time are very much unrelated to each other. The following equation gives the merit of a feature subset S consisting K features:

$$Merit_{S_k} = \frac{kr_{cf}}{\sqrt{k + k(k-1)r_{ff}}} \qquad (1)$$

Where,

k = Number of features in feature subset S

r_{cf} = average of feature-classification correlations

r_{ff} = average of feature-feature correlations

Best First Feature selection: This algorithm provides the globally best subset of features from a

dataset. It may start with an empty subset and search forward for the best features or it may start with all the attributes and work its way backwards searching for the best possible features and eliminating others.

Greedy Stepwise: This algorithm works on the problem solving method of selecting the locally optimum solution and hoping that it is globally optimum. This algorithm adds the best feature or deletes the worst one in each iteration of its execution.

B. *Information Gain*

This feature selection method is based on the work of Claude Shannon on Information Theory. This algorithm focuses on minimizing the number of steps required to classify the tuples in a given data set. The expected information needed to classify a tuple D is given by:

$$info(D) = -\sum_{i=1}^{\infty} p_i log(p_i) \qquad (2)$$

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.

Ranker: This feature selection filter ranks all the features in the given data set by the degree of relevant information that they provide for classifying each tuple. This method does not eliminate but simply prioritizes the features and provides one the freedom to choose features depending on a defined criterion.

IV. DATA PRE-PROCESSING

Data pre-processing is a very important step in classification of data. The methods of data-gathering are usually very loosely controlled and thus result in out-of range (e.g. marks=-2), missing or even (at times) impossible data combinations (e.g. present=false, pass=true). Classifying such unreliable data will no doubt end up with un-reliable results. Thus to avoid such issues in classification Data pre-processing stage is carried out.

Data pre-processing consists of data cleaning, data integration & transformation, data reduction, normalization, discretization, etc. This paper studies the effects of following Data pre-processing methods on performance of Naïve Bayesian classifier.

A. *Discretize:*

This pre-processing method discretizes a range of numeric attributes in the given data set into nominal values. If specified it will skip the class feature.

B. *Nominal To Binary:*

It converts all the nominal values to numeric binary values. An attribute with N nominal values is converted into N numeric binary attributes.

C. *Normalize:*

This pre-processing method normalizes all the attributes in the provided data set except class attribute(if so specified).

D. *Numeric to Binary:*

This converts all the numeric attributes into binary attributes. If the value of numeric attribute is other than 0, then value of new binary attribute will be 1.

E. *Numeric to Nominal:*

Unlike Discretize, this pre-processing method takes all the numeric values and adds them to the list of nominal values of that attribute.

F. *PKIDiscretize:*

Uses Equal Frequency binning to discretize numeric values and uses number of bins equal to the square root of number of non-missing values.

G. *Standardize:*

The output of this pre-processing method has zero mean and unit variance.

V. NAÏVE BAYESIAN CLASSIFIER

Naïve Bayesian classifier is a simple probabilistic classifier which works by applying the Baye's theorem along with naïve assumptions about feature independence. It assumes value of any feature is independent of values of other features. This assumption is also known as Conditional Independence. Despite the naïve assumption and over simplification, Naïve Bayesian classifiers have proved to be quite useful in complex real world conditions.

Probabilistic model of Naïve Bayesian Classifier is as follows:

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i)$$
$$= (x_k|C_i)P(x_1|C_i) * P(x_2|C_i) * \dots \qquad (3)$$

Naïve Bayesian considers the probability of a tuple X belonging to a class C_i is equal to the multiplication of probabilities that each attribute of the tuple X belongs to the classC_i,where the probabilities$P(x_1|C_i)$,$P(x_2|C_i)$ and so on can be easily calculated from the training data sets.

