© 2014,
Scienceline Publication
www.science-line.com

ISSN: 2322-5114
Journal of World's Electrical Engineering and Technology
J. World. Elect. Eng. Tech. 3(1): 01-05, 2014

JWEET

Presenting a Model for Making a Comparison of Bayesian Networks and Decision Tree Algorithms in Intrusion Detection Systems-Based on Data Mining

Hasan Fazli-Maghsoudi[1*], Hossein Momeni[2]

[1]*Master student in Information Technology, Department of Information Technology, University of Science and Technology of Mazandaran, Babol, Iran*

[2]*Assistant Professor; Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran*

*Corresponding author's Email: Fazli.m099@gmail.com

Abstract *– By development of information technology, network security is considered as one of the main issues and has great challenges. Intrusion detection systems are a major component of a secure network. Traditional intrusion detection systems cannot adapt themselves to the new attacks thus today's intrusion detection systems have been introduced based on data mining. Identifying patterns in large volumes of data is a great help to us. Data mining techniques by identifying a binary label (normal packet, abnormal packet) and specifying attributes by classification algorithms can recognize the abnormal data. Therefore, the precision and accuracy of intrusion detection systems will increase, thereby network security increases. In this paper, we present a proposed model that examines various decision tree algorithms and Bayesian networks on their data sets in which the results of simulation suggest that J48 algorithm has the highest precision of 85.49% for the intrusion detection system.*

Keywords*: abnormal packet, Bayesian networks, data mining, decision tree, intrusion detection systems, normal packet.*

Received 17 Jan. 2014
Accepted 30 Jan. 2014

ORIGINAL ARTICLE

INTRODUCTION

By rapid growth of computer networks and growth of the internet, network attacks, especially on the Internet has increased. Intrusion detection systems have been developed to ensure secure storage and processing of data on the network. Denial of service attacks on the Web is one of the most important attacks. A secure network must have features like data integrity, data availability, and data accuracy. The availability of data is the same as addressing and prevention of service denial attacks. Considering renovation of the attacks we must use learning systems in intrusion detection system which has capability of mining pattern of previous attacks, and can detect new attacks [1]. In this paper we provide a model based on data mining. Initially dataset preprocessing is conducted and then using decision tree and Bayesian network algorithms with respect to the accuracy parameters we propose the best algorithm. In this paper we initially present introduction, in the Section II the fundamental principles and concepts, related works are presented in the Section III.

In Section IV, the proposed model, algorithms and data preprocessing are explained, Section V includes the simulation results, and finally Section VI concludes the paper.

PRINCIPLES AND FUNDAMENTAL CONCEPTS

Intrusion detection is the process of monitoring the events occurring on a network or computer system in order to detect the deviation of security policies. This system is an application with the ability to identify, detect and respond to unauthorized or abnormal activities associated with the system. Intrusion detection process is shown in Fig. 1.

Two general approaches exist to implement intrusion detection test; misuse detection, and anomaly detection [1]. The performance of misuse detection is to identify the attack and define a model for analyzing engine and search for series of events which correspond with predetermined pattern. The performance of anomaly detection is to identify the normal operation of the system and provision of a profile of normal behavior for analysis engine and search for operating abnormalities [1]. Data mining is the process of finding knowledge from vast amounts of data stored in databases, data warehouses or other information repositories [2].

To cite this paper: Fazli-Maghsoudi H, Momeni H,. 2014. Presenting a Model for Making a Comparison of Bayesian Networks and Decision Tree Algorithms in Intrusion Detection Systems-Based on Data Mining. *J World Elec. Eng. Tech.*, 3 (1): 01-05.
Journal homepages: http://www.jweet.science-line.com/

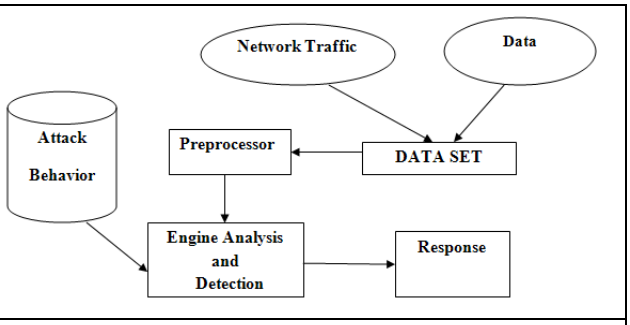

Fig. 1. Block diagram of Intrusion detection

A. Decision tree

Decision trees are a method to display a series of laws that would lead to a category or value. One of the differences among the methods of building decision tree is that how the distance is measured. Decision trees that are used to predict a set of variables, called classification trees because the samples are placed in categories or classes. Decision trees that are used to predict continuous variables are called regression trees [2].

