Anomaly Based Network Intrusion Detection Using Bayes Net Classifiers

Ashalata Panigrahi, Manas Ranjan Patra

Abstract: Ever increasing cybercrimes have posed a real challenge for ensuring secured internet based applications be it military, government, e-commerce, bank, or any other sensitive applications. Therefore, along with the development of innovative applications, security systems are being designed in the same pace to deal with cybercrimes. Intrusion Detection System (IDS) plays an effective role to achieve higher security in detecting malicious activities for a couple of years. Anomaly based intrusion detection approaches have the advantage of being able to detect previously unknown network attacks, but they suffer from the difficulty of building robust models of acceptable behavior which may result in a large number of false alarms. Developing flexible and adaptive security oriented approaches is a severe challenge. Bayesian belief networks are popular for adaptive learning. In this work, bayes net classifier techniques have been proposed for building a robust and accurate IDS using four classifiers, namely, K2 search, Tabu Search, Hill Climbing search, Tree Augmented Naive-Bayes, Further, three entropy based feature selection methods, viz., Information Gain, Gain ratio, Symmetrical Uncertainty, three statistical based methods, viz., Chi squared Attribute Evaluator, Relief-F, and one-R have been employed to select the most relevant features before classification. The performance of the model has been evaluated in terms of accuracy, precision, detection rate, F-value, and false alarm rate.

Index Terms: Adaptive learning, Bayesian networks, Directed acyclic graph, Feature selection, Hill climbing, Probabilistic reasoning, Tabu search.

1. INTRODUCTION

Now-a-days the fast rising networks proliferation, data transfer rate, and an unpredictable Internet usage have added more anomaly problems. Yahoo, Amazon.com, eBay, and some other popular World Wide Web (WWW) sites were targets of what appears to have been a coordinated "denial-of-service" attack. At the same time, every organization that uses computers faces the threat of hacking from individuals within the organization. Employees or former employees with malicious intent or who want to obtain information such as employee salaries or view other employee's files are also a threat to an organization's computers and networks. Many methods have been developed by organizations and play very important roles to secure network infrastructure and communications via the Internet such as through the use of firewalls, anti-virus software packages and intrusion detection systems. Current firewalls cannot defend against every category of intrusion, whereby some intrusions take advantages of computer system vulnerabilities [1] . Various Intrusion detection systems (IDSs) have been widely used to overcome security threats in computer networks. The main goal of an intrusion detection system is to detect the attacks efficiently. Furthermore, it is equally important to detect attacks at a beginning stage in order to reduce their impacts. Anomaly-based approaches have the advantage of being able to detect previously unknown attacks, but they suffer from the difficulty of building robust models of acceptable behaviour which may result in a large number of false positives and false negatives caused by incorrect classification of events in current systems. for network intrusion detection. This paper proposes an intrusion detection method based on Bayesian network.

- *Ashalata Panigrahi, Roland Institute of Technology, berhampur, India, ashalata.panigrahi@yahoo.com*
- *Manas Ranjan Patra, Berhampur University, Berhampur, India, mrpatra12@gmail.com*

Bayesian networks [2] are directed acyclic graphs that allow efficient and effective representation of the joint probability distribution over a set of random variables. Each vertex in the graph represents a random variable, and edges represent direct correlations between the variables. More precisely, the network encodes the following conditional independence statements: each variable is independent of its nondescendants in the graph given the state of its parents. These independencies are then exploited to reduce the number of parameters needed to characterize a probability distribution, and to efficiently compute posterior probabilities given evidence. Probabilistic parameters are encoded in a set of tables, one for each variable, in the form of local conditional distributions of a variable given its parents. Using the independence statements encoded in the network, the joint distribution is uniquely determined by these local conditional distributions. When learning Bayesian networks from dataset, we use nodes to represent dataset features.

2. RELATED WORK

Mazzuchi et al.[3] proposed an extended version of Naïve Bayes called Hidden Naïve Bayes (HNB) classifier to classify network attacks correctly. The authors integrate HNB with various discretization and feature selection methods to increase other methods such as decision tree, neural network, and sequential minimal optimization. Mukherjee et al., [4] studied the importance of reduced input features in building IDS that is computationally efficient and effective. They have investigated the performance of three standard feature selection methods using Correlation-based Feature Selection (CFS), Information Gain (IG), and Gain Ratio (GR). They have proposed the Feature Vitality Based Reduction Method (FVBRM) to identify important reduced input features and have applied the Naive Bayes classifier on the reduced intrusion dataset. Empirical results show that feature subset identified by CFS has improved Naive Bayes classification accuracy as compared to IG and GR. Further, FVBRM method shows considerable improvement on classification accuracy as compared to CFS but it takes more time.

