International Journal of Distributed and Cloud Computing
6 (2) December 2018, 19-23
http://www.publishingindia.com/ijdcc

Intrusion Detection Using Data Mining in Cloud Computing Environment

Pinki Sharma[1*], Jyotsna Sengupta[2] and P. K. Suri[3]
[1]Research Scholar, Department of Computer Science, Punjabi University, Patiala, Punjab, India.
Email: pinkisharma@gmail.com
[2]Professor, Department of Computer Science, Punjabi University, Patiala, Punjab, India.
Email: jyotsna.sengupta@gmail.com
[3]Former Professor, Kurukshetra University, Kurukshetra, Haryana, India. Email: pksurikuk@gmail.com
*Corresponding Author

Abstract: Nowadays cloud computing is widely accepted a paradigm. At present large amount of data is transferred between cloud and user and vice-versa. That data in transient is exposed to various intrusions. Therefore, security is the primary concern of cloud computing environment. Firewall and other security techniques can act as first line of defence and cannot provide a robust security solution. Intrusion detection systems proved to be best solutions to various attacks. Data mining techniques have emerged to make it less vulnerable and thus to analyze data and to determine various kind of attack. Both signatures based and anomaly based techniques effectively and efficiently used data mining techniques for any kind of attack detection. This paper presents various data mining techniques used in intrusion detection. This paper also reviews various cloud intrusion detection systems that uses data mining techniques for attack detection.

Keywords: Cloud computing, Cloud intrusion detection systems, Data mining, IDS, Intrusion detection.

I. Introduction

In last few years cloud computing has changed the IT world with rapidly emerging and widely accepted paradigm for computing systems. Cloud computing provides on demand services, less maintenance cost, offers data and services availability assurance, dynamically scalable and virtualised resources, reconfigurable aggregation, sharing, computational storage across network anytime, anywhere [1,2]. Cloud computing provides three basic abstraction layers. First is system level layer that is a virtual machine generalization of the server. Second is platform level layer which is a virtualised operating system of a server. And third is application level layer that include web applications. It also offers Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and Software as a Service (Saas). PaaS provides platform to users to develop various applications and run them. IaaS facilitates user by maintaining large infrastructure services to users like hosting servers, managing network and other resources for client. SaaS ensures use of hassle free installing and running software services on its machine [1]. Today some leading cloud service providers are Google, salesforce.com and Amazon who provide computation and application, services for storage as pay per basis.

Cloud provides a rapid and location independent information processing. Due to distributed information processing, trust is one of the major issue among the cloud users for using its resources. As indicated by IDC survey, security is the major concern of cloud computing [3]. Lockheed Martin's [4] cyber security division white paper states that foremost important security concern after information security is attack detection and prevention in cloud infrastructure. Abuse of vulnerabilities existing in cloud influences the confidentiality, availability and integrity of cloud resources and services offered. Most important security issue in cloud computing is to protect against various network attacks. The possible attacks are DNS poisoning, IP spoofing, ARP spoofing, man in the middle attack, RIP attack, port scanning, DoS, DDoS etc. [2]. Various technologies such as message encryption and firewall protect the network and can be used as first line of defence.

To overcome this situation an efficient Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) should be incorporated in cloud infrastructure to lessen these attacks. The role of IDS in the security of cloud is very important in view of the fact that it acts as additional preventive layer of security and apart from detecting only known intrusions, it can be able to detect variants of several known attacks and unknown attacks.

One of the important data mining applications is intrusion detection. Various techniques of data mining can enhance the working of an intrusion detection and prevention system in various ways. Both signature-based or misuse detection and anomaly-based or behaviour based detection can employ data mining techniques for the intrusion detection purpose. Training data can be labelled as either "normal" or "intrusion" in misuse or signature based detection. A classifier would then be able

to be determined to identify known intrusions. Signature based has the application of association rule mining, classification algorithms, and cost-sensitive modelling. In anomaly-based detection methods, the models of normal behaviour are constructed that automatically detects significant deviations from the normal behaviour. Anomaly based methods includes the application of classification algorithms, clustering, outlier analysis, and statistical approaches [5].

Rest of the paper is organized as follows: In section 2, we briefly describe the intrusion detection systems. Section 3 discusses possible data mining techniques for intrusion detection. Section 4 describes utilisation of data mining techniques for cloud Intrusion Detection Systems (IDS). Section 5 concludes the paper.

II. Intrusion Detection System

Intrusion Detection System in simple term is the system or hardware / software which can detect or prevent the system from insider and outsider attacks where it is placed. To prevent network infrastructure from web threats, cybercrimes, and of course from internal attacks, IDS has been proven to be the most important intrusion detection tool. IDS proved to be a major tool which is used for network administrator to defend networks from threats, worms, and insider attacks. An Intrusion Detection System (IDS) has been proposed for years as a most effective security measure [5]. For intrusion detection purpose basic two traditional IDS techniques are used: i) Signature Based IDS and ii) Anomaly Based IDS.