A.1. Decision tree learning algorithms

Most algorithms for learning decision tree operate based on a top-down search in the space of existing trees. In the decision tree (ID3) a statistical quantity which is called Information Gain is used to define how much a feature is able to isolate training examples in terms of classification. Entropy specifies purity degree (irregularities or lack of purity) of a set of [2]. If the set S consists of positive and negative examples of a target concept, S entropy with respect to Boolean classification is defined as follows.

$$Entropy\,(S) = -p^{+}\log_2(p^{+}) - p^{-}\log_2(p^{-}) \quad (1)$$

Information gain of a feature is the entropy loss due to separation of examples and can be obtained through this feature [2]. In other words, information gain (S, A), for a feature such as A with respect to example sets of S is defined as follows:

$$Information\ Gain = Entropy\,(S) - \sum_{v \in Values\,(A)} \frac{S_v}{|S|} Entropy\,(S) \quad (2)$$

where Values(A) is a collection of all the features of A and SV is a subset of S for which A has value V. In the above mentioned definition the first term is the amount of data entropy, and the second term is expected amount of entropy after separation of given data [2].

A.2. Regression trees

Learning duty in regression trees includes forecasts of real numbers instead of a discrete set of values. It shows this by having real values in the leaf nodes. In which average values of training samples in the target leaf node is obtained. These types of trees are easy to interpret and can approximate a piecewise constant function altogether. More sophisticated version of regression trees are model trees which shows a regression function by having internal or final nodes. (Each node has linear regressions functions). After regression tree was fully constructed, linear regression practice is applied to the samples of this node and only a subset of the attributes (attributes that will be seen In the sub tree) are used for this task. Because of using a subset of the attributes of each node, overhead of linear regressions will not be high[2].

A.3. Method of Bayesian network classification

At first we explain the simple Bayes and then Bayesian network is explained. Suppose A1 to An are attributes with discrete values, these values are used to predict a class of discrete C [2]. Our goal is to predict and select the category that equation (1) is the maximum.

$$P\left(C = c \mid A_1 \cup A_2 \cup ... \cup A_n\right) \quad (3)$$

Using the Bayesian rule, we have:

$$P\left(C = c \mid A_1 \cup A_2 \cup ... \cup A_n\right) = \frac{P\left(A_1 = a_1 \cup ... \cup A_n = a_n \mid C = c\right) P(C = c)}{P\left(A_1 = a_1 \cup ... \cup A_n = a_n\right)} \quad (4)$$

The denominator is ineffective for making decisions. Because C is the same for all values.On the other hand due to our independent feature set

$$P\left(A_1 = a_1 \cup ... \cup A_n = a_n \mid C = c\right) = P\left(A_1 = a_1 \mid C = c\right) \times ... \times P\left(A_n = a_n \mid C = c\right) \quad (5)$$

In general, if we consider C as indicator for classification issues, the goal is to maximize the value of $P(X \mid C\_i) \times P(C\_i)$ which X is other characters. The advantages of simple Bayes are easy implementation and good results for many applications. The disadvantages one can say that perhaps all the features are not independent of each other and there is dependence, in this case the model is poor[2].

Bayesian networks describe the conditional dependencies between variables (attributes).By using this network, prior knowledge in the field of dependencies between variables with training data classification model are combined [2].

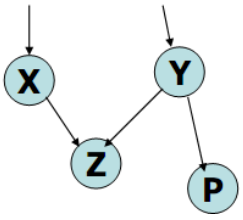

Fig. 2. Bayesian networks [2]

On Bayesian networks, nodes are variables that each of them has a specific set of conditions which are incompatible pairwise. Arcs (edges) indicate variable dependencies to each other. There are local distribution probabilities for each node which depends on the node

and is independent from parents' status [2]. An important assumption of a simple Bayes is conditional independence of classes from each other. But in practice, there are dependencies between variables. Possible Bayesian networks explore these possibilities.

A Bayesian network consists of two parts: the non-cyclic graph and conditional probabilities[2]. If an arc is connected to the *Y* from *Z*, it means that *Y* is the father *Z*. Each arc shows cause and effect relationship between associated variables. For each variable *A* with parents B_1, B_2, ...,B_n, there is a conditional probability table. In this table, for each variable its relationship considered with its parents [2]. Suppose given x with attributes x_1, x_2, ...,x_n, in this case:

$$p\left(x_1, x_2, ..., x_n\right) = \prod_{i=1}^{n} p\left(x_i \mid parents\left(y_i\right)\right) \qquad (6)$$

To learn these networks several scenarios exist. One is giving experts to fill the conditional probability table and drawing the related graph. Another method is to use heuristic methods such as hill climbing [2].

