Comprehensive study on Machine Learning Techniques for IDS in Cloud Computing

Ms. Pinal J. Patel [*1], Dr. J. S. Shah[#2] , Mr. Meghal Patel[#3]
* C.E. Department, Government College of Engineering, Bhavnagar
# C.E. Department, Gujarat Technology University, Gujarat, India
# CE Department, igate patni computers, Banglore

Abstract: - Cloud computing is one of the popular technologies where the user can easily use the resources of cloud services providers to perform their tasks and pay for the resources that they use. An easy accessibility condition cause computer network's vulnerable against several threats. Threats to networks are numerous and potentially overwhelming. So, Intrusion Detection Systems (IDS) have been used to detect malicious behaviours in network communication and hosts. We survey the existing types, techniques of Intrusion Detection Systems, and aspects of IDS in cloud computing in the literature. Finally, we compare these techniques.

Keywords: - Intrusion Detection, Cloud Computing

I. INTRODUCTION

There are many companies, labs are engaged in intrusion detection system research, and have completed development of prototype systems and products. Some research institutions and network security Products Company also carried out related research, but the domestic intrusion detection products are less. However, there are some drawbacks in current intrusion detection systems have like as i) insufficient detection rates, ii) too many intrusions detected or missed and the basic IDS have insufficient information to detect complex intrusions like distributed or coordinated attacks. A new intrusion detection system is based on cloud computing, with it, on any network site, a local detection engine analyzes the data collected by cloud computing center to find intrusion patterns. Afterwards, all the generated alerts are processed by a global intrusion detection engine to find more complex intrusions and to give a global view of the network security. [1]

The concept Cloud Computing, a new paradigm of software development and deployment has emerged. Cloud Computing extends an enterprise ability to meet the computing demands of its everyday operations, while offering flexibility, mobility and scalability. And it can be interpreted as the sum of Software as a Service (SaaS) and Utility Computing. There are some Cloud providers, which offer a specific virtualized infrastructure, and Cloud users, which use the provided services and infrastructure. Furthermore, there are three layers involved in Cloud Computing: i) The System Layer ii) The Platform Layer, and iii) The Application Layer.

The hardware layer is the basis for Cloud computing and is not provided directly to the user. Therefore we do not consider it as a part of the Cloud itself.

The system layer includes the virtual machine (VM) abstraction of a server and related virtual networks. The platform layer includes the infrastructure software, e.g., a virtualized operating system (OS) of a server or the runtime and the API of a specific programming language. Finally, the application layer includes other software running in the Cloud, such as a web application. An important property of Cloud computing is that the user has only partial control over the infrastructure being used, i.e., the user can only control the used services on the specific layer. Important examples for the different layers are: Google App Engine, Windows Azure and Amazon EC2. [2].

II BACKGROUND

Before discussing the most relevant approaches to IDS, we proceed to describe the fundamental elements inside the intrusion detection problem.

2.1. Attack definition and classification

A computer attack can be defined as the intelligence of evading or evading attempt of computer security policies, acceptable use policies, or standard security practices. In the security research community, the terms attack and intrusion are often used with the same meaning [18]

By Kendall [19], in which attacks can be classified into four categories:

Probing: Attacks oriented to gather information about the system, for further intrusion. These attacks include network traffic sniffing and port/address scanning.

Denial of Service (DoS): Attacks attempting to diminish or totally interrupt the use of a system or a service to their legitimate users.

User to Root (U2R): Attacks that aim to gain super user access to the system by means of exploiting vulnerabilities

in operating systems or software applications. The attacker has a valid account in the system.

Remote to Local (R2L): Attacks oriented to gain local access from outside the network.

2.2. A simple IDS architecture
In general, from an architectonic point of view, IDS is based on the following modules:

Traffic Data Acquisition: This module is used in the data collection phase. In the case of IDS, the source of the data are raw network frames or information from upper protocol layers (i.e. IP or UDP protocols).

Traffic Features Generator: This module is responsible for extracting a set of selected traffic features from captured traffic.