Upendra et al. [5] proposed classification of intrusion detection based on various machine learning algorithms like J48, Naïve Bayes, One-R, and Bayes Net. They found that the Decision tree algorithm J48 is most suitable which yields high positive rate and low false positive rate. Elngar et al. [6] proposed an effective PSO-Discretize-HNB intrusion detection system which combines Particle Swarm Optimization (PSO) and Information Entropy Minimization (IEM) discretize method with the Hidden Naıve Bayes (HNB) classifier. To evaluate the performance of the proposed network IDS several experiments were conducted on the NSL-KDD network intrusion dataset. A comparative study of applying Information Gain (IG) which is a well-known feature selection algorithm with HNB classifier was accomplished. To validate the classifier, it is also compared with different feature selection methods such as Principal Component Analysis (PCA) and Gain Ratio. The results obtained showed the efficiency of the proposed network IDS which could reduce the number of features from 41 to 11 leading to high intrusion detection accuracy of 98.2%.

3. METHODOLOGY

3.1. K2 Search Algorithm

The K2 search Algorithm [7] is a greedy search algorithm that learns the network structure of the BN 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 will be constructed. The ordering is such that if node Xi comes prior to node Xj in the ordering, then node Xj cannot be a parent of node Xi. In other words, the potential parent set of node Xi can include only those nodes that precede it in the input ordering.

3.2. 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 [8]. Its strategic use of memory and responsive exploration is based on selected concepts that cut across the fields of artificial intelligence and operations research. Tabu search is viewed as “intelligent” search because it makes use of adaptive memory. The adaptive memory feature of TS allows the implementation of procedures that are capable of searching the solution space economically and effectively.

Tabu Search Algorithm:

Step 1: Select an initial solution x Є X, and let x* = x and x_0 = x,

Set iteration counter k = 0 and Tabu list TL = Ø.

Step 2: If S – TL = Ø, then go to Step 4;

else k = k + 1 and select s_k Є S – TL such that s_k (x_k − 1) = OPTIMUM (s(x_k − 1): s_k Є S – TL)

Step 3: Let x_k = s_k(x_k − 1). If c(x_k) < c(x*) where x* denotes the best solution currently found,

Let x* = x_k.

Step 4: If a chosen number of iterations has elapsed either in total number or since x* 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 = si, i> k − t, } where k is the iteration index and s-1is the inverse of the move s; i.e.,s^{-1} (s(x)) = x. In words, 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.

3.3. Hill Climbing Search

Hill climbing search [9] 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 B_0

i = 0

B_{best} ← B_0

while stopping criteria not met

{

for each possible operator application, a

{

B_{new} ← apply (a, B_i)

if score (B_{new}) > score(B_{best})

B_{best} ← B_{new}

}

++i

B_i ← B_{best}

}

3.4. Tree Augmented Naive Bayesian (TAN)

Tree Augmented Naive Bayesian (TAN) model imposes a restriction on the level of interaction between the variables to one. In a TAN model, all the variables are connected to the class variables by means of direct edges. Hence, it takes into account all the variable while determining P(C | X1,...,Xn). In addition to that, each variable can be connected to another variable in the network [10]. That is, each variable in the graph can have two parents viz., the class node and another variable node, except for one variable which is called root. The computational complexity of this model, is greatly reduced, as each variable has a maximum of two parents. "Thus TAN maintains the robustness and computational complexity of the Naive Bayes model and at the same time displays better accuracy". The tree construction procedure consists of 4 main steps [10] :

1. Compute Ip(Xi ; Xj | C) between each pair of attributes i ≠ j.
2. Build a complete undirected graph in which the vertices are the attributes X1,...,Xn. And annotate the weight of an edge connecting Xi to Xj by Ip(Xi ; Xj | C).
3. Build a maximum weighted spanning tree.
4. Transform the resulting undirected tree to a directed one by randomly choosing a root variable and setting the direction of all the edges outward from the root.

