VOL. 9, NO. 11, NOVEMBER 2014 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

EFFECTIVE INTRUSION DETECTION SYSTEM FOR CLOUD ARCHITECTURE

P. Padmakumari, K. Surendra, M. Sowmya and M. Sravya
CSE, School of Computing, SASTRA University, Thanjavur, Tamilnadu, India
E-Mail: padmalec.sastra@cse.sastra.edu

ABSTRACT

Cloud computing enables the end users to easily access internet based applications and data storage services. With the increasing popularity of cloud, providing security to cloud environment has become an important issue. In order to provide security for a cloud environment we require more than the traditional security methods like firewalls, user authentication, access control and confidentiality in data transmission. Hence Intrusion Detection System (IDS) becomes a needful component in terms of cloud security. Many methods are being utilized for the development of effective intrusion detection systems, but none of them is completely secure. In this process of betterment, here we present an intrusion detection system, by applying k-means clustering for anomaly detection and integrate it with a frequent attacks generation module using apriori algorithm to detect frequently occurring attacks in various network environments. To evaluate the performance, KDD 99 CUP dataset has been used in our system.

Keywords: cloud, security, intrusion detection **s**ystem, KDD 99, anomaly based detection.

INTRODUCTION

Security becomes a major issue for various organizations because of rapid growth in internet and network based services. Even though exchanging the information over the network had improved the efficiency, it also had given an opportunity for cyber-attacks. There are many ways in which an attacker can attack network of an organization. These can be accessing the information for which he is not authorized, bringing down the whole network, etc. Many methods are being introduced to improve the security of a network. Some of the systems make use of cryptographic methods, firewalls etc. But none of the methods is completely secure. Hence we cannot totally rely on attack prevention techniques. Hence we need some methods so that we can detect whether attack is being done or not so that we can take corrective action. The system which use these methods are said to be *Intrusion Detection Systems* (IDS). *Intruder* can be defined as a program, system or a person who tries to gain unauthorized access to a network or an information system and perform an illegal action [1]. In other words, intrusion can be defined as an action that aims to compromise the security, confidentiality and integrity of a network or a system [1]. Intrusion detection system (IDS) is a software application or a device that monitors the network traffic to detect any network intrusions, malicious activities that involves in compromising the security of a system or violation of system policies. IDS perform various activities like scanning network traffic, finding unauthorized access to resources if any and alerting admin or appropriate persons when an intrusion was found so that pre-cautionary measures can be taken [2].

There are two approaches for detecting intrusions in a system: signature based detection and anomaly based detection [3]. Some of the intrusion detection systems make use of both signature based and anomaly approaches which are called as *hybrid systems.* Signature based IDSs uses a rules or signature database to detect malicious attacks [4]. When the signature of incoming network packet matches with the pattern present in the database it is detected as an attack. The strength of these systems lies in signature database and hence the signature database has to be updated regularly with signatures of new intrusions. The limitation of these types of IDSs is that they cannot detect unknown attacks. Anomaly based IDS defines the normal behavior of the system. Any network packet that deviates the normal behavior of the system is detected as an attack. The advantage of anomaly based IDS over signature based IDS is that it can detect novel (unknown) attacks. The limitation of anomaly approach is that it has tendency of generating high false alarm rate [4].

Besides these limitations of traditional intrusion detection systems, there are some other problems which are faced by many of the intrusions systems. They are:

Fidelity problem: During the examination of attacks, IDS make use of the data related to network packets which is stored in log files. During the transmission of data from source (log files) to the place where IDS is positioned, data may be modified by the intruder. This may result in missing of some of the events. This refers to *fidelity* problem.

Resource usage problem: Most of the intrusion detection systems are designed in such a manner that all of its components work all the time even though when there is no sign of intrusions found in the network. This may result in heavy utilization of resources by the components of IDS which is of no use. This is known as *resource usage* problem.

Reliability problem: This is the biggest problem facing by most of the existing intrusion detection systems. IDS is used to detect intrusions but what happens when the intruder directly modify IDS in such a way that it can't detect legitimate intrusions anymore. This problem is

referred as *reliability* problem. The only solution is to reduce human intervention in the operation of intrusion detection system.

