RESEARCH ARTICLE OPEN ACCESS

Augmenting Search based Feature Selection to Enhance Efficacy of Bayesian Classifiers for Building Network Intrusion Detection Models

Ashalata Panigrahi*
* *Roland Institute of Technology, Berhampur, India*

ABSTRACT:
With the advent of sensor networks, Internet of Things, and social networks there has been flooding of data across computer networks. This has led to hackers being active in the network creating all kinds of nuisance, viz., password cracking, peer-to-peer attack, eavesdropping attack, DOS attack etc. by exploiting system vulnerabilities. Day-by-day cyber-attacks are becoming more and more sophisticated, posing serious challenge for security experts to identify unknown attacks. Thus, there is a need for building effective intrusion detection systems(IDS) to detect and classify unforeseen and unpredictable cyber-attacks. The objective of this paper is to build an intrusion detection system based on four Bayes net classifiers,viz., Hill Climbing search, K2 search, Tabu Search, and Tree Augmented Naive-Bayes, combined with three bio-inspired feature selection methods, viz., ant search, genetic search, particle swarm optimization search, two informed search feature selection methods, viz., best first search and greedy stepwise search, random search, vote harmony search, EDA search, and rank search. The best combination has been identified to build an effective IDS after evaluating the effectiveness of each combination in terms of accuracy, precision, detection rate, false alarm rate, and efficiency.

Keywords: Hill climbing, Informed search, Particle swarm optimization, Ant search, Bayesian network.

Date of Submission: 18-07-2021 Date of Acceptance: 03-08-2021

I. INTRODUCTION

Today, it is not possible to imagine a world without Internet. Internet is expanding at an amazing rate and plays an important role in almost all fields such as entertainment, education, research and development, business transactions, social networks including Facebook, WhatsApp, Instagram, Twitter. The unstoppable growth of Internethas led to security issues, thereby forcing organizations to continuously assess the network vulnerabilities and adopt different defense mechanisms such as user authentication, encryption, firewall etc to protect their systems from cyber-attacks. As cyber-attacks are becoming more sophisticated day-by-day, it has become a real challenge to identify unknown attacks. There has been an increase in security threats such as zero-day attacks designed to target internet users. Many countries have been significantly impacted by the zero-day attacks. According to the 2017 Symantec Internet Security threat report [1] more than three billion zero-day attacks were reported in 2016. Intrusion detection system have been developed to provide early warning of a possible intrusion, so that appropriate measures can be taken to quickly detect before any serious damage is caused. The basic types of intrusion detection systems fall into two categories, signature based and heuristic or anomaly based. Signature based intrusion detection system perform simple pattern matching and detect known attack types. Heuristic intrusion detection techniques identify both known and unknown attacks. Since at times, it is difficult to find the distinction between the behavior of an attacker and authorized user, the biggest challenge lies in the effectiveness of an anomaly IDS towards false positives and false negatives.

Bayesian networks are efficient probabilistic directed acyclic graphical models that can be used to build models from variables. They can be applied in different fields such as gene regulatory network, biomonitoring, medicine, document classification, image processing, spam filter, anomaly detection, decision making under uncertainty, etc. In the Bayesian network classifier [2] the assumption is that every variable is independent from the rest of the variables. This technique assigns probability values to each of the variables and defines the dependency among the variables. Let {N1, N2, N3, ………., Nn} be variables which can be represented as nodes. If one variable has dependency on another variable, then an arc is drawn from one variable to another which represent direct correlation between the variables. Bayesian networks are popular methods for modeling uncertain and complex domains which can be used to build a robust and mathematically coherent framework for analysis. The main aim of this paper is to experimentally verify the impact of different search based feature selection methods on Bayesian classifier. In this paper we propose to develop an adaptive network intrusion detection system using a Bayesian network to detect unknown intrusion attempts ensuring low false alarms. While learning Bayesian networks from dataset, variables have been used to represent dataset features.

The rest of the paper is organized as follows: The related works done by other authors briefly represented in Section 2. Section 3 describes different techniques based on Bayesian classification. Section 4 presents proposed model adopted in this study. Section 5 divided into three subsections. Section 5.1 describes NSL-KDD dataset on which the experiments are conducted, Section 5.2 briefly describe different search based feature selection methods, Section 5.3 briefly describes confusion matrix. Section 6 describes the conducted experiments and summarises the results of the proposed model. Section 7 makes some concluding results and proposes for future research.

II. RELATED WORK

Feature selection is one of the prior requirements to deal with huge data sets in order to select only those features which are useful for further processing. Several techniques have been proposed for feature selection in order to find the minimum set of features in a dataset. An efficient and effective model has been developed using feature selection methods and C4.5 classification techniques [3]. Four different feature selection techniques namely, relief-F, correlation, info gain, and symmetrical uncertainty have been applied on NSL-KDD dataset to select important features. Experimental results show that C4.5 with info gain feature selection gives highest accuracy of 99.68% with 17 features.