4. THE PROPOSED HYBRID MODEL

The objective of the proposed model is to combine different techniques to build a hybrid intrusion detection system which can achieve better accuracy, high detection rate and low false alarm rate . The model as depicted in figure 1 comprises of two levels. The first level consists of feature selection methods with an objective of identifying, and removing irrelevant attributes from the intrusion dataset. Three entropy based methods namely Information gain, Gain Ratio, Symmetrical Uncertainty and three statistical based methods namely Chi Squared attribute evaluator, Relief-F, and One-R have been applied for selection of relevant attributes. In the second level the reduced data set obtained from level-1 is classified using four Bayesian network (BN) classifiers such as as K2 search, Tabu Search, Hill Climbing search , and Tree augmented Naive-Bayes.

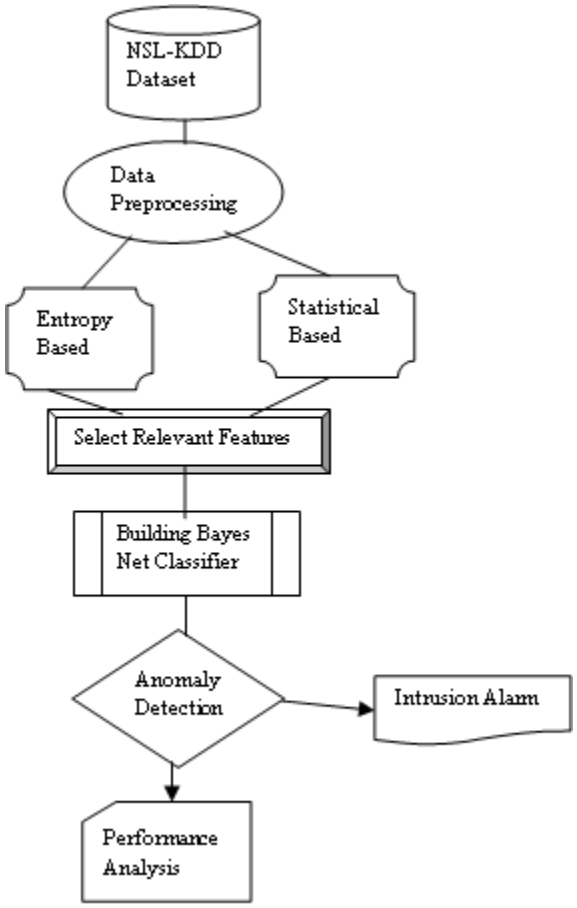

Fig.1 *The proposed hybrid model*

5. EXPERIMENTAL SETUP

5.1. NSL-KDD Dataset

The NSL- KDD dataset proposed by Tavallace et al. [11] is a reduced version of the original KDD CUP 99 dataset. NSL-KDD dataset consists of same features as KDD CUP 99 dataset but has some advantages over the original KDD'99 dataset. The NSL-KDD intrusion dataset which consists of 41 feature attributes has been used for our experimentation. The total number of records in the data set is 125973 out of which 67343 are normal and 58630 are attacks. The dataset contains 24 different attack types which can be classified into four main categories namely, Denial-of-Service (DOS), Probe, Remote-to-Local (R2L), and User-to-Root (U2R). Each record is categorized as normal or attack with exactly one particular type of attack. DOS: In this type of attack an attacker makes some computing/memory resources too busy or too full to handle legitimate requests, or denies legitimate users access to a machine. There are different ways to launch DOS attacks: by abusing the computers legitimate features; by targeting the bugs; or by exploiting the system's misconfiguration. Probe: In this type of attacks 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. There are different types of probes; some of them abuse the computer's legitimate features; some of them use social engineering techniques. This class of attacks is the most commonly heard and requires very little technical expertise. R2L: In this type of attack an 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. U2R: In this type of attacks 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. Most common exploits in this class of attacks are regular buffer overflows, which are caused by regular programming mistakes and environment assumptions.

5.2. Feature Selection

In problems dealing with high dimensional feature space, some of the features may be redundant or irrelevant. Removing these redundant or irrelevant features is very important for effectiveness. Effective feature selection is an important research challenge for classifying intrusion data as normal or intrusive. This has tremendous impact on the performance of any intrusion detection system. Here three entropy based and three statistical based feature selection methods are applied for selection of important features.

