ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

Intrusion Detection System for Cloud Network Using FC-ANN Algorithm

Swati Ramteke[1], Rajesh Dongare[2], Komal Ramteke[3]

Student, Department of Information Technology, VIIT, Pune, India [1]

Student, Department of Information Technology, MIT COE, Pune, India [2]

Assistant Professor, Department of Information Technology, Rajiv Gandhi College of Engineering and Research, Nagpur, India [3]

Abstract: Due to increasing incidents of cyber attacks, building effective intrusion detection systems are essential for protecting information systems security, and till now it remains an elusive goal and a big challenge. We present the state-of-the-art of the evolution of intrusion detection systems and address some of the research challenges to design efficient and effective intrusion detection systems . In existing IDS system the rule sets of various attack patterns are stored in databases and the whole network traffic is matched against it to avoid any unauthorized and illegal activities. if any attack happens, and the pattern of that attack is not stored in IDS rule sets then that attack pattern is need to be manually updated in the database to prevent it next time. Hence we are proposing such a system in which there is no need to manually updating the attack pattern in IDS rule sets, the proposed FC-ANN algorithm will automatically capture the patterns of new attack and store it in IDS database which will reduces the human time as well as effort to learn new attacks pattern manually.

Keywords: Cloud computing, Intrusion detection system, Attacks, security.

I. INTRODUCTION

The term cloud is analogical to "Internet". The term cloud computing is based on cloud drawing used in the past to represent telephone networks & later to depict internet in. Cloud computing is internet based computing where virtual shared servers provide software, infrastructure, platform, devices and other resources and hosting to customer as a service on pay-as you-use basis. Fig. 1 shows the concept [1]. With the coming of Internet age, network security has become the key foundation to web applications such as online retail sales, online auctions, etc. Intrusion detection attempts to detect computer attacks by examining various data records observed in processes on the network. Detection precision and detection stability are two key indicators to evaluate intrusion detection systems (IDS). In order to enhance the detection precision and detection stability, many researchers have been done and in the early stage, the research focus lies in using rule-based expert systems and statistical approaches.

But when encountering larger datasets, the results of rule-based expert systems and statistical approaches become worse. Thus a lot of data mining techniques have been introduced to solve the problem. Among these techniques, Artificial Neural Network (ANN) is one of the widely used techniques and has been successful in solving many complex practical problems and ANN has been successfully applied into IDS [2]. However, the main drawbacks of ANN-based IDS exist in two aspects: (1) lower detection precision, especially for low-frequent attacks, e.g., Remote to Local (R2L), User to Root (U2R), and (2) weaker detection stability.

Fig. 1 Cloud Computing

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

For the above two aspects, the main reason is that the distribution of different types of attacks is imbalanced. For low-frequent attacks, the leaning sample size is too small compared to high-frequent attacks. It makes ANN not easy to learn the characters of these attacks and therefore detection precision is much lower. In practice, low-frequent attacks do not mean they are unimportant. Although prior research has proposed some approaches, when encountering large datasets, these approaches become not effective.

To solve the above two problems, we propose a novel approach for ANN-based IDS, FC-ANN, to enhance the detection precision for low-frequent attacks and detection stability. The general procedure of FC-ANN approach has the following three stages. In the first stage, a fuzzy clustering technique is used to generate different training subsets. Based on different training sets, different ANNs are trained in the second stage. In the third stage, in order to eliminate the errors of different ANNs, a meta-learner, fuzzy aggregation module, is introduced to learn again and combine the different ANN's results. The whole approach reflects the famous philosophy ''divide and conquer". By fuzzy clustering, the whole training set is divided into subsets which have less number and lower complexity. Thus the ANN can learn each subset more quickly, robustly and precisely, especially for low-frequent attacks, such as U2R and R2L attacks.

A. Related work on IDS

IDS is split into two categories: misuse detection systems and anomaly detection systems [3]. Misuse detection is used to identify intrusions that match known attack scenarios. However, anomaly detection is an attempt to search for malicious behavior that deviates from established nor-mal patterns. In order to detect the intrusion, various approaches have been developed and proposed over the last decade. In the early stage, rule-based expert systems and statistical approaches are two typical ways to detect intrusion. A rule-based expert IDS can detect some well-known intrusions with high detection rate, but it is difficult to detect new intrusions, and its signature database needs to be updated manually and frequently. Statistical-based IDS, employs various statistical methods including principal component analysis, cluster and multivariate analysis, Bayesian analysis, and frequency and simple significance tests. But this type of IDS needs to collect enough data to build a complicated mathematical model, which is impractical in the case of complicated network traffic. To solve the limitations of above methods, a number of data mining techniques have been introduced. Among these techniques, ANN is one of the most used techniques and has been successfully applied to intrusion detection. According to different types of ANN, these techniques can be classified into the following three categories: supervised ANN-based intrusion detection, unsupervised ANN-based intrusion detection, and hybrid ANN-based intrusion detection. Supervised ANN applied to IDS mainly includes multi-layer feed-forward (MLFF) neural networks and recurrent neural net-works. Ryan et al. (1998) and Tan (1995) used MLFF neural networks for anomaly detection based on user behaviors. But in practice the number of training set is very large and the distribution of training set is imbalanced, the MLFF neural networks is easy to reach the local minimum and thus stability is lower. Especially, for low-frequent attacks, the detection precision is very low. Supervised ANN had been shown to have lower detection performance than SVM and MARS.