A hybrid intrusion detection mechanism has been proposed based on binary particle swarm optimization (PSO) and random forest (RF) technique called PSO-RF [4]. Binary PSO is used to select most important features from the NSL-KDD dataset and RF is used as a classifier. Experimental results show that the average IDR and average FPR values are much better as compared to other techniques used.

A new learning technique has been proposed for developing a novel intrusion detection system using modified k-means algorithm [5] . KDD CUP 99 dataset is used to analyze the performance of the proposed model. Results show that high efficiency is achieved in attack detection and accuracy which is 95.75%.

An adaptive and robust intrusion detection system has been proposed using Hypergraph based Genetic Algorithm (HG-GA) for parameter setting and support vector machine for feature selection [6] . For performance analysis NSL-KDD dataset is used under two conditions: by taking all the features and by considering only relevant features obtained from HG-GA. Experimental results show the prominence of HG-GA SVM over the existing techniques in terms of classifier accuracy, detection rate, false alarm rate, and runtime analysis.

An effective intrusion detection has been proposed [7] based on support vector machine with augmented features. On experimenting with NSL-KDD dataset,the results show better performance in terms of detection rate, accuracy, and low false alarm rate.

III. METHODOLOGY

Here, we discuss various Bayesian network based techniques which we have used as classifiers.

3.1. Hill Climbing Search

Hill climbing search [8] begins with an initial network, i.e., an empty network or a randomly generated structure and repeatedly apply single edge operations, including addition, deletion, and reversal until a locally optimal network is found. The search is not restricted by an order on the variables.

Hill Climbing Search Algorithm

```
Given, Data set D, Initial network X_0
    j = 0
    X_best ← X_0
    while stopping criteria not met
      {
        for each possible operator application, b
          {
            X_new ← apply( b, X_j )
            if score (X_new) >score(X_best)
            X_best ← X_new
          }
         ++j
        X_j ← X_best
      }
```

3.2. K2 Search Algorithm

The K2 search Algorithm [9] is a greedy search algorithm that learns the network structure of the Bayesian network from the data presented to it. It attempts to select the network structure that maximizes the networks posterior probability given the experimental data. The K2 algorithm reduces this computational complexity by requiring a prior ordering of nodes as an input, from which the network structure can be constructed. The ordering is such that if node Y_i comes prior to node Y_j in the ordering, then node Y_j cannot be a parent of node Y_i. In other words, the potential parent set of node Y_i can include only those nodes that precede it in the input ordering.

3.3. Tabu Search

Tabu Search is a meta-heuristic strategy that is able to guide traditional local search methods to escape the trap of local optimality with the assistance of adaptive memory [10]. Tabu Search's adaptive memory feature allows the implementation of procedure that are capable of searching the solution space more effectively.

Tabu Search Algorithm

Step 1: Select an initial solution $y \in Y$, and let $y^* = y$ and $y_0 = y$,
Set iteration counter m = 0 and Tabu list TL = Ø.
Step 2: If S – TL = Ø, then go to Step 4;
else m = m + 1 and select $s_m \in S - TL$ such that $s_m(y_m - 1) = \text{OPTIMUM}(s(y_m - 1)$: $s_m \in S - TL$)
Step 3: Let $y_m = s_m(y_m - 1)$. If $c(y_m) < c(y^*)$ where y^* denotes the best solution currently found, Let $y^* = y_m$.
Step 4: If a chosen number of iterations has elapsed either in total number or since y^* was last improved or S − TL = Ø; upon reaching this step from step 2, stop.
Otherwise, update TL and return to Step 2.
Tabu list (TL) is given by
$TL = \{s^{-1} : s = s_r, \; r > m - t, \}$ where m is the iteration index and s^{-1} is the inverse of the move s; i.e.,$s^{-1}(s(y)) = y$. TL is the set of those moves that would undo one of those moves in the t most recent iterations. It is called the Tabu tenure.

3.4. Tree Augmented Naive Bayesian (TAN)

In a TAN model, all the variables are connected to the class variables using direct edges. Hence, it takes into account all the variables while determining $P(C \mid Y_1,...,Y_n)$. Also each variable can be connected to another variable in the network [11]. The computational complexity of this model is very low because each variable has a maximum of two parents. Thus TAN maintains the robustness and computational complexity of the Naive Bayes model and also displays better accuracy.

The tree construction procedure consists of four steps [11] :
Step 1. Compute $I_p(Y_i ; Y_j \mid C)$ between each pair of attributes $i \neq j$.
Step 2. Build a complete undirected graph in which the vertices are the attributes $Y_1,...,Y_n$. And annotate the weight of an edge connecting Y_i to Y_j by $I_p(Y_i ; Y_j \mid C)$.
Step 3. Build a spanning tree of maximum weight.
Step 4. The resulting undirected tree transform to directed tree by choosing a root variable randomly and setting the direction of all the edges outward from the root.

