ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

An Effective Intrusion Detection on Cloud Virtual Machines Using Hybrid Feature Selection and Multiclass Classifier

[1]S. Muthurajkumar, [2]S. Ganapathy, [1]M. Vijayalakshmi, [1]A. Kannan

[1]*Department of Information Science and Technology, College of Engineering Guindy, Anna University, Chennai, India*
[2]*Department of Computer Applications, Maulana Azad National Institude of Technology, Bhopal, MP, India.*

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form 26 December 2014
Accepted 29 January 2015
Available online 10 February 2015

Keywords:
Cloud computing, intrusion detection, genetic feature selection algorithm.

ABSTRACT

Recently, cloud computing has increasing rapidly gained success over the internet. Therefore, Cloud security issues are providing more challenges to the users and developers. In the past, various security services have been developed to improve the cloud security. However, they did provide the significant security for the cloud data. In this paper, we introduce a new hybrid intrusion detection system by combining a new hybrid feature selection and a new hybrid multiclass classification algorithm for providing enhanced security to cloud data. This Hybrid Multiclass Classification algorithm is proposed by combining the IAEMSVM and HNB. In addition, we propose a new hybrid feature selection method is called Hybrid Genetic based Feature Selection Algorithm (HGFSA) for effective classification. Experiments have been carried out on NSL-KDD'99 datasets for testing the proposed Genetic based attributes selection algorithm and Hybrid Multiclass Classifier.

To Cite This Article: S. Muthurajkumar, S. Ganapathy, M. Vijayalakshmi, A. Kannan., An Effective Intrusion Detection on Cloud Virtual Machines Using Hybrid Feature Selection and Multiclass Classifier. ***Aust. J. Basic & Appl. Sci.,*** *9(6): 38-41, 2015*

INTRODUCTION

Computer networking is becoming a major infrastructure of our daily communications. The network is interconnects the various people together to make a borderless world. Specially, cloud computing usage is increased in our regular life and the need for the cloud services to be secure and resilient to cyber-attacks and malicious attacks. Urgent need of the world is pleasant security software to safeguard the emerging data. Current software is enables attacks to spread rapidly and thus exposing the systems to major attacks by well-informed attackers. Even though, the threats and attackers are also increasing the level of computer networking technology growth. This is right time to possess the efficient intrusion detection system for securing the cloud and internetworking data.

Intrusion detection system (IDS) can be defined as a software system which can be used to detect malicious users in computer networks. IDS have been classified to two methods namely anomaly detection and misuse detection. Anomaly detection is the counterpart of misuse detection. The lack of misuse detection is it cannot detect newly attack that on no account learn or different from the stored signatures. Therefore, anomaly detection approach is the complement to misuse detection by the property that it can detect new attack without using signatures. Anomaly detection approach is to establish the normal behavior profile. The performance of misuse detection system is always higher than anomaly detection systems according to the nature of the model that learns from known knowledge. Misuse detection is a rule-based or signature-based depends on the algorithm used the known rules on the attack are kept in the systems. IDS distinguished into two type's namely host-based intrusion detection systems (HIDS) and network-based intrusion detection systems (NIDS). Host based Intrusion Detection System monitors the user behavior and state of a computer system where as Network based Intrusion Detection System monitoring all the packets which are transferred through the system to find wary patterns in the network traffic.

In this paper, a new intrusion detection system is proposed which is developed by combining an effective hybrid feature selection algorithm and hybrid Multiclass classifier. Hybrid Multiclass Classifier is the combination of a Hidden Naïve Bayes Classifier and Intelligent Agent based Enhanced Multiclass Support Vector Machine (IAEMSVM) for multiclass classification responsibilities. The proposed hybrid multiclass

Corresponding Author: S. Muthurajkumar, Department of Information Science and Technology, College of Engineering Guindy, Anna University, Chennai, India

classifier finds the difficult instances in the training data using a HNB classifier and removes these instances from the training set before apply the IAEMSVM algorithm and making the decision. In addition, we apply a new hybrid feature selection method for effective pre-processing and improve the classification accuracy.