The second category uses unsupervised ANN to classify input data and separate normal behaviors from abnormal or intrusive ones. Using unsupervised ANN in intrusion detection has many advantages. The main advantage is that unsupervised ANN can improve their analysis of new data without retraining. Just like using supervised learning ANN, the performance of unsupervised ANN is also lower. Especially for low-frequent attacks, unsupervised ANN also gets lower detection precision. The third category is hybrid ANN which combines supervised ANN and unsupervised ANN, or combine ANN with other data mining techniques to detect intrusion. The motivation for using the hybrid ANN is to overcome the limitations of individual ANN. For ANN-based intrusion detection, hybrid ANN has been the trend. But different ways to construct hybrid ANN will highly influence the performance of intrusion detection. Different hybrid ANN models should be properly constructed in order to serve different aims.

Following this stream, we propose a hybrid ANN, called FC-ANN, to solve the two drawbacks of current ANN-based IDS mentioned in Section i.e., lower detection precision for low-frequent attacks and weaker detection stability. FC-ANN approach introduces fuzzy clustering technique into ordinary ANN. By using fuzzy clustering technique, the whole training set can be divided into subsets which have less size and lower complexity. Therefore based on these sub sets, the stability of individual ANN can be improved, the detection precision, specially for low-frequent attacks, can also be enhanced.

B. Security Issues in Cloud Computing

Cloud data confidentiality issue : Confidentiality of data over cloud is one of the glaring security concerns. Encryption of data can be done with the traditional techniques. However, encrypted data can be secured from a malicious user but the privacy of data even from the administrator of data at service provider's end could not be hidden. Searching and indexing on encrypted data remains a point of concern in that case. Above mentioned cloud security issues are a few and dynamicity of cloud architecture are facing new challenges
with rapid implementation of new service paradigm.

Network and host based attacks on remote Server : Host and network intrusion attacks on remote hypervisors are a major security concern, as cloud vendors use virtual

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

machine technology. DOS and DDOS attacks are launched to deny service availability to end users.

Cloud security auditing : Cloud auditing is a difficult task to check compliance of all the security policies by the vendor. Cloud service provider has the control of sensitive user data and processes, so an automated or third party auditing mechanism for data integrity check and forensic analysis is needed. Privacy of data from third party auditor is another concern of cloud security.

Lack of data interoperability standards : It results into cloud user data lock-in state. If a cloud user wants to shift to other service provider due to certain reasons it would not be able to do so, as cloud users data and application may not be compatible with other vendor's data storage format or platform. Security and confidentiality of data would be in the hands of cloud service provider and cloud user would be dependent on a single service provider.

II. PROPOSED MODEL

A. System Architecture

Our proposed multi-threaded NIDS model for distributed cloud environment is based on three modules: capture & queuing module, analysis/ processing module and reporting module. The capture module, receives the in-bound and out-bound (ICMP, TCP, IP, UDP) data packets. The captured data packets are sent to the shared queue for analysis.