5.3. Cross Validation

Cross validation calculates the accuracy of the model by separating the data into two different populations, a training set and a testing set. We have adopted a 10-fold cross-validation process wherein the dataset is randomly partitioned into 10 mutually exclusive approximately equal size folds; T1, T2, …..,T10 out of which 9 folds are used to training and the remaining for testing. The process is repeated 10 times in a turn-taking manner by changing the folds. The 10 sets of results thus obtained are averaged to produce a single model estimation.

5.4. Performance Measurement

The performance of an intrusion detection system is evaluated by its ability to make accurate prediction of attacks. Intrusion detection systems mainly discriminate between two classes, attack class (abnormal data), and normal class (normal data). Here, a confusion matrix has been built to measure the performance of different classifiers in terms of accuracy, precision, detection rate, F-value, and false alarm rate. Essentially, a Confusion matrix

is a tabular representation of false positives (FP), false negatives (FN), true Positives (TP), and true negatives (TN).

Table: 1 Confusion Matrix

		Predicated Class	
		Normal	Attack
Actual Class	Normal	True Negative (TN)	False Positive (FP)
	Attack	False Negative (FN)	True Positive (TP)

True Positive (TP): An actual attack is successfully detected by the IDS.
True Negative (TN): No attack has taken place and no IDS alert is raised.
False Positive (FP): An alarm/alert that indicates that an attack is in progress but in fact there was no such attack.
False Negative (FN) : A failure of IDS to detect an actual attack .
Next, we define some of the measures which are used to evaluate the performance of different classifiers using the values obtained from the confusion matrix.
Accuracy measures the probability that an algorithm can correctly predict positive and negative examples. Accuracy is calculated as:

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

Precision is a measure of the accuracy provided that a specific class has been predicted and it is calculated as:

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

Detection rate measures the probability that one can correctly predict positive scenarios:

Detection Rate = $\frac{TP}{TP+FN}$

F- value is the harmonic mean of Precision and Recall which measures the quality of classification:

F - Value = 2 * $\frac{(Precision * Recall)}{(Precision + Recall)}$

False Alarm Rate is the ratio of the number of normal instances incorrectly labelled as intrusion and the total number of normal instances:

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

6. RESULT ANALYSIS

Different combinations of four Bayes net classifiers namely, K2 search, Tabu Search, Hill Climbing search, and Tree augmented Naive-Bayes with two categories of feature selection methods were applied on the NSL-KDD dataset. The performance of different classifiers are evaluated on the basis of accuracy, precision, detection rate, F-value, and false alarm rate. 10-Fold cross validation has been used for training and testing. The results are summarized in Table No. 2 and 3.

***Table: 2** Comparison of four Bayes net classifiers using Entropy based feature selection methods*

Entropy based Feature Selection Method	Classifier Techniques	Evaluation Criteria				
		Accuracy in %	Precision in %	Recall / Detection Rate in %	F-Value in %	False Alarm Rate in %
Information Gain	BayesNet + K2	96.5866	96.5074	96.1453	96.3260	3.0293
	Bayesnet + Tabu Search	96.9382	96.686	96.7372	96.7116	2.8867
	Bayesnet + Hill Climbing	96.5866	96.5074	96.1453	96.326	3.029
	Bayesnet + TAN	**99.7198**	**99.6152**	**99.7854**	**99.6992**	**0.3356**
Gain Ratio	BayesNet + K2	95.5729	94.7199	95.8298	95.2716	4.6508
	Bayesnet + Tabu Search	96.6929	94.2768	98.8982	96.5322	5.227
	Bayesnet + Hill Climbing	95.5729	94.7199	95.8298	95.2716	4.6508
	Bayesnet + TAN	**99.2562**	**99.6045**	**98.7941**	**99.1977**	**0.3415**
Symmetrical Uncertainty	BayesNet + K2	96.7176	96.5682	96.3722	96.4701	2.9817
	Bayesnet + Tabu Search	96.7271	96.6518	96.3039	96.4775	2.9045
	Bayesnet + Hill Climbing	96.7176	96.5682	96.3722	96.4701	2.9818
	Bayesnet + TAN	**99.7087**	**99.6252**	**99.7493**	**99.6872**	**0.3267**