A. Signature Based IDS

Signature based IDS also termed as Misuse based IDS. In this type of intrusion detection technique redefined dataset / pattern which is generally called as signature provided by the system. This predefined data set has been generated by the security experts. In Signature based Intrusion Detection System (SIDS) can detect known attacks through matching signature in predefined attack pattern. Unfortunately, this technique is capable to identify only known and predefined patterned attacks. The Major drawback of this technique is that they are unable to catch totally new malicious activity or say unknown threats. Though, their execution of identification is extremely high. Thus, it comes to the new technique of IDS (i.e. Anomaly based IDS). These signatures are composed of several elements which are defined by network traffic.

B. Anomaly Based IDS

To monitor system's behaviour the IDS used are Anomaly based IDS (AIDS). Suppose, one pattern which is predefined which comes through the network traffic during the communication between two or more than that, at this time pattern changed or say intruder attacks on that which cannot identify by the signature based IDS. But when it used anomaly based IDS, it checks the whole behaviour of out coming packets as well as insider packets. If it founds any kind of behavioural changes in that it will deny the packet and send it to log and then send it to the reporting system. This technique overcomes the issue related to detect unknown and abnormal behavioural activities. But by using this technique system gets high false alarm rate. Various machine learning and data mining techniques / algorithms used for anomaly detection techniques [6].

Table I: Comparative Analysis of Signature Based and Anomaly Based Techniques

	Signature Based IDS	Anomaly Based IDS
Pros	Uses the patterns of known attacks (signatures) to identify intrusions.	Uses the deviation from normal usage patterns to identify intrusions.
Cons	Known attacks have to be stored first. Unable to detect totally new attacks. Need signatures updated regularly.	Has to study sequential interrelation between transactions and more false positives.
Alarm	Low	High
Rate Speed	High	Low
Resource Consuming	Low	High
Flexibility	Low	High
Reliability	High	Moderate
Scalability	Low	High
Robustness	Low	High

III. Data Mining Techniques for Intrusion Detection

Data mining algorithms can be utilised for both signature or misuse based detection and behaviour or anomaly based detection. In misuse detection, training data are labelled as either "normal" or "intrusion." After that a classifier would then be able to be inferred to recognize or to detect known attacks [7]. Classification algorithms, association rule mining, and cost-sensitive modelling are the various data mining techniques that are used for misuse or signature based intrusion detection applications. Anomaly detection based systems typical constructs models of entity's normal behaviour and naturally detects significant deviations from the normal behaviour. Either supervised or unsupervised learning can be used for the detection of intrusion using anomaly based detection [8]. Supervised approach constructs model based on training data that are known to be "normal." Unsupervised approach has no information about the training data, detection models learn themselves. Classification algorithms, statistical approaches, clustering, and outlier analysis are the various data mining techniques used in anomaly detection. The data mining methods utilized must be effective and versatile, and fit for taking

care of system information of high volume, dimensionality, and heterogeneity. Both signatures based and anomaly based detection centres around the varied classification of existing data mining techniques [5].

A. Clustering Based Detection Techniques

Clustering can be characterized as a division of information into collections of comparatively similar objects. In each cluster, objects that are similar to one another and dissimilar to objects are assembled in same cluster [7]. Clustering algorithms are capable to distinguish intrusion data without any prior information. Various clustering techniques are available that has applications in intrusion detection [8].

B. Classification Based Detection Techniques

Classification can be characterize as a problem of recognizing the category of objects on the basis of a training set of data containing observations whose category membership is known. The category can be termed as class label [8]. Many objects can belong to one or many of the class labels. In machine learning, classification is considered as an example of supervised learning. An algorithm that implements classification is known as a classifier. It is constructed to predict categorical labels or class label attribute. For the purpose of intrusion detection it will classify the data generally into two categories namely normal or abnormal.

C. Association and Correlation Based Detection Techniques

Association and correlation mining can be utilised to discover relationships between system attributes of network data. Such type of information can give insight into the selection of useful attributes for intrusion detection. New attributes derived from aggregated data may also be helpful in many ways, such as summary counts of traffic matching a particular pattern [5].

D. Stream Data Based Detection Techniques

The nature of intrusions and malicious attacks is transient and dynamic, because of it, it is critical to perform intrusion detection in the data stream environment. Furthermore, an event may be normal on its own, yet thought to be malevolent if seen as a major aspect of a succession of events. Accordingly it is important to analyse what sequences of events are as often as possible experienced together, find sequential patterns, and identify outliers. Other data mining methods for finding evolving clusters and building dynamic classification models in data streams are likewise important for ongoing intrusion detection [5].