Network traffic features can be classified in low-level features and high-level features. A low-level feature can be directly extracted from captured traffic (e.g. IP header). Whereas a high-level feature consists of traffic information deduced from captured traffic by a subsequent process. Features can be also classified according to the network traffic source used for generating them. Packet features are those directly obtained from network raw packets headers [18]

Incident Detector: This module processes the data generated by the Traffic Features Generator module to identify intrusive activities.

Finally, we summarize the four more relevant measures when considering true deployment feasibility:
Prediction accuracy: Measures how good is a IDS in detecting intrusion.

Processing time: Considers the rate at which events are processed. IDS should be able to perform detection as soon as possible.

Adaptability: Indicates the IDS capability to deal with new attacks techniques. IDS should be able to readapt it in the presence of new threats.

Resource consumption: Measures how much memory and storage resources are required by the system.

III TYPES OF IDS

Intrusion detection system sets off alerts about detected intrusions so that a system administrator or the system itself may take appropriate action. In general, IDS collects network traffics, analyzes these traffics, and makes response or alerts the network to the manager if there is an intrusion taking place. Thus, the aim of the IDS is to alert or notify the system that some malicious activities have taken place and try to eliminate it.

According to the method of the collection of intrusion data, all the intrusion detection systems can be classified into two types: host-based and network-based IDSs. Host-based intrusion detection systems (HIDSs) analyze audit data collected by an operating system about the actions performed by users and applications; while network-based intrusion detection systems (NIDSs) analyze data collected from network packets.

IDSs analyze one or more events gotten from the collected data. According to analysis techniques, IDS system is classified into two different parts: misuse detection and anomaly detection. Misuse detection systems use signature patterns of exited well-known attacks of the system to match and identify known intrusions. Misuse detection techniques, in general, are not effective against the latest attacks that have no matched rules or pattern yet. Anomaly detection systems identify those activities which deviate significantly from the established normal behaviours as anomalies. These anomalies are most likely regarded as intrusions. Anomaly detection techniques can be effective against unknown or the latest attacks. However, anomaly detection systems tend to generate more false alarms than misuse detection systems because an anomaly may be a new normal behaviour or an ordinary activity.

While IDS detects an intrusion attempt, IDS should report to the system administrator. There are three ways to report the detection results [6]: notification, manual response, and automatic response. In notification response system, IDS only generates reports and alerts. In manual response system, IDS provides additional capability for the system administrator to initiate a manual response. In automatic response system, IDS immediately respond to an intrusion through auto response system.

IV EXISTING TECHNIQUES

Many techniques of detection intrusions are reviewed in [16, 17]

3.1 Signature based detection

Signature based intrusion detection attempts to define a set of rules (or signatures) that can be used to decide that a given pattern is that of an intruder. As a result, signature based systems are capable of attaining high levels of accuracy and minimal number of false positives in identifying intrusions. Little variation in known attacks may also affect the analysis if a detection system is not properly configured (Brown et al., 2002). Therefore, signature based detection fails to detect unknown attacks or variation of known attacks.

3.2 Anomaly based detection

Anomaly (or behavioral) detection is concerned with identifying events that appear to be anomalous with respect to normal system behavior. A wide variety of techniques including data mining, statistical modeling and

hidden markov models have been explored as different ways to approach the anomaly detection problem. Anomaly based approach involves the collection of data relating to the behavior of legitimate users over a period of time, and then apply statistical tests to the observed behavior, which determines whether that behavior is legitimate or not. The ability of soft computing techniques to deal with uncertain and partially true data makes them attractive to be applied in intrusion detection (Moradi and Zulkernine, 2004). There are many soft computing techniques such as Artificial Neural Net- work (ANN), Fuzzy logic, Association rule mining, Genetic Algorithm (GA), etc. that can be used to improve detection accuracy and efficiency of signature based IDS or anomaly detection based IDS.

3.3 Artificial neural network (ANN) based IDS

The goal Of Using ANNs (Han and Kamber, 2006) for intrusion detection is to be able to generalize data (from incomplete data) and to be able to classify data as being normal or intrusive (Ibrahim, 2010). Types of ANN used in IDS areas (Ibrahim, 2010): Multi-Layer Feed-Forward (MLFF) neural nets, Multi-Layer Perceptron (MLP) and Back Propagation(BP).