RELATED WORKS

In this case, a lot of work has been done. Of course, data sets and algorithms used in this paper are somewhat different. In decision tree most papers have tried to use ID3 and C4.5 algorithms. In [3] a comparison between decision tree and support vector machine (SVM) for analyzing network protocol was performed and it became clear decision tree has a better answer compared with support vector machine. In [4] using a decision tree was studied which by combining pattern matching and protocol analysis methods was performed. The first method is most commonly used to identify patterns and the second method is to build a decision tree based on analysis of network traffic. In [5] classification of variety of attacks is done with decision tree. In this paper, four kinds of main attack classes and 22 minor attack classes were considered and varieties of attacks are investigated.

In [6] Bayesian model has been used for intrusion detection in a way that simple Bayesian model a little has been modified and the answer is more accurate. In [7] the method of game theory with Bayesian model for intrusion detection in wireless networks has been used. In this issue it is tried to build a model that can prevent energy loss in case of an attack. The two game modes are used in static and dynamic form in which the dynamic mode is closer to reality. In [8], the Bayesian network approach for intrusion detection in wireless networks has been used. The purpose of using signature detection is to detect abnormal packets. When this signature does not conform to attack packages, package is discovered. The main problem with this method is updating attacks. In [9] Bayesian neural network approach has been used to classify network traffic. Technique is a method of supervised classification by using data and features which are derived from the contents of the pack and the strength of method is based on this characteristics. In [10] an intrusion detection system based on a probabilistic model that uses Bayesian networks is suggested.This method is more focused on IP Spoofing attacks and characteristics and features of the package contents are used in. In this method agent-based architecture is used so that agents communicate and collaborate with each other and they have the ability to update.

PROPOSED ALGORITHM

In this model, we first consider the data set. These datasets are related to NSL-KDD which contain 42 features[11]. That feature Num_outbound_cmd is always zero in which this amount will be eliminated. The rest of the features, depending on the algorithm used for data pre-processing operations conducted and the data appropriately modified to work with the decision trees algorithms and Bayesian networks. The modeling on Rapid miner software adding to this setting WEKA algorithms[12], and preprocessing data has been done. From the dataset, 17000 data are used in training step and 3000 samples are used in test phase. The main task which has been done is to provide a number of data sets extracted from the original sets and also modeling of a laze model that so far has not been used for intrusion detection systems. This model, like any other model which is based on data mining is as Fig. 3.

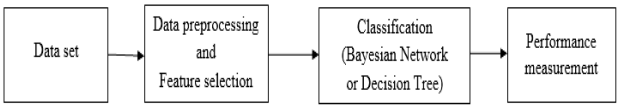

Fig. 3 - Proposed architecture by using decision tree and Bayesian network for the evaluation of intrusion detection.

A. Data preprocessing and features selection

Data set is considered and according to the nature of the algorithms is divided into two categories: we do preprocessing and discretization. Two data sets are required for processing.

A.1 Decision tree

A discrete set of data to work with ID3 and CHAID tree and the ADT and decision tree, all data are nominal. According to our data, the attributes are classified into the following

TABLE I
CONVERTED INDICATORS

Attribute name	Attribute type
PROTOCAL_TYPE	NOMIAL
FLAG	NOMIAL
SERVICE	NOMIAL
LAND	BINOMIAL
WRONG_FRAGMENT	NOMIAL
NUM_FAILED_LOGINS	NOMIAL
ROOT_SHELL	BINOMIAL
SU_ATTEMPTED	NOMIAL
NUM_SHELLS	NOMIAL
NUM_ACCESS_FILE	NOMIAL
IS_HOST_LOGIN	BINOMIAL
IS_GUEST_LOGIN	BINOMIAL

A.2 Bayesian network

Due to the nature of algorithms the data sets are divided into three categories as follows and we do preprocessing and discretization as

- INTEGER and REAL data set according to Table 1.
- NUMERICAL datasets except Index
- Discrete datasets

SIMULATION RESULTS

In Rapidminer software we use precision parameter (accuracy).