IV. THE PROPOSED MODEL

The objective of the proposed model is to apply different Bayes net based classifiers to build intrusion detection system that exhibit high detection rate and low false alarm rate. The overall model is depicted in figure 1.

Step 1: Load NSL-KDD dataset with all features.

Step 2: Apply six search based feature selection methods namely, bio-inspired search based, informed search based, random search, vote harmony search, EDA search, and Rank search feature selection methods for finding relevant important features.

Step 3: Different Bayes net based classifiers are applied on selected relevant features of the dataset for testing using 10-fold cross validation.

Step 4: Evaluate the model by comparing the performance in terms of accuracy, precision, detection rate, false alarm rate and efficiency.

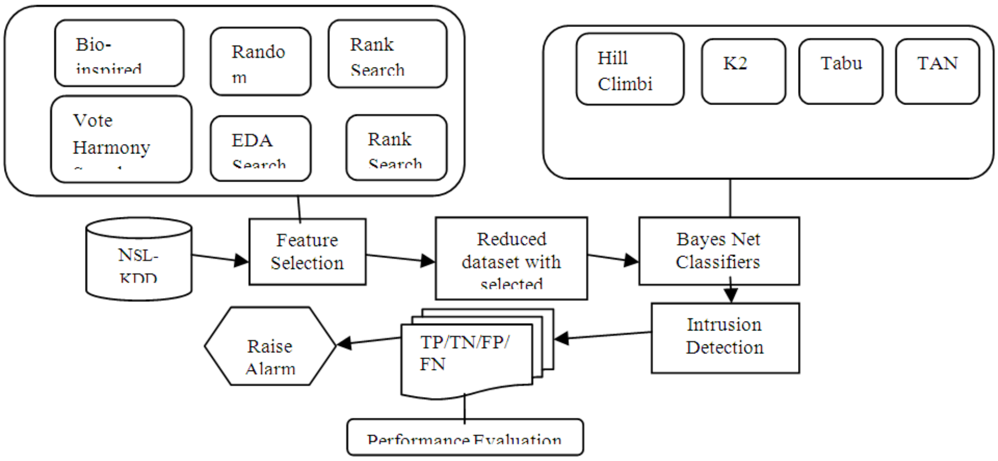

Fig. 1 Proposed Model

V. EXPERIMENTAL SET UP

5.1 . NSL-KDD Dataset

The NSL- KDD intrusion dataset is a refined version of KDD CUP 99 dataset [12] has been used for our experimentation. The data set consists of 41 feature attributes out of which 38 are numeric and 3 are symbolic. The total number of records in the data set is 125973 out of which 67343 (53.48%) are normal and 58630 (46.52%) are attacks. The attacks which fall into 24 different types, and can be classified into four attack categories namely, Denial of Service (DOS:36.45%), Remote to Local(R2L:0.78%),User to Root (U2R: 0.04%), and Probing(9.25%) as depicted in figure 2. In DoS attacks, attacker makes some computing/memory resources too busy or too full to handle legitimate requests, or denies legitimate users access to a machine, e.g. syn flood, Neptune, Smurf, Pod and Teardrop. In Remote to Local (R2L) attack, the attacker who does not have an account on a remote machine sends packets to that machine over a network and exploits some vulnerability to illegally gain
local access as a user of that machine, e.g. guessing password, Ftp-write, Imap and Phf. In User to Root (U2R) attack, an attacker starts out with access to a normal user account on the system and is able to exploit system vulnerabilities to gain root access to the system, e.g. Buffer-overflow, Load- module, Perl and Spy. In Probing, an attacker scans a network of computers to gather information or find known vulnerabilities. An attacker with a map of machines and services that are available on a network can use this information to look for exploits, e.g., port scanning, Portsweep, IPsweep, Nmap and Satan.