***Table: 3** Comparison of four Bayes net classifiers using Statistical based feature selection methods*

Statistical based Feature Selection Method	Classifier Techniques	Evaluation Criteria				
		Accuracy in %	Precision in %	Recall / Detection Rate in %	F-Value in %	False Alarm Rate in %
Chi Squared Attribute Evaluator	BayesNet + K2	97.2272	97.415	96.6058	97.0087	2.2319
	Bayesnet + Tabu Search	96.3714	96.944	96.2934	96.6176	3.5376
	Bayesnet + Hill Climbing	97.2272	97.415	96.6058	97.0087	2.2319
	Bayesnet + TAN	**99.6698**	**99.5135**	**99.7783**	**99.6457**	**0.4247**
Relief-F	BayesNet + K2	93.9701	92.6663	94.525	93.5864	6.5129
	Bayesnet + Tabu Search	91.6347	90.2242	91.9939	91.1004	8.678
	Bayesnet + Hill Climbing	93.9701	92.6663	94.525	93.5864	6.5129
	Bayesnet + TAN	**98.1972**	**98.4892**	**97.6241**	**98.0547**	**1.3038**
One-R	BayesNet + K2	96.1492	96.4421	95.2396	95.8371	3.059
	Bayesnet + Tabu Search	97.9535	97.6827	97.926	97.8042	2.0224
	Bayesnet + Hill Climbing	96.1492	96.4421	95.2396	95.8371	3.059
	Bayesnet + TAN	**99.7412**	**99.6796**	**99.7646**	**99.7221**	**0.2792**

It is clearly observed that TAN classification technique gives the highest accuracy, detection rate and low false alarm rate irrespective of the feature selection methods. TAN with one-R gives the highest accuracy of 99.7412%, highest detection rate of 99.7646%, and low false alarm rate of 0.2792%. These results suggest that TAN classification technique outperforms other three techniques, thus qualify to be a potential candidate for building and effective IDS.

7. CONCLUSIONS

In this paper, we have proposed a hybrid intrusion detection model based on four Bayes net based classifiers and three different categories of feature selection methods. The performance of the model was analyzed along different evaluation criteria on the intrusion dataset. It was observed that the TAN classifier with One-R feature selection gives the highest accuracy, detection rate and low false alarm rate.

REFERENCES

[1]. J.T. Oh , S. K. Park, J.S. Jang, and Y.H. Jeon "Detection of DDoS and IDS Evasion Attacks in a High-Speed Networks Environment" , International Journal of Computer Science and Network Security (IJCSNS), VOL.7 No.6, pp.124-131, 2007.

[2]. J. Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco, CA, USA, 1988.

[3]. T.A.Mazzuchi, L.Koc, and S.Sarkani. "A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier", Expert System with Applications vol 39, no. 18, pp.13492-13500, 2012.

[4]. S. Mukherjee. and N.Sharma "Intrusion Detection using Naïve Bayes Classifier with Feature Reduction" Procedia Technology 4, pp.119 – 128, 2012

[5]. Upendra and Y.K.Jain, " An Efficient Intrusion Detection Based on Decision Tree classifier using feature Reduction" , International Journal of Scientific and Research Publications, vol.2, no.1, pp. 1- 6, 2012.

[6]. A. Elngar, A. Mohamed, and. F.M.Ghaleb, "A Real Time Anomaly Network Intrusion Detection System with High Accuracy",Information sciences Letters, An International Journal, vol. 2, no. 2, pp.49-56, 2013.

[7]. G. F. Cooper and E. Herskovits, A Bayesian method for the induction of probabilistic networks from data, Mach. Learn, vol. 9, pp. 309-347, 1992.

[8]. F.Glover, "Tabu Search - part I", ORSA Journal on Computing, vol.1, no. 3, pp.190–206, 1989

[9]. W.L. Buntine, " A guide to the literature on learning probabilistic networks from data", IEEE Transactions on Knowledge and Data Engineering, vol.8:, pp.195-210, 1996.

[10]. N. Friedman, D. Geiger, and M. Goldszmidt. "Bayesian Network Classifiers", Machine Learning, vol.29, pp.131-163, 1997.

[11]. M. Tavallaee, E. Bagheri, Wei Lu, and A. Ghorbani, " A detailed analysis of the KDD CUP 99 data set" , Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense applications (CISDA 2009), pp. 1-6, 2009.