High false alarm rate: Almost all of the intrusion detection systems which are based on anomaly detection suffer from this problem. To overcome this problem anomaly based IDSs has to be combined with signature based IDSs to significantly reduce the false alarm rate. Or a robust anomaly based IDS using advanced techniques has to be designed to lower the false alarm rate.

RELATED WORK

In [5], H.M. Shiraji wrote a well-known paper that uses a combination of both memetic algorithm and Bayesian networks to detect intrusions in network environments. In [1] Mohammad Sazzadul Hoque *et al*. proposed an intrusion detection system using genetic algorithm. Stand deviation equation is used to evaluate the fitness of chromosome.

In [6], Patel Hemant *et al* discussed various data mining algorithms that are used for network intrusions. These algorithms include Naïve Bayesian, NBTree and Decision tree. Advantages and disadvantages of each algorithm are explained in brief. Special characteristics of individual attacks are used to build decision tree using Naïve Bayesian algorithm. On the other hand, NBTree is used to scan the network and analyse the complex features of attacks behaviours. The results of this analysis are used to detect attacks in network environments. In [7] Vaidehi Kasarekar *et al*. proposed an intrusion detection and a real time response system which makes use of both anomaly detection and signature based detection. This system was designed with purpose of increasing the security of various wireless networks. This system yields much better results when it is integrated with a highly robust anomaly detection engine.

Aritra Kundu in [8] proposed a hybrid IDS with the intention of lowering the false rate in anomaly detection. Important attributes are selected by using entropy based feature selection algorithm and irrelevant attributes are removed. Combination of two classifiers Naïve Bayesian and K-nearest neighbour are used to detect intrusions. Fuzzy algorithms are used to reduce misclassification of normal and anomalous data.

David J. Day *et al*. in [9] proposed CONDOR (COmbined Network intrusions Detection ORientiate), a hybrid intrusion detection system which combines the merits of both misused and anomaly detection approaches. This system was proposed with the aim to increase detection rate and to reduce the intervention of administrator in the operation of IDS.

Based on the above research, although anomaly based intrusion detection is not a new one, but performing anomaly detection using K-means clustering is still a fresh concept and provides the scope for researcher to work on it. This research uses K-means clustering technique for anomaly detection and an apriori algorithm to detect frequently occurring attacks. This research ensures to improve the system security and also alert the admin or appropriate persons to decide on future plan of actions when an attack is detected.

INTRUSION DETECTION SYSTEM OVERVIEW

The following subsections present the overview of various attacks in networking, various classifications and approaches of Intrusion Detection.

Attacks in networking

Denial of service (DoS): It refers to an attempt by an attacker making the host machine or the system resources unavailable to the legitimate users of a service. This attack can be implemented as consuming system resources making it unavailable for the legitimate/intended users or by blocking the communication between the legitimate users and the target machines, making any further communication inadequate.

Backdoor: It is the method of obtaining remote access to the system by bypassing the security policies of the system such as authentication, thereby gaining access to the text within the system while attempting to remain unknown. The implementation of backdoor can be done by compromising the compiler, which not only compiles the source code with inclusion of backdoor as normal code but also places the threat within the system.

Probing: It is an attack in which the attacker attempts to compromise the target machine by scanning the entire system in order to exploit the vulnerabilities.

Major classifications of intrusion detection system

Host-based intrusion detection system (HIDS): These run on and monitor packets to and from individual devices on the network. They evaluate information found on multiple host systems. It even determines the accessibility of system resources.

Network-based IDS (NIDS): It is an intrusion detection system which analyses the network packets or the network flow to determine the illegitimate access to a computer network. These are placed within the network to monitor and analyse the network traffic for matches in database of known attacks. They evaluate information from network communication. The example of such NIDS is *Snort.*

Approaches for intrusion detection

Misuse-based intrusion detection: In this type of detection, the network packets are checked against the database of known attacks for any pattern matching. The major limitation of this type of detection is that it only determines the known attacks leaving the unknown future intrusions undetected.