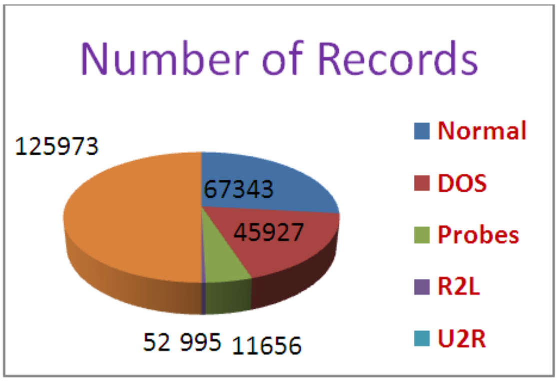

Fig.2 Distribution of Records

5.2 . Feature Selection

The unpredictable behavior, nonlinear character of intrusion attempts and a large number of features in the problem makes intrusion detection a difficult task. Identifying important and key features in the dataset which can help in detecting intrusions is essential. Therefore, the use of suitable feature selection methods to identify and remove irrelevant and redundant features from the dataset that do not contribute to the accuracy of a predictive model is crucial. Feature selection methods have several advantages [13] such as improving the performance of the machine learning algorithms, data understanding, gaining knowledge about the process and helping to visualize it, data reduction, limiting storage requirements, and helping in reducing processing costs. In this work three bio-inspired based feature selection methods namely, ant search, genetic search, PSO search, two informed search based feature selection methods namely, best first search and greedy stepwise search, and random search method have been employed to select the important features. Bio-inspired algorithms [14] are based on the principles of the behavior or the phenomena in living organisms and creatures, such as gene evolution, insect swarming, bird swarming, food foraging, and the like. Bio-inspired algorithms are well known for their applicability to optimization problems. Each individual in a bio-inspired algorithm represents a candidate solution to the problem, and the algorithm converges to the optimal solution (under certain assumptions) through the evolutionary interactions of the individuals in the solution space. Heuristic search [15] has Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS). SFS starts from an empty set. Each time a feature is added to the feature subset so that the evaluation metric could be optimized. SBS starts from the universal set and delete a feature each time. Both SFS and SBS are greedy algorithms that are likely to fall into the local optimum. Heuristic search provides the direction regarding the solution. When no start state is supplied, random search starts from a random point and reports the best subset found. If a start state is supplied, then the technique searches randomly for subsets that are as good or better than the start point with the same or fewer attributes. Vote harmony search [16] performs a random search in the space of feature subsets. If no start state is supplied, the search method starts from a random point and selects the best subsets. EDA search method selects the best features from the data set using estimation of distribution algorithms. From the estimated distributed algorithm, the feature ranking set is derived. Rank search [17] feature selection method rank all features of the dataset. If a subset evaluator is specified, then a forward selection search is used to generate the rank of all features. From the rank list of features, subsets of increasing size are evaluated. Finally, the best feature subset is selected.

5.3 . Confusion Matrix

Intrusion detection systems mainly discriminate between two classes, attack class and normal class. The confusion matrix reports the number of False Positives (FP), False Negatives (FN), True Positives (TP), and True Negatives (TN). True Positives (TP) is the number of attacks that are detected successfully and alarm is raised. False positives (FP) is the number of normal records wrongly detected as attacks and false alarm is raised. True Negatives (TN) is the number of normal records detected as normal and alarms are not raised. False Negative (FN) is the number of attack records detected as normal

Based on these values the following performance measurements can be made:

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FN+FP}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Detection Rate or Recall} = \frac{TP}{TP+FN}$$

$$\text{False Alarm Rate} = \frac{\text{FP}}{\text{TN+FP}}$$

$$\text{Efficiency} = \frac{Total\ Detected\ Attack}{Total\ Attack} \times 100$$

$$\text{Rate of Attack} = \frac{\text{Number of attack detected correctly}}{\text{Total number of attacks}}$$

VI. RESULT ANALYSIS

Different combinations of four Bayes net classifiers namely, Hill Climbing search, K2 search, Tabu Search, and Tree Augmented Naive-Bayes with six categories of search based feature selection methods namely, random search, bio-inspired based search and informed based search, vote harmony search, EDA search, and rank search methods were applied on the NSL-KDD dataset. The performance of different classifiers is evaluated on the basis of rate of attack of four different types of attacks, accuracy, precision, detection rate, and false alarm rate. 10-Fold cross validation has been used for training and testing. Table 1 depicts the attack rate of four attacks namely, DOS, R2L, U2R, and probes.

Table 1 Attack rate of four different attacks

Feature Selection Method	Classifiers	Rate of Attack in %			
		DOS	R2L	U2R	Probes
Ant Search	Bayesnet + Hill Climbing	90.4326	78.995	13.4615	96.594
	Bayesnet + K2	90.4326	78.995	13.4615	96.5940
	Bayesnet + Tabu Search	90.4348	78.4925	7.6923	96.7313
	Bayesnet + TAN	**99.6756**	**75.9799**	**44.2308**	**98.7045**
Genetic Search	Bayesnet + Hill Climbing	94.7482	94.4724	36.5385	98.4042
	Bayesnet + K2	94.7482	94.4724	36.5385	98.4042
	Bayesnet + Tabu Search	94.7482	94.4724	36.5385	98.4042
	Bayesnet + TAN	**99.8345**	**96.0804**	**53.8461**	**98.6016**