Table II: Comparison of Data Mining Techniques

Data Mining Technique	Pros	Cons
Clustering	Able to detect intrusions without prior knowledge. Ability to discover some or all hidden patterns. More efficient.	If the anomalies in the data form clusters by themselves, these techniques will not be able to detect such anomalies. Expensive training time.
Classification	Effective for both misuse detection and anomaly detection, but more frequently used for misuse detection.	Needs labelled data, expensive training time, Less efficient High false negatives and false positives.
Association and correlation	Does not classifies the data.	Needs two threshold minimum support and confidence.
Stream data	Maintains profiles of normal activities. Identifies novel attacks. Identifies clusters and outlers in traffic data real time intrusion detection.	High rate of false positive.

IV. Cloud Intrusion Detection System Using Data Mining Techniques

In recent times, there has been an expanded enthusiasm for data mining based approaches to construct detection models for intrusion detection systems in cloud computing environment. These are generalizing from both well known attacks and normal behaviour of the system in order to detect unknown or new attacks. These models can be constructed in a quicker and more automated fashion than manually encoded models that require difficult and critical analysis of audit data by domain experts. A few compelling data mining techniques for detecting intrusions have been developed for cloud computing by many researchers, considerable number of which perform better than and closed to systems engineered by domain experts.

P. Ganesh Kumar and N. Pandeeswari [9] developed a hypervisor model ANFIS to detect the malicious activities. Due to fail in construction of model for target data, the ANFIS model provided the learning and adaptation proficiency of neural network. Thus, the hypervisor detector was designed

and developed by the hybrid algorithm of back propagation gradient descent technique with the least square method. According to authors, the ANFIS model achieved the minimum false negative rate and high detection accuracy. But, the small number of instance led to degrade the detection performance.

Chirag Modi *et al.* [10] design and integrate Bayesian Classifier and Snort based Network Intrusion Detection System (NIDS) in Cloud. The proposed NIDS framework intends to detect network intrusions in Cloud environment. The proposed framework uses Bayesian classification algorithm and Snort. Bayesian classifier predicts that weather the given event is attack or not by constantly examining the previously stored network events and the Snort is used to detect known attacks. This module ensures low false positives and low false negatives with affordable computational cost.

Chirag Modi *et al.* [11] proposes a snort and signature apriori based framework integrating Network Intrusion Detection System (NIDS) in the Cloud. Proposed framework generates new rules from captured packets and that new rules are appended in the Snort configuration file to improve efficiency of Snort. This integrated framework can detect known attacks and as well as the derivative of known attacks in Cloud by monitoring network traffic, while ensuring low false positive rate with reasonable computational cost.

N. Pandeeswari and Ganesh Kumar [12] deploys a hypervisor detector at the virtual machine monitor layer. The Hypervisor Detector is constructed with a combination of Fuzzy C-Means clustering and Artificial Neural Network. The first phase of FCM-ANN is fuzzy clustering module which is used to divide the large dataset into small clusters so as to improve the learning capability of ANN. Fuzzy clustering module enhances the performance of artificial neural network. In second phase, various ANN modules are trained according to their cluster values. In third phase, Fuzzy aggregation module is used to combine the results of various ANN. Authors compares the hypervisor detector with Naïve Bayes and classic ANN. The authors claim that FCM-ANN outperforms the Naïve Bayes and the classic ANN algorithms even for low frequent attacks.

The authors, Vereia *et al.* [13] have proposed a Grid and Cloud Computing Intrusion Detection System (GCCIDS). GCCIDS integrates both knowledge and behaviour analysis to discover the intrusions. This system makes use of an event auditor that captures data from various resources like system logs, node messages and services. Based on the captured data, the IDS service can be used to detect intrusions by using behaviour based and knowledge based techniques. GCCIDS uses artificial neural network for behaviour analysis.

Z. Chiba *et al.* [14] described the Cooperative and hybrid based network IDS system (CH-NIDS) using the Back Propagation Neural network (BPN). They developed the BPN model based on Snort and Optimized method. The snort prior in the BPN was used to detect the unknown attacks. Due to low convergence of BPN, they exploited the optimization algorithm to optimize the parameters which enhanced the detection rate and accuracy. Also, the snort and optimized based BPN was also used to detect the DoS and DDoS attacks by sharing alerts in central log. Thus, simulation results were evaluated to improve the detection rate and mitigate the false rate.

Sivakami *et al.* [15] proposes the usage of type-2 fuzzy neural network based on genetic algorithm to incorporate intrusion detection techniques into cloud. These systems are intelligent to gain knowledge of fuzzy sets and fuzzy rules from data to detect intrusions in a cloud environment.