Anomaly-based intrusion detection: In this type of detection, the network packets are monitored for any abnormal or unusual behaviour using statistical approaches or anomaly heuristics. The major limitation of this type of detection is that it may result in high number of false positives.

EXISTING SYSTEMS

The table below presents an overview of the existing Intrusion detection techniques.

Table-1.

Name of the IDS	Description
SNORT	It is a free and open source network intrusion detection system that uses protocol analysis to detect various vulnerability exploit attempts.
OSSEC	It is a host–based IDS that performs integrity checking and has a powerful log analysis engine.
OSSIM	Open Source Security Information Management aims to provide a detailed view of all aspects of network and host devices.
Suricata	Open source–based IDS
Bro	Unix–based IDS which uses parsing technique for extracting semantics.
BASE	The basic analysis and security engine is a PHP–based analysis engine that processes a database of security events.
Sguil	It is used for network security analysis and facilitates event driven analysis.

PROPOSED SYSTEM

The proposed system works as follows:

A dataset containing 1000 records is considered and a clustering algorithm is applied to detect attacks. Then the signatures of the detected attacks are mapped to a series of binary values. A variation of apriori algorithm is then applied on the binary values to find the signatures that repeatedly lead to an attack.

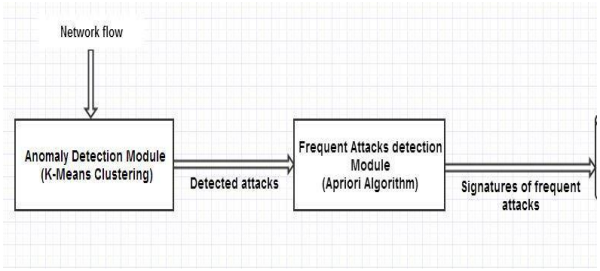

Figure-1. Architecture.

Anomaly -based detection

The input to the anomaly -based detection module is the KDD-99 data set. This dataset is comprised of 1000 connection records. A connection record consists of various attributes and their respective values. A few of the major attributes are duration, source bytes, destination bytes and protocol. These values are then passed to a clustering algorithm which uses the concept of Euclidean distance to cluster the values. Applying the clustering algorithm separately for different connection attributes (duration, source bytes and destination bytes) improves the detection quality. The signatures of detected attacks can be retrieved and stored in a database.

There are 4 steps in K-means clustering algorithm:

a) Consider K clusters.
b) Arbitrarily divide all elements into K clusters and calculate their centroids.
c) Iterate over all elements and compute the Euclidean distance to the centroids of all clusters.
d) Assign each element to the cluster with nearest centroids.
e) For both the modified clusters, recalculate the centroids.

Mapping

The output of the anomaly-based detection module is a set of signatures of detected attacks stored in a database. These signatures are then mapped into a series of binary values each specifying the category to which a particular attribute belongs.

Table-2.

Duration			Protocol		Service				Source bytes			
1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	1	0	1	0	0	0	0	0	0	1
0	1	0	1	0	1	0	0	0	1	0	0	0
0	0	1	0	1	0	1	0	0	0	1	0	0
0	1	0	1	0	0	0	1	0	0	0	0	1
0	0	1	0	1	0	1	0	0	0	0	1	0
1	0	0	0	1	0	0	0	1	1	0	0	0

1: Duration < 500 2: Duration < 1000
3: Duration >=1000
4: Protocol-TCP 5: Protocol-UDP
6: Service-FTP 7: Service-HTTP
8: Service-Telnet 9: Service-Others
10: Source bytes<10000
11: Source bytes<100000
12: Source bytes<280000
13: Source bytes>280000

Frequent attacks detection

The basic version of the apriori algorithm is modified to accept binary values as input and accordingly function. This new version of apriori algorithm is used to find the signatures that repeatedly lead to an attack.