PSO Search	Bayesnet + Hill Climbing	90.2345	89.2462	3.8461	97.4948
	Bayesnet + K2	90.2345	89.2462	3.8461	97.4948
	Bayesnet + Tabu Search	90.2345	89.2462	3.8461	97.4948
	Bayesnet + TAN	**99.8171**	**94.1708**	**9.6154**	**99.0048**
Best First Search	Bayesnet + Hill Climbing	97.0453	95.2764	65.3846	98.4386
	Bayesnet + K2	97.0453	95.2764	42.7515	98.4386
	Bayesnet + Tabu Search	97.3392	94.3718	50	97.7865
	Bayesnet + TAN	**99.9412**	**95.7789**	21.1538	**98.6187**
Greedy Stepwise Search	Bayesnet + Hill Climbing	96.4901	94.4724	48.0769	97.9839
	Bayesnet + K2	96.4901	94.4724	48.0769	97.9839
	Bayesnet + Tabu Search	96.4552	94.3718	9.6154	97.9753
	Bayesnet + TAN	**99.9412**	**96.0808**	23.0769	**98.593**
Random Search	Bayesnet + Hill Climbing	96.8319	93.6683	50	97.8895
	Bayesnet + K2	96.8319	93.6683	50	97.8895
	Bayesnet + Tabu Search	96.8319	93.6683	9.6154	97.8723
	Bayesnet + TAN	**99.5645**	**96.0804**	15.3846	**98.8246**
Vote Harmony Search	Bayesnet + Hill Climbing	99.5819	93.0653	46.1538	98.0782
	Bayesnet + K2	99.4883	93.3668	51.9231	97.8809
	Bayesnet + Tabu Search	99.5819	93.0653	46.1538	98.0782
	Bayesnet + TAN	**99.8589**	**95.4774**	**59.6154**	**98.3699**
EDA Search	Bayesnet + Hill Climbing	99.6473	92.8643	38.4615	95.5216
	Bayesnet + K2	98.9265	92.8643	46.1538	96.2766
	Bayesnet + Tabu Search	99.6473	92.8643	38.4615	95.5216
	Bayesnet + TAN	**99.6712**	**94.0703**	**48.0769**	**96.4739**
Rank Search	Bayesnet + Hill Climbing	99.6647	94.3718	46.1538	97.9667
	Bayesnet + K2	95.1357	94.4724	48.0769	96.9629
	Bayesnet + Tabu Search	99.6647	94.3718	46.1538	97.9667
	Bayesnet + TAN	**99.8106**	**95.5779**	**55.7692**	**98.4386**

In Table 1,it is observed that TAN classifier has better attack rate in comparison to other Bayes net classifiers irrespective of the feature selection techniques used.

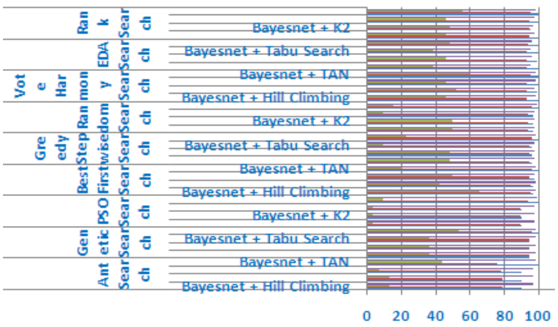

Fig. 3 Rate of attack of different attack types

Table 2 depicts the performance of four Bayes Net techniques with random search feature selection method. The criteria are accuracy, precision, detection rate, false alarm rate, and efficiency. TAN classification demonstrates the lowest false alarm rate of **0.4395%** and highest detection rate of **99.6862%.**

Table 3 depicts the performance of four Bayes Net techniques with Bio-inspired feature selection methods. TAN classification with genetic search feature selection gives the lowest false alarm rate of **0.2792%** and highest detection rate of **99.6725%.**

Table 2 Performance Comparison of Bayes Net Classifiers using Radom Search Feature Selection Method

Feature Selection Method	Classifier Techniques	Evaluation Criteria				
		Accuracy in %	Precision in %	Recall / Detection Rate in %	False Alarm Rate in %	Efficiency in %
Random Search	Bayesnet + Hill Climbing	98.2369	98.1856	98.0232	1.577	96.9469
	Bayesnet +	98.2369	98.1856	98.1856	1.577	96.9469

	K2					
	Bayesnet + Tabu Search	98.298	98.2489	98.0914	1.522	96.9077
	Bayesnet + TAN	**99.619**	**99.4961**	**99.6862**	**0.4395**	**99.2836**

Table 3 Performance Comparison of Bayes Net Classifiers using Bio-inspired Search based Feature Selection Method

Feature Selection Method	Classifier Techniques	Evaluation Criteria				
		Accuracy in %	Precision in %	Recall / Detection Rate in %	False alarm Rate in %	Efficiency in %
Ant Search	Bayesnet + Hill Climbing	95.5927	98.1686	92.2514	1.4983	91.3952
	Bayesnet + K2	95.5927	98.1686	92.2514	1.4983	91.3952
	Bayesnet + Tabu Search	93.7463	94.3962	92.0263	4.7562	91.4105
	Bayesnet + TAN	**99.2562**	**99.2437**	**99.1574**	**0.6578**	**99.0312**
Genetic Search	Bayesnet + Hill Climbing	97.1637	97.042	96.8583	2.5704	95.4187
	Bayesnet + K2	97.1637	97.042	96.8583	2.5704	95.4187
	Bayesnet + Tabu Search	97.1637	97.042	96.8583	2.5704	95.4187
	Bayesnet + TAN	**99.6983**	**99.6793**	**99.6725**	**0.2792**	**99.485**
PSO Search	Bayesnet + Hill Climbing	97.7051	96.823	98.2944	2.808	91.5845
	Bayesnet + K2	97.7051	96.823	98.2944	2.808	91.5845
	Bayesnet + Tabu Search	97.7051	96.823	98.2944	2.808	91.5845
	Bayesnet + TAN	**99.4014**	**99.1474**	**99.5702**	**0.7454**	**99.48**