A. Bayesian network algorithms

Kernel naive Bayesian is an algorithm makes use of probability theory of simple Bayesian with kernel density. Weight function used in kernel uses estimation techniques without parameter. This model uses the data set type 1, and its accuracy is equal to 79.50%. Naive Bayesian algorithm uses simple probability theory of Bayesian and variables should be independent which uses data set type 2 and its accuracy is equal to 77.24%.Waode is a model based on the average estimate of weigh make dependency among the parameters which uses data set type 3 [14] and its accuracy is equal to 81.88%. Aode achieve average classification accuracy by replacing a simple Bayesian approach which does not have simple Bayesian terms. Like independence of variables which uses data set type 3 [15] and its accuracy is equal to 82.09%.Aodesr is the same as Aode, with these features at classification time that specification between the two features are detected and removed at the time of Generalization. The data set type 3 is used. And its accuracy is equal to 80.70%. And its accuracy is equal to 81.99%. Bayesenet model uses different kinds of algorithms to enhance the precision accuracy. This algorithm does not use ADT tree data structure and uses type three of data sets and its accuracy is equal to 81.73%.In HNB, a hidden parent is created for each attribute by combining its effects on other traits. This algorithm is known as HNB that uses the weight of the inter-dependencies [16] and its accuracy is equal to 83.29%.Dmnbtext is a model for making a simple Bayesian using polynomial division (discriminative multinomial). Data set type 1 is used and its accuracy is equal to 75.35% [17]. Bayesian Logic Regression model implements Bayesian function by Laplace and Gauss estimation, and uses type 3 datasets, and its accuracy is equal to 76.65%.

B. Decision tree algorithms

Chaid tree is like a decision tree with the difference that instead of information gain the chi-square is used. This algorithm does not accept numerical data and data should be nominal. Its accuracy is equivalent to 77.33%. In our model, we used only nominal data. Because the obtained answer for decision tree (simple) is better than other data and its accuracy is equal to 80.46%. ID3 tree make spruned ID3 decision tree which only accepts nominal data and its accuracy is equivalent to 79.45%. ADT tree makes a decision tree based on Freund *et al* algorithm [18]. In the algorithm only nominal data is used. This tree uses heuristic methods. By default, each node has 3 children with accuracy equal to 76.54%. BF tree is a decision tree based on the first level of the search syntax [20], and its accuracy is equal to 78.81%. FT decision tree which uses a regression function in the leaves and nodes [21] with accuracy equal to 78.81%. J48 tree pruned or not pruned tree which derived from the tree C4.5 [19], and its accuracy is equal to 85.49%. LAD tree makes a decision tree based on Logic Boost in strategy[22]. And its accuracy is equal to 75.84%. LMT tree is decision tree where the leaves have regression functions [23]. And its accuracy is equal to 81.32%. M5P tree builds a tree based on the M5 algorithm [24]. And its accuracy is equal to 74.34%. NB tree is a decision tree where in its leaves Bayes algorithm is used [25] where its accuracy is equal to 82.10%.REEP tree is a decision tree using Information Gain/Variance and pruning is used and its accuracy is equal to 81.00%. Simplecart tree makes a decision tree with minimal pruning cost [26]. And its accuracy is equal to 80.70%. In Fig. 4, obtained results for different algorithms are shown.

CONCLUSION

Due to the increasing development of Internet and Internet attacks, the existence of intelligent intrusion detection system becomes necessary. Many systems were proposed to perform the intrusion detection approach, such as decision tree and Bayesian network-based algorithms. In this paper, we provide the comprehensive study on the performance of the above mentioned algorithms, where several algorithms based on the data mining on training data and the test was evaluated. Test results indicate that the decision tree-based J48 algorithm provide the highest accuracy equal to 85.49% on the considered dataset.

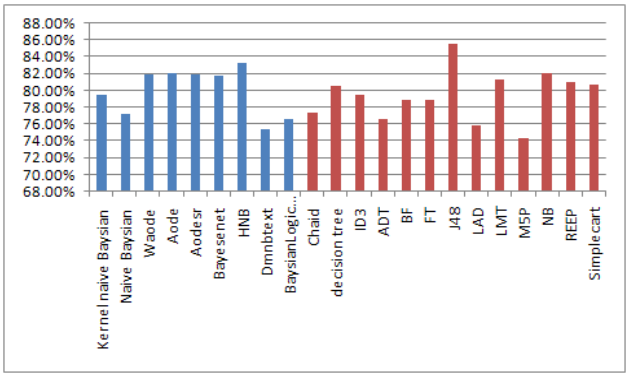

Fig. 4 - The results of comparison between decision tree algorithms and Bayesian networks (red bars are decision tree algorithms and blue bars are Bayesian network)

REFERENCES

[1] M. Bishop, "*Introduction to Computer Security*," Prentice Hall, 2004.