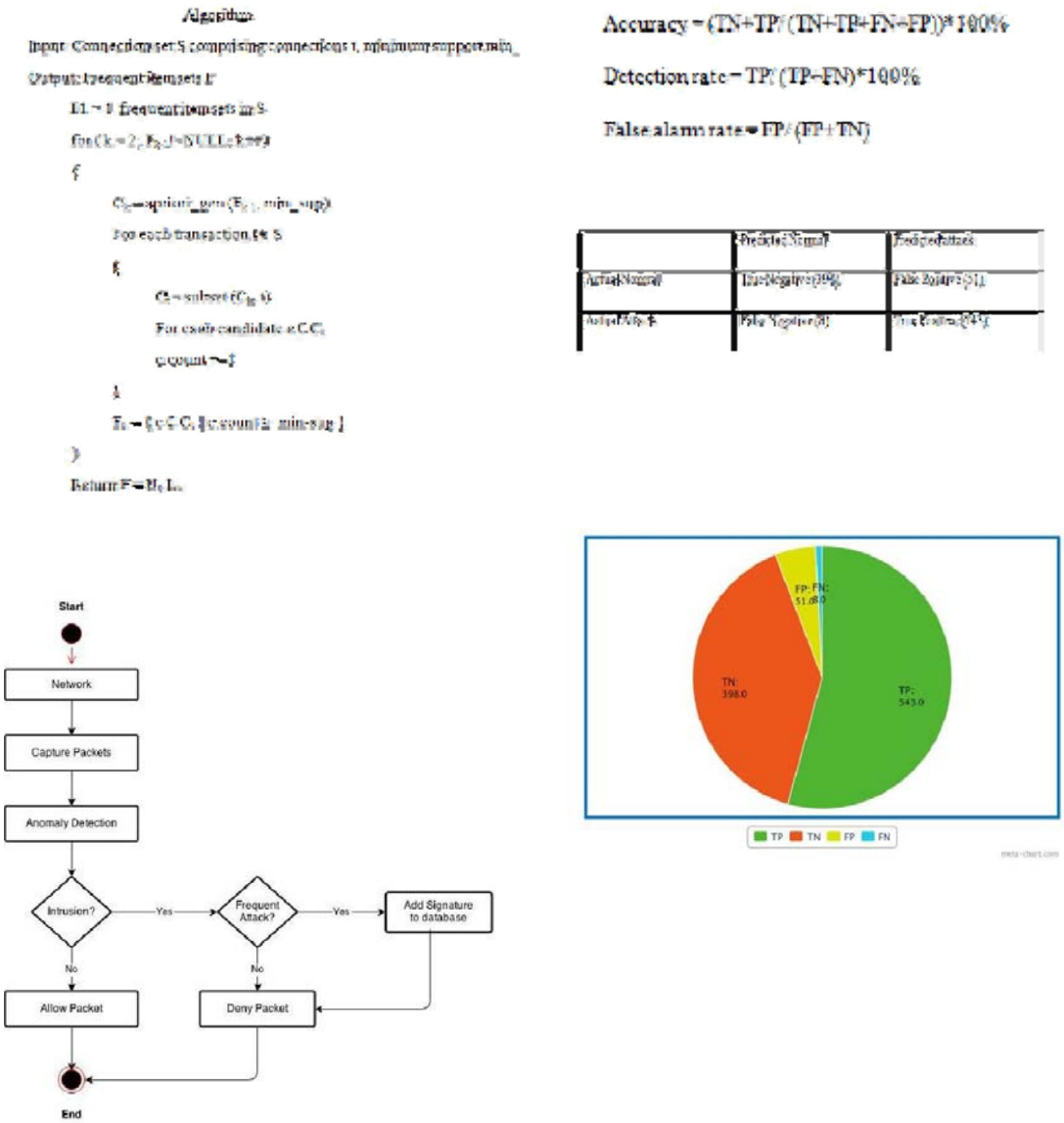

Figure-2. Flow chart of proposed IDS.