Table 4 Performance Comparison of Bayes Net Classifiers using Informed Search based Feature Selection Method

Feature Selection Method	Classifier Techniques	Evaluation Criteria				
		Accuracy in %	Precision in %	Recall / Detection Rate in %	False Alarm Rate in %	Efficiency in %
Best First Search	Bayesnet + Hill Climbing	98.1075	97.9097	98.0266	1.822	97.2642
	Bayesnet + K2	98.1075	97.9097	98.0266	1.822	97.2642
	Bayesnet + Tabu Search	98.125	97.9252	98.0488	1.8087	97.3358
	Bayesnet + TAN	**99.5277**	**99.2849**	**99.7032**	**0.6251**	**99.5378**
Greedy	Bayesnet +	98.4132	98.3935	98.1938	1.3958	96.7099

Stepwise Search	Hill Climbing					
	Bayesnet + K2	98.4132	98.3935	98.1938	1.3958	96.7099
	Bayesnet + Tabu Search	98.4362	98.4505	98.1852	1.3454	96.6451
	Bayesnet + TAN	**99.5332**	**99.2917**	**99.7083**	**0.6192**	**99.5395**

Table 5 Performance Comparison of Bayes Net Classifiers using Vote Harmony Search Feature Selection Method

Feature Selection Method	Classifier Techniques	**Evaluation Criteria**				
		Accuracy in %	Precision in %	Recall / Detection Rate in %	False Alarm Rate in %	Efficiency in %
Vote Harmony Search	Bayesnet + Hill Climbing	99.0911	98.5441	99.5173	1.28	99.125
	Bayesnet + K2	99.1093	98.6217	99.4764	1.2102	99.0227
	Bayesnet + Tabu Search	99.0911	98.5441	99.5173	1.28	99.125
	Bayesnet + TAN	**99.6372**	**99.5705**	**99.6503**	**0.3742**	**99.4525**

Table 6 Performance Comparison of Bayes Net Classifiers using EDA Search Feature Selection Method

Feature Selection Method	Classifier Techniques	**Evaluation Criteria**				
		Accuracy in %	Precision in %	Recall / Detection in %	False Alarm Rate in %	Efficiency in %
EDA Search	Bayesnet + Hill Climbing	99.1562	98.7798	99.415	1.0691	98.8948
	Bayesnet + K2	99.0664	98.6601	99.3433	1.1746	98.25
	Bayesnet + Tabu Search	99.1562	98.7798	99.415	1.0691	98.6577
	Bayesnet + TAN	**99.5682**	**99.597**	**99.4747**	**0.3504**	**98.8948**

Table 7 Performance Comparison of Bayes Net Classifiers using Rank Search Feature Selection Method

Feature Selection Method	Classifier Techniques	**Evaluation Criteria**				
		Accuracy in %	Precision in %	Recall / Detection in %	False Alarm Rate in %	Efficiency in %
Rank Search	Bayesnet + Hill Climbing	98.9776	98.2725	99.5531	1.5235	99.1898
	Bayesnet + K2	98.3854	98.3264	98.2021	1.4552	95.446
	Bayesnet + Tabu Search	98.9776	98.2725	99.5531	1.5235	99.1898
	Bayesnet + TAN	**99.6213**	**99.4363**	**99.5974**	**0.4908**	**99.4269**

Table 4 depicts the performance of four Bayes Net techniques with informed search feature selection method. TAN classification with greedy stepwise search feature selection gives the lowest false alarm rate of **0.6192%** and highest detection rate of **99.7083%.**

Table 5 depicts the performance of four Bayes Net techniques with vote harmony search feature selection method. It is observed that TAN classification with vote harmony search feature selection gives the lowest false alarm rate of **0.3742%** and highest detection rate of **99.6503%.**
Table 6 depicts the performance of four Bayes Net techniques with EDA search feature selection method. It is evident that TAN classification with EDA search feature selection gives the lowest false alarm rate of **0.3504%** and highest detection rate of **99.4747%.**

Table 7 depicts the performance of four Bayes Net techniques with rank search feature selection method. It is observed that TAN classification with rank search feature selection gives the lowest false alarm rate of **0.4908%** and highest detection rate of **99.5974%.**