[2] J. Han and M. Kamber, "*Data Mining: Concepts and Techniques*", San Diego, Academic Press, 2001.

[3] A. SnehalandP.R.Devale,"Intrusion Detection System using Support Vector Machine and Decision tree," *International journal computer application*, vol. 3, no. 3, pp. 40-48, 2010.

[4] T. Abbas and A. Bouhoula,"Protocol Analysis in Intrusion Detection System Using Decision tree," in *Proc. International Conference in Information technology*, pp. 404-408,2004.

[5] D.M. Farid, N. Harbi, E. Bahri, M.Z. Rahman, and C.M. Rahman, "Attacks classification in adaptive intrusion detection using decision tree," *World Academy of Science, Engineering and Technology*, vol. 63, pp. 86-90, 2010.

[6] C. Kruegel, D. Mutz and W. Robertson, F. Valeur, "Bayesian event classification for intrusion detection," in *Proc. IEEE Computer Security Applications Conference*, pp. 14-23, 2003.

[7] Y. Liu, C. Comaniciu, and H. Man, "Game theory for communications and networks," in *Proc. workshop ACM New York*, pp .34-43, 2006.

[8] F. Jemili, M. Zaghdoud, and A. Ben, "A Framework for an Adaptive Intrusion Detection System using Bayesian Network," in *Proc. of IEEE intelligence and Security Informatics conference*, pp. 66–70, 2007.

[9] T. Auld, A.W. Moore, and S.F. Gull, "Bayesian Neural Networks for Internet Traffic Classification," *IEEE Trans. on* Neural Networks, vol. 18, pp. 223-239, 2007.

[10] V. Gowadia, and C. Farkas, "PAID: A Probabilistic Agent-Based Intrusion Detection system," *Computers & Security*, vol. 24, no. 7, pp. 529–545, 2005.

[11] *http://nsl.cs.unb.ca/NSL-KDD/*

[12] *http://www.rapid-i.com/content/view/181/*

[13] L. Jiang, H. Zhang, "Weightily Averaged One-Dependence Estimators," in *Proc. of biennial pacific rim international conference on artificial intelligence*, pp. 970-974, 2006.

[14] G. Webb, J. Boughton, and Z. Wang, "Not So Naive Bayes: Aggregating One-Dependence Estimators," *Machine learning*, vol. 58, no. 1, pp. 5-24, 2005.

[15] F. Zheng and I. Geoffrey, "Webb: Efficient lazy elimination for averaged-one dependence estimators," in *Proc. of the international conference on Machine Learning*, pp. 1113-1120, 2006.

[16] H. Zhang, L. Jiang, and J. Su, "Hidden Naive Bayes," in *Proc. of conference on artificial intelligence*, pp. 919-924, 2005.

[17] J. Su, H. Zhang, C.X. Ling, S. Matwin, "Discriminative parameter learning for Bayesian networks," in *Proc. of international conference on Machine learning*, pp. 1016-1023, 2008

[18] Freund, Y., Mason, L.: ''The alternating decision tree learning algorithm''. In: Proceeding of the Sixteenth International Conference on Machine Learning, Bled, Slovenia, 124-133, 1999.

[19] J.R. Quinlan, "C4.5: Programs for Machine Learning,'' *Morgan Kaufmann Publishers*, 1993.

[20] S. Haijian,"*Best-first Decision Tree Learning*" Thesis Master of Science (MSc) The University of Waikato, 1993.

[21] J. Gama, "Functional Trees," *journal of Machine Learning*, vol. 55, no. 3 , pp 219-250, 2004.

[22] G. Holmes, B. Pfahringer, R. Kirkby, E. Frank, and M. Hall," Multiclass alternating decision trees," in ECML, pp. 161-172, 2001.

[23] M. Sumner, E. Frank, and M. Hall, "Speeding up logistic model tree induction," in *Proc. of European conference on principles and practice of knowledge discovery in databases*, pp. 675-683, 2005

[24] Y. Wang and I.H. Witten, "Inducing model trees for continuous classes,"in *Proc. of the European conference on machine learning*, pp. 128-137, 1997.

[25] R. Kohavi, "Scaling up the accuracy of naive-Bayes classifiers: a decision-tree hybrid," in *Proc. of second international conference on knowledge discovery and data mining*, pp. 202-207, 1996.

[26] L.B.J.F.R. Olshen and C.J. Stone, "Classification and regression trees," *Wadsworth International Group*, 1984