Literature survey:

Tjhai *et al*. (2010) proposed two levels of classification method using Self-Organizing Map (SOM) based neural networks and k-means clustering algorithm. They achieved high reduction in false Positive Rate (FPR). Mansour *et al* (2010) proposed a data mining technique is called Growing Hierarchical Self organized Map (GHSOM) for effective intrusion detection in cloud. The proposed system reduces the false positive rate and false negatives significantly for the real world data Filtering algorithms. Amjad *et al* (2013) proposed an anomaly Intrusion Detection System using machine learning approach for virtual machines on cloud computing. Their proposal is feature selection over events from Virtual Machine Monitor to detect anomaly in parallel to training the system so it will learn new threats and update the model.

Zubair A. Baig *et al* (2013) proposed a new technique based on GMDH for classifying network data into normal and abnormal. Their proposed system have two types of techniques namely Monolithic and Ensemble-based, were tested on the KDD-99 dataset. The data set was pre-processed by using Information Gain, GMDH and Gain Ratio feature ranking methods. Their obtained results proved that the proposed intrusion detection system provides better detection rates when it is compared with existing classification techniques. Panos Louvieris *et al* (2013) proposed a novel intrusion detection technique used to detect the attacks on a network by identifying the important features. This feature selection method uniquely combines Naïve Bayes feature selection, k-means clustering and C4.5 classification for identifying attacks.

Reda M. Elbasiony *et al* (2013) proposed a data-mining-based network intrusion detection system. Two data-mining techniques are used by the authors in misuse, anomaly and hybrid detection. The random forests algorithm is used as a data mining classification algorithm into a misuse detection method to build intrusion patterns from a balanced training dataset, and to classify the captured network connections to the main types of intrusions due to the built patterns. Gisung Kim *et al* (2014) proposed a new hybrid intrusion detection model which is used to integrate the misuse and anomaly detection models. Koc *et al* (2012) applied a hidden naive Bayes (HNB) classifier for identifying the network attacks. Their system provides significant improvement for the detection of denial-of-services (DoS) attacks. Seyed Reza Hasani *et al* (2014) introduced a new wrapper optimization method of LGP_BA based on L-Genetic Programming (LGP) and Bees Algorithm to achieve an efficient feature selection algorithm.

System architecture:

The architecture of the proposed system for effective intrusion detection is shown in *Fig. 1*, It consists of four major components namely NSL-KDD 1999 Cup Data Set, Data collection Module, Hybrid Feature Selection Module, Hybrid classification module and result. All these components are responsible for performing intrusion detection effectively.

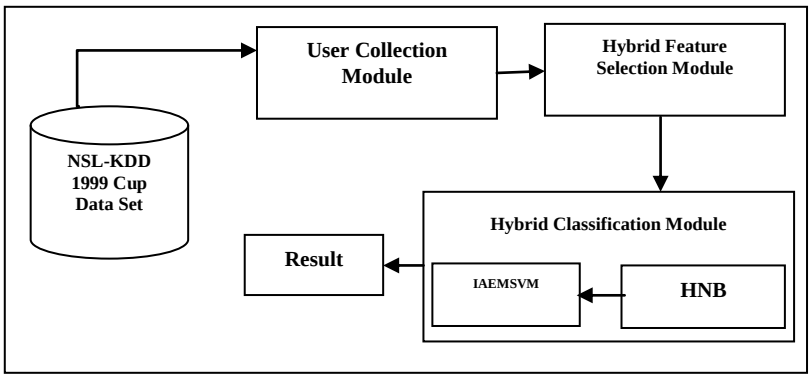

Fig. 1: System architecture.

NSL-KDD 1999 Cup Data Set contains 10% of data set from full data set. The data collection module collects the necessary data from the NSL-KDD data set. Hybrid Feature selection module contains two feature selection algorithms which is used to select the important features from NSL-KDD Cup'99 data set which has 41 features. Classification module is contains a hybrid multi class classification algorithms for classifying the data. The result module is contains the result of the system that are either normal or attacks.

Proposed method:

This section introduces a new genetic based hybrid feature selection algorithm and Hybrid Multiclass Classifier for effective intrusion detection.

Firstly, a hybrid feature selection algorithm based on Genetic and Discrete Particle Swarm Optimization (DPSO) (Alper Unler, Alper Murat, 2010) to achieve an efficient feature selection. Second, a new Hybrid Multiclass Classification algorithm is also proposed for effective classification which is combining the IAEMSVM (Snehal A. Mulay, 2010; Ganapathy, S., 2012) and HNB.