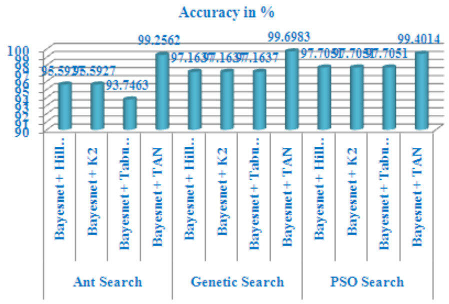

Fig.4 Comparison of accuracy among the classifiers

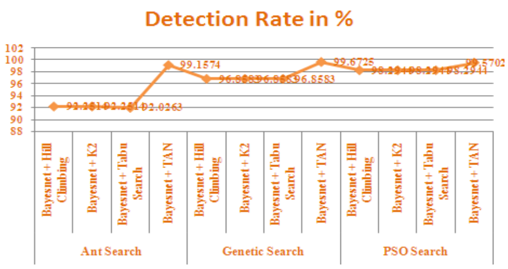

Fig.5 Comparison of Detection Rate among the classifiers

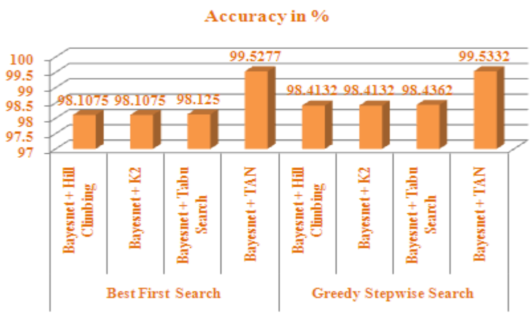

Fig.6 Comparison of accuracy among the classifiers

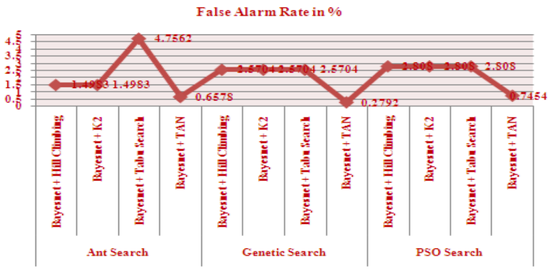

Fig.7 Comparison of False Alarm Rate among the classifiers

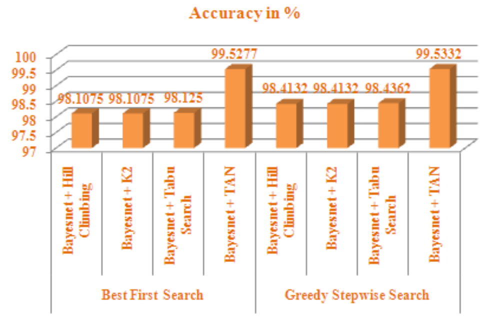

Fig.8 Comparison of accuracy among the classifiers

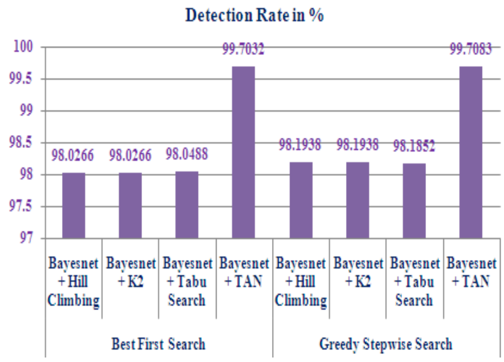

Fig. 9 Comparison of Detection Rate among the classifiers

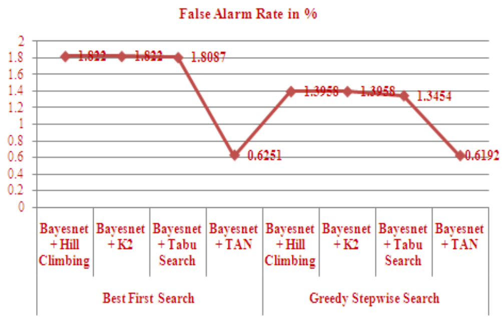

Fig.10 Comparison of False Alarm Rate among the classifiers

Table 8 Comparative analysis of our results with that of results obtained by other authors/works (NR: Not Reported)

Author	Feature Selection Method	Classifier Techniques	Accuracy in %	Detection Rate / Recall in %	False Alarm Rate in %	Dataset
[18]	Information Gain	SSPV-SVDD	NR	77.5	NR	NSL-KDD
[19]		Fuzzy Genetic Algorithm	98.2	NR	0.5	NSL-KDD
[20]	Ant Colony Optimization + Feature weighting SVM	SVM	NR	95.75	NR	KDD CUP 99
[21]	Cuttle fish algorithm	C 5.0 + One class SVM	98.20	Nr	1.405	NSL-KDD
[22]	Intelligent Water Drop (IWD)	SVM	NR	99.40	1.405	KDD CUP 99
[23]	Filter based	SVM	99.94	NR	NR	NSL-KDD
[24]	Hybrid Kernel PCA + GA	Multilayer SVM	NR	99.22	NR	KDD CUP 99
[25]	Wrapper subset evaluator	Fuzzy Rough NN	97.513	95.139	0.4202	NSL-KDD
Present Work	**Genetic Search**	**TAN**	**99.6983**	**99.6725**	**0.2792**	**NSL-KDD**

VII. CONCLUSION

In this paper, we have discussed the need for suitable intrusion detection systems to guard against malicious attempts to access network resources. In an attempt to develop an intrusion detection model, we have used four Bayes net based classifiers and six different categories of search based feature selection methods to find out the suitability of the most effective combination. The performance of each combination was measured in terms of various evaluation criteria on the NSL-KDD intrusion dataset. The results indicate that TAN classification with genetic search feature selection emerges as the best combination with the lowest false alarm rate of **0.2792%.** In future we wish to explore other classification methods along with various feature selection techniques.