VI. EXPERIMENTAL RESULTS

Intel Core-i5 (1.70 GHz) machine having 4 GB Ram is used to perform experiments. Performance of Naïve Bayesian classifier in combination with Feature Selection and Data pre-processing methods is evaluated using KDD 99 dataset [15].Records belonging to known DoS/DDoS attacks and Normal behavior were extracted from training and testing dataset provided by KDD99 to create training and testing dataset for experimentation. All the experiments were performed using Weka tool (Version 3.7.9) [16].Throughout these experiments, the heap size allocated to Weka tool was 1408 MB using the Java –Xmx command.

Figure 1 shows accuracy of Naïve Bayesian classifier when training and testing dataset were pre-processed using various data pre-processing methods.

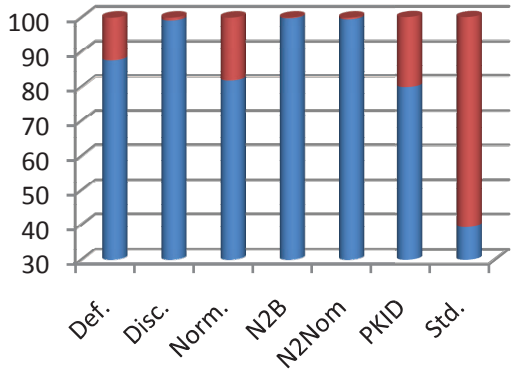

Fig.1.Without Attribute Selection

Table 1 describes abbreviations used in figure 1.

TABLE I
ABBREVIATIONS USED IN GRAPHS

Abbreviation	Description
Def.	Default
Disc.	Discretize
Norm.	Normalize
N2B	Numeric To Binary
N2Nom	Numeric to Nominal
PKID	PKI Discretize
Std.	Standardize

Color Reference-

- Red – Incorrectly classified,
- Blue – Correctly classified

It can be observed from figure 1 that, Naïve Bayesian classifier provides the accuracy of 99.8418% when combined with Numeric to Binary Pre-processing. Features selected by Correlation Feature Selection (CFS) and Information Gain attribute selection methods when applied on training set are shown in Table 2.

TABLE II

SELECTED ATTRIBUTES

FEATURE SELECTION	SELECTED FEATURES
CFS + Best First	2,3,4,5,6,7,8,23,30,36
CFS + Greedy Stepwise	2 ,3, 4, 5, 6, 7, 8, 23 , 30, 36
Information Gain	2, 3, 4, 5, 6,23,24,29, 30, 33,34,35, 36,38

Figure 2 shows accuracy of Naïve Bayesian classifier when selected features of training and testing dataset by CFS + Best first method were pre-processed using various data pre-processing methods.

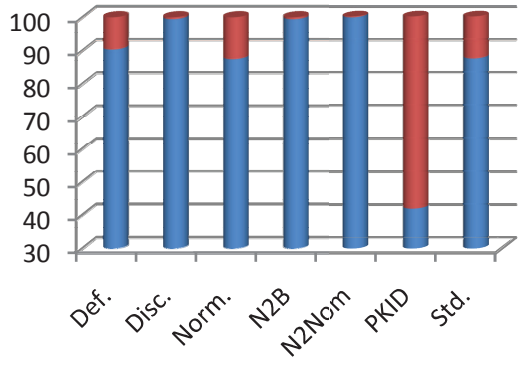

Fig.2 Best First Feature Selection

Naïve Bayesian in combination with CFL + Best First feature selection and numeric to nominal pre-processing gives the maximum accuracy of 99.76%.

Figure 3 shows accuracy of Naïve Bayesian classifier when selected features of training and testing dataset by CFS + Greedy Stepwise method were pre-processed using various data pre-processing methods.

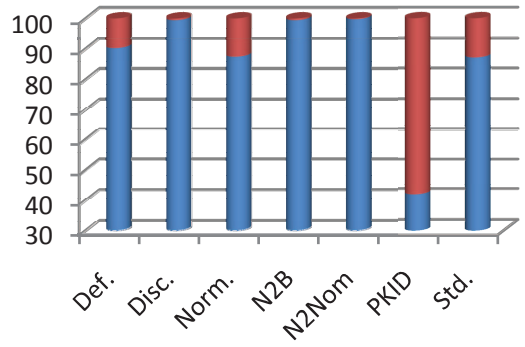

Fig.3. Greedy Stepwise Feature Selection

Results received with Greedy Stepwise feature selection method are identical to the ones found in Best First Attribute Selection method.