PERFORMANCE EVALUATION

Three factors have been considered for evaluating the performance of the proposed intrusion detection system:

a) Accuracy
b) Detection rate
c) False alarm rate

RESULTS

The above mentioned factors of our proposed Intrusion Detection System (IDS) are compared against those of Naive Bayesian. The proposed model yielded better results. The accuracy and detection rate of proposed system is more than that of Naive Bayesian and false alarm rate was determined to be lesser. The comparison is shown in the figure below:

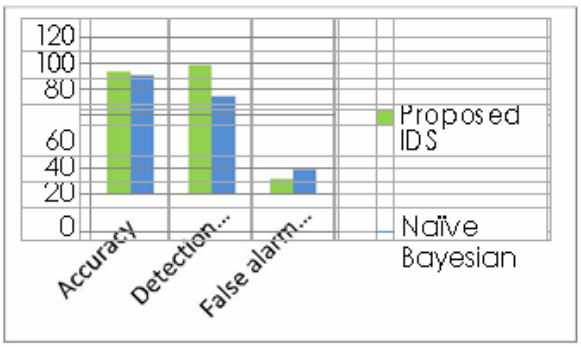

Figure-3.

CONCLUSIONS

The need for reliable defences is a crucial element in cloud architecture. Traditional IDS such as Signature -based IDS are incapable of detecting future unknown attacks. Anomaly - based IDS can detect those attacks. Applying the clustering algorithm separately for different connection attributes (duration, source bytes and destination bytes) improves the detection quality. The frequent attack detection module detects the frequent attacks, ensuring low false alarm rate and hence increasing reliability.

FUTURE WORK

The proposed IDS architecture uses an apriori algorithm to detect frequent attacks. The future research is to include a feedback mechanism such that the frequent attacks detected by the IDS are updated to the signature database. This will ensure that it doesn't remain as an unknown intrusion in future.

REFERENCES

[1] Mohammad Sazzadul Hoque, Md. Abdul Mukit and Md. Abu Naser Bikas. 2012. An implementation of intrusion detection system using genetic algorithm. International Journal of Network Security Its Applications (IJNSA). 4(2).

[2] Kanubhai K. Patel and Bharat V. Buddhadev. 2013. An Architecture of Hybrid Intrusion Detection System. International Journal of Information and Network Security (IJINS). 2(2).

[3] Farah Jemili, Dr. Montaceur Zaghdoud and Pr. Mohamed Ben Ahmed. 2010. A Framework for an Adaptive Intrusion Detection System using Bayesian Network. RIADI Laboratory, ENSI, Manouba University Manouba, Tunisia.

[4] Safwan Mawlood Hussein, Fakariah Hani Mohd Ali and Zolidah Kasiran. Evaluation Effectiveness of Hybrid IDS Using Snort with Naïve Bayes to Detect Attacks.

[5] H. M. Shirazi, A. Namadchian and A. khalili Tehrani. 2012. A Combined Anomaly Base Intrusion Detection Using Memetic Algorithm and Bayesian Networks. International Journal of Machine Learning and Computing. 2(5).

[6] Patel Hemant, Bharat Sarkhedi and Hiren Vaghamshi. 2013. Intrusion Detection in Data Mining With Classification Algorithm. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. 2(7).

[7] Vaidehi Kasarekar and Byrav Ramamurthy. Distributed Hybrid Agent Based Intrusion Detection and Real Time Response System. University of Nebraska Lincoln.

[8] Hari Om and Aritra Kundu. 2012. A Hybrid System for Reducing the False Alarm Rate of Anomaly Intrusion Detection System. 1st Int'l Conf. on Recent Advances in Information Technology, RAIT.

[9] David J. Day, Denys A. 2012. Flores and Harjinder Singh Lallie. CONDOR: A Hybrid IDS to Offer Improved Intrusion Detection. IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications.

[10] Kanubhai K. Patel, Bharat V. Buddhadev. 2013. An Architecture of Hybrid Intrusion Detection System.

[11] Poonam Dabas, Rashmi Chaudhary. 2013. Survey of Network Intrusion Detection Using K-Mean Algorithm.

[12] Maheshkumar Sabhnani, Gursel Serpen. KDD Feature Set Complaint Heuristic Rules for R2L Attack Detection.