Genetic based hybrid feature selection algorithm:

The feature selection algorithm proposed in this paper is the based on combination of two optimization methods namely Genetic Algorithm (GA) and Discrete Particle Swarm Optimization (DPSO) (Alper Unler, Alper Murat, 2010) algorithm, which called HGFSA. This proposed method uses the genetic based random search technique for feature deduction. In this work, Genetic algorithm provides random selection of features among all features in dataset to perform new generations.

Algorithm:

Step 1: Read n records from the data set
Step 2: Initialize the population using these first n records.
Step 3: Randomly select m features from n (Select the features)
Step 4: Calculate the fitness value using the equation
Fitness Value = $m - n / m + n$
Step 5: If the Fitness Value > 0.5 then
Step 6: Call the Discrete PSO algorithm (Alper Unler, Alper Murat, 2010)
Step 7: Apply cross over operator to the chromosome
Step 8: Apply mutation operation to the chromosome
Step 9: Compute the fitness again, if it is less than < 5 and previous current < 0.01 then go to step 4.
Step 10: End if.

First, the population is selected by m selection from n possibilities and the first stage fitness will be calculated. The main criterion in the GA part is the fitness value must be more than 50. If the condition has been met by the fitness function the chromosome will be passed to the BA process otherwise GA applies the crossover and mutation to create some generation until the proper fitness values achieved. Parents in the original algorithm are composed of the n number of individuals selected by a selecting algorithm, while parents in queen-bee evolution consist of the n/2 number of couples of a queen-bee. All parts of the individuals in genetic algorithm are mutated into small mutation probability while only some parts of individuals in GHFSA are mutated into normal probability and others with strong probability.

Hybrid Multiclass Classification:

1) Hidden Naïve Bayes classifiers:

Hidden Naïve Bayes (HNB) classifier (Dewan M. Farid, 2014) is an extension proposed on the Naïve Baye's classifier. This algorithm relaxes the conditional probability to use in the basic Baye's classifier. It creates another layer to represent the hidden parent of each attribute and combines all the other attributes with this hidden parent. In the HNB model, each attribute has a hidden parent which influences from all the attributes.

This HNB method uses hidden parent that influences all the other attributes through conditional probability. It includes the influence of difficult attributes dependencies in large datasets.

2) Intelligent Agent Based Enhanced Multiclass Support Vector Machine:

The IAEMSVM (Ganapathy, S., 2012) algorithm is used intelligent agent for achieving better detection accuracy than the existing multiclass classification algorithms. Minkowsi distance measurement formula is used in this algorithm for reducing the execution time. Moreover, it achieves better detection accuracy through binary decision tree using the intelligent agents.

Experimental results:

We implemented our method in java by using the NB Tree original implementation (Seyed Reza Hasani, 2014) and tested through the NSL-KDD of KDD'99 datasets (Reda M. Elbasiony, 2013). The experimental results indicate that the models based on pre-processing and classification, while the results of the hybrid model provide better performance than the existing hybrid approaches. The class distributions on the NSL-KDD 1999 Cup data set balanced dataset differ in the training and test data. From the various experimental results, we can observe that the proposed hybrid intrusion detection model is fast and perform well where the huge data load has cloud virtual machine. Table 1, shows the list of selected features from 41 features in the NSL-KDD'99 Cup data set using the Hybrid Genetic Feature Selection Algorithm (HGFSA).

Table 1: List of selected features.

Ipsweep, Nmap, Portsweep, Satan, Back, Land, Neptune, Pod, Smurf, Teardrop, Buffer_overflow, Perl, Rootkit, Ftp_write, Guess_passwd, Multihop, Spy, Warezclient

Table 2: Comparison of the overall performance of hybrid intrusion detection systems.

Exp.No.	Name of the Hybrid Model	Accuracy (%)
1	Proposed HFSMC	95.85
2	IAFSH (Ganapathy, S., 2012)	92.81
3	HNNA (Li Xiangmei, Qin Zhi, 2011)	85.67
4	Hybrid IDS (DT +NB) (Dewan M. Farid, 2014)	81.91

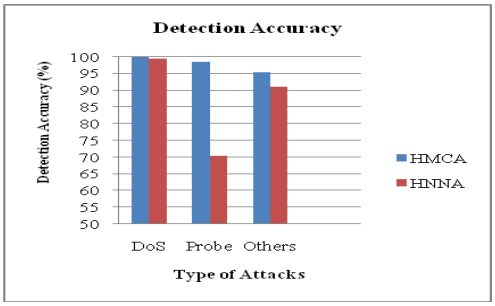

Fig. 2: Results comparison between HMCA and HNNA.