Table III: Various Cloud Intrusion Detection System Using Data Mining for Attack Detection

Authors	Methods	Advantages	Disadvantages
P. Ganeshkumar and N. Pandeeswari [9]	Adaptive Neuro Fuzzy Inference System	Resolves the non-deterministic and non-linear problems.	Due to minimum number of instances, the performance gets degraded.
Chirag Modi *et al.* [10]	Bayesian Classifier snort	Low false positive. Low false negative. Affordable computational cost.	Complexity increased due to integration of both, signatures and anomalies.
Chirag Modi *et al.* [11]	Signature apriori Snort	Can detect derivatives of known attacks.	Cannot detect unknown or new attacks.
N. Pandeeswari and Ganesh Kumar [12]	Fuzzy clustering based ANN	Attains highest detection accuracy.	The number of instances is not sufficient to analyse the performance.
Veira *et al.* [13]	Neural networks	Detect and prevent only specify kind of attack. False rate for unknown attack is lower since ANN used.	Low learning cannot detect insider attack. Low efficiency. Requires more training time and samples for detection accuracy.

Authors	Methods	Advantages	Disadvantages
Z. Chiba *et al.* [14]	Back propagation neural network	Enhances the detection rate.	It is not applicable in an open source cloud environment.
Sivakami Raja and Saravanan Ramaiah [15]	Type-2 Fuzzy neural network	The accuracy of fuzzy rule is enhanced through the optimal learning algorithm.	Increased complexity in terms of speed.

V. Conclusion

The security of Cloud computing model must be thought about essentially for its success. In this paper, we have studied various intrusions detection system that uses data mining techniques for intrusion detection. Then we have comprehensively illustrated different types of IDS in cloud environment. Data mining techniques for intrusion detection proves to be a better choice for both misuse detection and anomaly detection. Data mining techniques are helpful in detecting totally new attack and derivatives of well known attacks.

References

[1] I. Gul, and M. Hussain, "Distributed cloud intrusion detection model," *International Journal of Advanced Science and Technology*, vol. 34, September 2011.

[2] U. Oktay, and O. K. Sahingoz, "Attack types and intrusion detection systems in cloud computing," In *6th International Information Security & Cryptology Conference*, Turkey, pp. 71-76, 20-21 September 2013.

[3] F. Gens, "New IDC IT cloud service survey: Top benefits and challenges, IDC exchange," 2009. Available: http://blogs.idc.com/ie/?p=730S

[4] L. Martin, White Paper, 2010. Available: /http://www.lockheedmartin.com/data/assets/isgs/documents/CloudComputingWhitePaper.pdf

[5] J. Han, and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed., Morgan Kaufmann, 2006.

[6] H. J. Patel, and R. Patel, "A Survey on intrusion detection system in cloud," *International Journal of Engineering and Technical Research (IJETR)*, vol. 2, no. 5, pp. 38-39, May 2014.

[7] P. Berkhin, "A survey of clustering data mining techniques," *Grouping Multidimensional Data*, Springer Berlin Heidelberg, pp. 25-71, 2006.

[8] S. Agrawal, and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Procedia Computer Science*, vol. 60, pp. 708-713, 2015.

[9] P. Ganeshkumar, and N. Pandeeswari, "Adaptive neuro-fuzzy-based anomaly detection system in cloud," *International Journal of Fuzzy Systems*, vol. 18, no. 3, pp. 367-378, June 2016.

[10] C. N. Modi, D. R. Patel, A. Patel, and R. Muttukrishnan, "Bayesian classifier and snort based network intrusion detection system in cloud computing," *Third International Conference on Computing, Communication and Networking Technologies*, 26-28 July 2012.

[11] C. N. Modi, D. R. Patel, A. Patel, and R. Muttukrishnan, "Integrating signature apriori based Network Intrusion Detection System (NIDS) in cloud computing," *Procedia Technology*, vol. 6 pp. 905-912, 2012.

[12] N. Pandeeswari, and G. Kumar, "Anomaly detection system in cloud environment using fuzzy clustering based ANN," *Mobile Networks and Applications,* vol. 21, no. 3, pp. 494-505, June 2016.

[13] K. Vieira, A. Schulter, C. B. Westphall, and C. M. Westphall, "Intrusion detection for grid and cloud computing," *IT Professional*, IEEE, vol. 12, no. 4, pp. 38-43, July-August 2010.

[14] Z. Chiba, N. Abghour, K. Moussaid, A. E. Omri, and M. Rida, "A cooperative and hybrid network intrusion detection framework in cloud computing based on snort and optimized back propagation neural network," *Procedia Computer Science*, vol. 83, pp. 1200-1206, 2016.

[15] S. Raja, and S. Ramaiah, "An efficient fuzzy-based hybrid system to cloud intrusion detection," *International Journal of Fuzzy Systems*, vol. 19, no. 1, pp. 62-77, February 2017.