REFERENCES

[1]. Symantec, Internet security threat report 2017 April, 2017, Available: https://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017, vol.22, 2017

[2]. C.M. Rahman, D.Md. Farid, and M.Z. Rahman (2011) Adaptive Intrusion Detection based on Boosting and Naïve Bayesian Classifier. International Journal of Computer Applications, 24(3), 12-19

[3]. H. Hota and A.K. Shrivas, (2014). Decision tree techniques applied on NSL-KDD data and its comparison with various feature selection techniques. Advanced Computing Networking and Informatics , 205-211.

[4]. J. Malik, W. Shahzad and F. A. Khan (2012), Network intrusion detection using hybrid binary PSO and random forests algorithm. Security and Communication Networks.

[5]. AI. Yaseen, Z.A. Othman, and M.Z.A. Nazri, (2017). Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Systems with Applications , 296-303.

[6]. M.R.G. Raman, N. Somu,, K., Kirthivasan, R.. Liscano,, V.S.S. Sriram, (2017). An efficient intrusion detection system based on Hypergraph-Genetic algorithm for parameter optimization and feature selection in support

vector machine. Knowledge based System, 1-12.

[7]. H. wang, J. Gu, and S. Wang (2017). An effective intrusion detection framework based on SVM with feature augmentation. Knowledge based system..

[8]. W.L. Buntie (1996). A guide to the literature on learning probabilistic networks from data. IEEE Transactions on Knowledge and Data Engineering, 195-210.

[9]. G.F. Cooper and E. Herskovits (1992). A Bayesian method for the induction of probabilistic networks from data. Machine Learning, 309-347

[10]. F. Glover (1989). Tabu Search Part-1. Informs Journal on Computing. 1(3), 190-206.

[11]. N. Friedman, D. Geiger, and M. Goldszmidt, (1997) Bayesian Network Classifiers. Machine Learning, 131-163.

[12]. M. Tavallaee, E. Bagheri, W. Lu,, and Ghorbani (2009). A detailed analysis of the KDD CUP 99 Dataset. Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defence Applications , 1-6.

[13]. I., Guyon, S. Gunn, M. Nikravesh, L. Zadeh, (2006). Feature Extraction, Foundations and Applications Springer.

[14]. S. Olariu, and A.Y. Zomaya, (2006). Handbook of Bioinspired Algorithms and Applications, CRC Press, Boca Raton, Taylor and Francis Group.

[15]. P.A. Devijver and J. Kittler (1982). Pattern Recognition: A Statistical Approach. Prentice Hall.

[16]. H. Liu and Setiono (1996). A probabilistic approach to feature selection – A filter solution. In 13th International Conference on Machine Learning.

[17]. M. Hall, G. Holmes (2003). Benchmarking attribute selection techniques for discrete class data mining. IEEE Transactions on Knowledge and Data Engineering, 15(6), 1437-1447.

[18]. M.EI. Boujnouni (2018). New intrusion detection system based on support vector domain description with information gain metric. International Journal of Network Security, 25-34.

[19]. G., Javadzadeh and R. Azmi (2015). Introducing an intrusion detection using hybrid fuzzy genetic approach. International Journal of Network Security , 754-770.

[20]. W. Xingzhu (2015). ACO and SVM selection feature weighting of network intrusion detection method. International Journal of Security and Its Applications,129-270.

[21]. M.S. Rani, S.B. Xavier (2015). A hybrid intrusion detection based on C5.0 decision tree and one-class SVM. International Journal of Current Engineering and Technology, 2001-2007.

[22]. N. Acharya, S. Singh (2018). An IWD-based feature selection method for intrusion detection system. Soft Computing, 4407-4416.

[23]. M.A. Ambusaidi, X. He, P. Nanda, and Z. Tan, (2016). Building an intrusion detection system using a filter-based feature selection algorithm. IEEEE Transactions on Computers , 2986-2998

[24]. F. Kuang, W., Xu and S. Zhang (2014). A novel hybrid KPCA and SVM with GA model for intrusion detection. Applied Soft Computing, 178-184.

[25]. A. Panigrahi, M.R. Patra (2018) .Rough set based network intrusion detection with Wrapper subset evaluator. International Journal of Engineering Science Invention (IJESI) , 51-57.