Table 2, shows that the performance of proposed hybrid approach is better performance in terms of accuracy than the existing hybrid systems.

Fig. 2, Shows the comparison of detection accuracies for DoS, Probe and other attacks for the Hybrid Neural Network Algorithm (HNNA) model (Li Xiangmei, Qin Zhi, 2011) based on the proposed model HMCA. From this figure, it can be seen that the proposed HMCA model achieves better detection accuracy than the existing HNNA (Li Xiangmei, Qin Zhi, 2011).

Conclusions:

In this paper, we proposed the hybrid approaches for feature selection and classification which can perform well in detecting intrusions in virtual machine environments on cloud. First the proposed hybrid feature selection algorithm is used to identify the valuable attributes from the NSL-KDD'99 Cup data set. Second, we proposed an effective hybrid classifier for handling the multiclass problem while classifying the data. These two hybrid approaches helps to perform the higher accuracy than the existing approaches. The proposed hybrid intrusion detection system has been evaluated over the NSL-KDD of KDD'99 datasets after solving the problems of categorical and different scales features.

REFERENCES

Alper Unler, Alper Murat, 2010. A Discrete Particle Swarm Optimization Method for Feature Selection in Binary Classification Problems. European Journal of Operational Research, 206: 528-539.

Amjad Hussain Bhat, Sabyasachi Patra, Debasish Jena, 2013. Machine Learning Approach for Intrusion Detection on Cloud Virtual Machines. International Journal of Application or Innovation in Engineering & Management (IJAIEM), 2(6): 57-66.

Dewan M. Farid, Li Zhang, Chowdhury Mofizur Rahman, M.A. Hossain, 2014. Rebecca Strachan. Hybrid decision tree and naive Bayes classifiers for multi-class classification tasks. Expert Systems with Applications, 41: 1937-1946.

Ganapathy, S., P. Yogesh, Kannan Arputharaj, 2012. Intelligent Agent Based Intrusion Detection System Using Enhanced Multiclass SVM. International Journal of Computational Intelligence and Neuroscience, 2012: 195-202.

Gisung Kim, Seungmin Lee, Sehun Kim, 2014. A Novel Hybrid Intrusion Detection Method Integrating Anomaly Detection with Misuse Detection. Expert Systems with Applications, 41: 1690-1700.

Koc, L., T.A. Mazzuchi, S. Sarkani, 2012. A Network Intrusion Detection System Based on a Hidden Naive Bayes Multiclass Classifier. Expert Systems with Applications, 39: 13492-13500.

Li Xiangmei, Qin Zhi, 2011. The Application of Hybrid Neural Network Algorithms in Intrusion Detection System. IEEE Conference on Trends in Neural Network.

Mansour, N., M. Chehab, A. Faour, 2010. Filtering intrusion detection alarms. Cluster Computing, 13: 19-29.

NSL-KDD University of New Brunswick, Canada http://nsl.cs.unb.ca/NSL-KDD/.

Panos Louvieris, Natalie Clewley, Xiaohui Liu, 2013. Effects-Based Feature Identification for Network Intrusion Detection. Neurocomputing, 121: 265-273.

Reda M. Elbasiony, Elsayed A. Sallam, Tarek E. Eltobely, Mahmoud M. Fahmy, 2013. A Hybrid Network Intrusion Detection Framework based on Random Forests and Weighted K-Means. Ain Shams Engineering Journal, 4: 753-762.

Seyed Reza Hasani, Zulaiha Ali Othman, Seyed Mostafa Mousavi Kahaki, 2014. Hybrid Feature Selection Algorithm for Intrusion Detection System. Journal of Computer Science, 10(6): 1015-1025.

Snehal A. Mulay, P.R. Devale, G.V. Garje, 2010. Intrusion Detection System using Support Vector Machine and Decision Tree. International Journal of Computer Applications, 3: 0975-8887.

Tjhai, G.C., S.M. Furnell, M. Papadaki, N.L. Clarke, 2010. A preliminary two-stage alarm correlation and filtering system using SOM neural network and K-means algorithm. Computers & Security, 29: 712-23.

Zubair A. Baig, Sadiq M. Sait, Abdul Rahman Shaheen, 2013. GMDH-Based Networks for Intelligent Intrusion Detection Engineering Applications of Artificial Intelligence, 26: 1731-1740.