ANN based IDS is an efficient solution for unstructured network data. The intrusion detection accuracy of this approach is based on number of hidden layers and training phase of ANN. An approach proposed by Vieira et al. (2010), uses ANN based anomaly detection technique for Cloud environment, which requires more training samples as well as more time for detecting intrusions effectively.

3.4 Genetic Algorithms (GAs)

Genetic algorithms mimic the natural reproduction system in nature where only the fittest individuals in a generation will be reproduced in subsequent generations, after undergoing recombination and random change. The application of GAs in IDS research appeared as early as 1995 [12], and involves evolving a signature that indicates intrusion. A related technique is the Learning Classifier System (LCS), where binary rules are evolved, that collectively recognizes patterns of intrusion.

3.5 Decision Tree

It is a model of decisions and also can be used to show possible consequences for particular occurrences where there are conditional probabilities for each occurrence. Those occurrences of attacks form a tree-based structure that contains root node and a number of leaf nodes. Decision tree generally performs very efficiently even if dealing with a large amount of data [13].

3.6 Bayesian Network

A set of transition rules are represented as probabilistic independencies in a graphical model. Each node contains the state of random variable and a conditional probability table, which determine the probabilities of the node in a state, given a state of its parent [14]. An advantage of the approach is that it can deal with incomplete data.

3.7 Fuzzy logic

It is a set of concepts and approaches designed to handle vagueness and imprecision. A set of rules can be created to describe a relationship between the input variables and the output variables, which may indicate whether an intrusion has occurred. Fuzzy logic uses membership functions to evaluate the degree of truthfulness [15].

3.8 Association Rules

Some intrusion attacks are formed based on known attacks or variant of known attacks. To detect such attacks, signature apriori algorithm (Han et al., 2002) can be used, which finds frequent subset (containing some features of original attack) of given attack set. Han et al. (2002) proposed network based intrusion detection using data mining technique. In this approach, signature based algorithm generates signatures for misuse detection. However, drawback of the proposed algorithm is its time consumption for generating signatures. Zhengbing et al. (2008) solved the database scanning time problem examined in Han et al. (2002). They proposed scanning reduction algorithm to reduce number of database scans for effectively generating signatures from previously known attacks. However, it has very high false positive alarm rate since unwanted patterns are produced. Lei et al. (2010)

3.8 Hybrid Technique

Hybrid techniques use the combination of two or more of above techniques. Hybrid technique can combine two low level components ex. fuzzy logic for misuse detection and neural networks for anomaly detection. It is an effective model, which does not require dynamic updates of rules.

V COMARISIONS

IDS Techniques	Characteristics/ Advantages	Limitations/challenges
Signature based detection	-It Identifies intrusion by matching captured patterns with predefined knowledge base -High detection accuracy for known attack	-Cannot detect new attack -High false alert rate for unknown attacks
Anomaly detection base IDS	-Uses statistical test on collected behavior to identify intrusion -Can lower false alert rate for unknown attacks	-More time is required to identify attack -Detection accuracy is based on amount of collected behavior
ANN based IDS	-Classifies unstructured n/w packet efficiently -Multiple hidden layers in ANN increase efficiency of classification	-Requires more time and more samples training phase
Fuzzy logic based IDS	-Used for quantitative features -Provides better flexibility to some uncertain problems	-Detection accuracy is lower than ANN
Association rules based IDS	-Used to detect known attack signature	-It cannot detect totally unknown attacks -It requires more number of database scans to generate rules
GA based IDS	-It is used to select best features for detection -Has better efficiency	-It is complex method
Decision Tree based IDS	-It is simple to use and easy to understand	-Small change in input data can cause large changes in tree
Bayesian Network based IDS	-It is robust in nature	-It is computationally difficult to explore a previously unknown attack
Hybrid techniques	-It is an efficient approach to classify rules accurately	-Computational cost is high

Table 1 [17]

CONCLUSION

We discussed several intrusions which can threat integrity, confidentiality and availability. Firewall only may not be adequate to solve security issues. This paper emphasized the basic terminology of IDS and usage of alternative options to incorporate intrusion detection into Cloud. Recent soft computing incorporating IDS in Cloud have been discussed with their advantages and disadvantages. The adoption of soft computing techniques in IDS can improve the security.