Figure 4 shows accuracy of Naïve Bayesian classifier when selected features of training and testing dataset by Information Gain method were pre-processed using various data pre-processing methods.

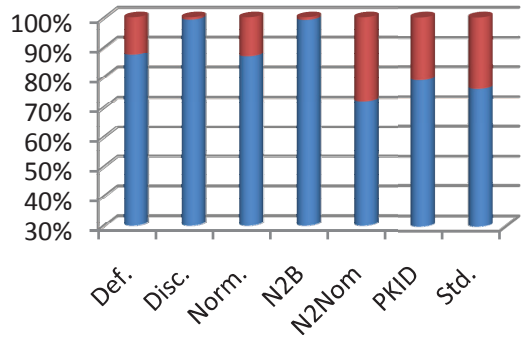

Fig.4. Information Gain Attribute Selection

Naïve Bayesian in combination with Information Gain feature selection and discretize pre-processing gives the maximum accuracy of 99.26%.

VII. COMPARISON

Table 3 shows comparison between NB+N2B classifier and solutions proposed by other researchers. FC-ANN and NFBoost 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. HNB with multi class classifier approach has used Cross Fold 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, NFBoost uses Min-Max Normalization and HNB Multi Class Classifier uses PKIDiscretize 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 and NFBoost use multiple classifiers, which require additional computational resources as compared to NB+N2B.

VIII. CONCLUSION

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

TABLE III

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%
NFBoost	Randomly Selected	Dynamically Decided	Min-Max Normalization	98.20%
HNB Multi Class Classifier	Cross Fold (10) Validation	1	PKIDiscretize	99.6
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%
NB + N2B	All records belonging to DOS/DDoS attacks and Normal connection present in KDD 99	1	Numeric To Binary	99.84%

IX. 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] D.E. Denning, An intrusion detection model, *IEEE Transactions on Software Engineering* vol. SE-13 (February (2)) (1987) pp. 222–232.

[5] Y.Y. Chung, N. Wahid, A hybrid network intrusion detection system using simplified swarm optimization (SSO), *Appl. Soft Comput. J.* (2012), http://dx.doi.org/10.1016/j.asoc.2012.04.020

[6] 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

[7] A. Zainal, M.A. Maarof, S.M. Shamsuddin, Ensemble classifiers for network intrusion detection system, *Journal of Information Assurance and Security* vol. 4 (2009) pp. 217–225.

[8] S. Mukkamala, A. Sung, A. Abraham, Intrusion detection using an ensemble of intelligent paradigms, Journal of Network and Computer Applications vol. 28 (April (2)) (2005) pp. 167–182.

[9] P. Arun Raj Kumar, S. Selvakumar, Detection of distributed denial of service attacks using an ensemble of adaptive and hybrid neuro-fuzzy systems, *Computer Communications* vol. 36 (2013) pp. 303–319

[10] 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

[11] LeventKoc, Thomas A. Mazzuchi, ShahramSarkani, A network intrusion detection system based on a Hidden Naïve Bayesian multiclass classifier, *Expert Systems with Applications* vol. 39 (2012) pp. 13492–13500

[12] 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

[13] Srinivas Mukkamala, Andrew H. Sung, Ajith Abraham, Intrusion detection using an ensemble of intelligent paradigms, *Journal of Network and Computer Applications* vol. 28 (2005) pp. 167–182

[14] Xuan Dau Hoang, Jiankun Hu, Peter Bertok, A program-based anomaly intrusion detection scheme using multiple detection engines and fuzzy inference, *Journal of Network and Computer Applications* vol. 32 (2009) pp. 1219–1228

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

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