REFERENCES

[1] WANG Xin, HUANG Ting-Iei, LIU Xiao-yu **"**Research on the Intrusion Detection Mechanism based on Cloud Computing" 2010 IEEE
[2] Sebastian Roschke, Feng Cheng, Christoph Meinel "Intrusion Detection in the Cloud" 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing
[3] G. Carl, G. Kesidis, R.R, Brooks, and S. Rai, "Denial-of-serive attack-detection techniques," IEEE Transaction on Internet Computing, Vol.10, issue 1, 2006, pp.82-89
[4] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer Networks*, 2007.
[4] J. Haggerty, S. Qi, and M. Merabti, "Early detection and prevention of denial-of-service attacks: a novel mechanism with propagated traced-back attack blocking," IEEE Journal on Selected Areas in Communications, Vol.23, Issue 10, Oct.2005, pp.1994-2002
[5] M.H., Islam, K. Nadeem, S.A., Khan, "Efficient placement of sensors for detection against distributed denial of service attack," International Conference on Innovations in Information Technology, 2008, 16-18 Dec. 2008, pp.653-657
[6] D.J. Ragsdale, C.A. Carver, Jr. J.W. Humphries, U.W. Pooch, "Adaptation techniques for intrusion detection and intrusion response systems," 2000 IEEE International Conference on Systems, Man, and Cybernetics, Vol.4 , 8-11 Oct. 2000 p.2344-p.2349
[7] E.H, Spafford and D. Zamboni, "Intrusion Detection Using Autonomous Agent," Computer Networks, vol.34, issue 4, 2000, pp.547-570
[8] S. Cheung, R. Crawford, and M. Dilger et al., "The Design of GrIDS: A Graph-Based Intrusion Detection System," Technical Report CSE- 99-2, U.C. Davis Computer Science Department, January 1999
[9] S.R. Snapp, J. Brentano, G.V. Dias, T.L. Goan, T. Grance, L.T. Heberlein, C.L. Ho, K.N. Levitt, B. Mukherjee, D.L. Mansur, K.L. Pon, and S.E. Smaha, "A system for distributed intrusion detection," Compcon Spring'91, Feb-March 1991, pp.170-176
[10] D. Curry and H.Debar, "Intrusion Detection Message Exchange Format Data Model and Extensible Markup Language (XML) Document Type Definition," draft-ietf-idwg-idmef-xml-06.txt, Feb. 2002
[11] Igor Muttik,Chris Barton, "Cloud security technologies," information security technical report, 2009 Elsevier Ltd All rights reserved, April 2009
[12] Abdoul Karim Ganame,Julien Bourgeois,Renaud Bidou,Francois Spies, "A global security architecture for intrusion detection on computer networks", computers & security 27(2008)30-47,. March 2008
[13] SUN Yun, HUANG Hao , "Hybrid Network Intrusion Detection System," Computer Engineering, vo1.34,NO.9, May 2008
[14] M.AliAydin,A.HalimZaim,K.G6khanCeylan , "A hybrid intrusion detection system design for computer network security," Computers and Electrical Engineering 35(2009)517-526, February 2009.
[15] User Mode Linux (UML), Website: http://user-modelinux. sourceforge.net/ (accessed Oct 2009).
[16] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," Computer Networks, 2007.
[17] Chirag Modi a,n, DhirenPatel, BhaveshBorisaniya , HirenPatel , Avi Patel , MuttukrishnanRajarajan, "A survey of intrusion detection techniques in Cloud", Journal of Network and Computer Applications 36 (2013) 42–57
[18] Carlos A. Catania a,⇑ , Carlos García Garino," Automatic network intrusion detection: Current techniques and open issues", Computers and Electrical Engineering 38 (2012) 1062–1072
[19] Kendall K. A database of computer attacks for the evaluation of intrusion detection systems. Master's thesis, AAI3006082; 1999.