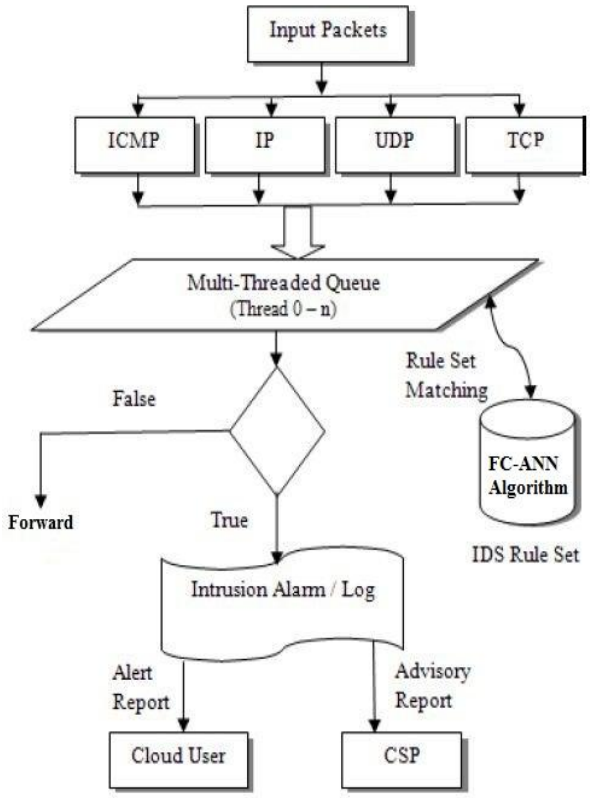

Fig. 2 Flowchart of cloud IDS

As shown in fig. 2 the main process will receive TCP, IP, UDP and ICMP packets and multiple threads would concurrently process and match those packets against pre-defined set of rules. Through an efficient matching and analysis the bad packets would be identified and alerts generated. Reporting module would read the alerts from shared queue and prepares alert reports. The third party monitoring and advisory service having experience and resources would immediately generate a report for cloud users information and sends a comprehensive expert advisory report for cloud service provider.

Cloud computing provides application and storage services on remote servers. The clients do not have to worry about its maintenance and software or hardware up-gradations. Cloud model works on the „concept of virtualization" of resources, where a hypervisor server in cloud data center hosts a number of clients on one physical machine. Deploying HIDS in hypervisor or host machine would allow the administrator to monitor the hypervisor and virtual machines on that hypervisor. But with the rapid flow of high volume of data as in cloud model, there would be issues of performance like overloading of VM hosting IDS and dropping of data packets. Also if host is compromised by an offending attack the HIDS employed on that host would be neutralized. In such a scenario, a network based IDS would be more suitable for deployment in cloud like infrastructure.

The system architecture of proposed model is shown below in Fig. 3 where NIDS would be placed outside the VM servers on bottle neck of network points such as switch, router or gateway for network traffic monitoring to have a global view of the system.

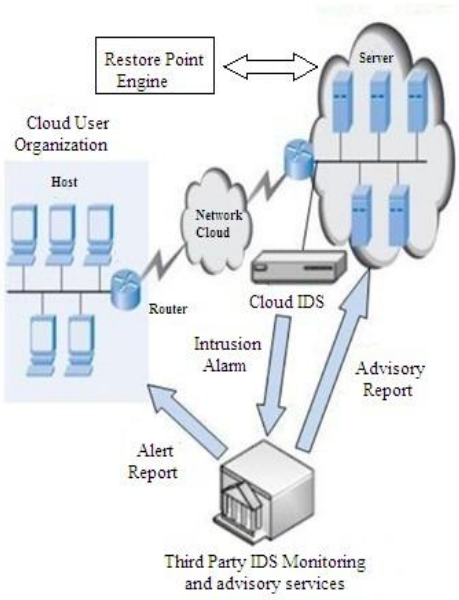

Fig. 3 Architecture of Cloud IDS Model

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

Such NIDS would still be facing the issue of large amount of data through network access rate in cloud environment. To handle a large number of data packets flow in such an environment a multi-threaded IDS approach has been proposed in this paper. The multi-threaded IDS would be able to process large amount of data and could reduce the packet loss. After an efficient processing the proposed IDS would pass the monitored alerts to a third party monitoring service, who would in turn directly inform the cloud user about their system under attack. The third party monitoring service would also provide expert advice to cloud service provider for misconfigurations and intrusion loop holes in the system [4]. The cloud user accesses its data on remote servers at service provider's site over the cloud network. User requests and actions are monitored and logged through a multi-threaded NIDS. The alert logs are readily communicated to cloud user with an expert advice for cloud service provider.

B. Framework of FC-ANN

FC-ANN firstly divides the training data into several subsets using fuzzy clustering technique. Subsequently, it trains the different ANN using different subsets. Then it determines membership grades of these subsets and combines them via a new ANN to get final results. The whole framework of FC-ANN is illustrated in Fig. 1. As typical machine learning framework, FC-ANN incorporates both the training phase and testing phase. The training phase includes the following three major stages :

Stage I : For an arbitrary dataset DS, it is firstly divided into training set TR and testing set TS. Then the different training subsets TR_1, TR_2, .. TR_k are created from TR with fuzzy clustering module.

Stage II : For each training subset Tr_i (i=1, 2, ...k) , the ANN model, ANN_i ,(i=1,2,....k) is training by the specific learning algorithm to formulate k different base ANN models.

Stage III : In order to reduce the error for every ANN_i, we simulate the ANN_i using the whole training set TR and get the results. Then we use the membership grades, which were generated by fuzzy clustering module, to combine the results. Subsequently, we train another new ANN using the combined results.

In the testing phase, we directly input the testing set data into the k different ANN_i and get outputs. Based on these outputs, the final results can then be achieved by the last fuzzy aggregation module. The three stages of FC-ANN framework raise three important is-sues: (1) how to create k different training subsets from the original training dataset TR; (2) how to create different base model ANN_i with different training subsets; (3) how to aggregate the different results produced by different base model ANN_i.

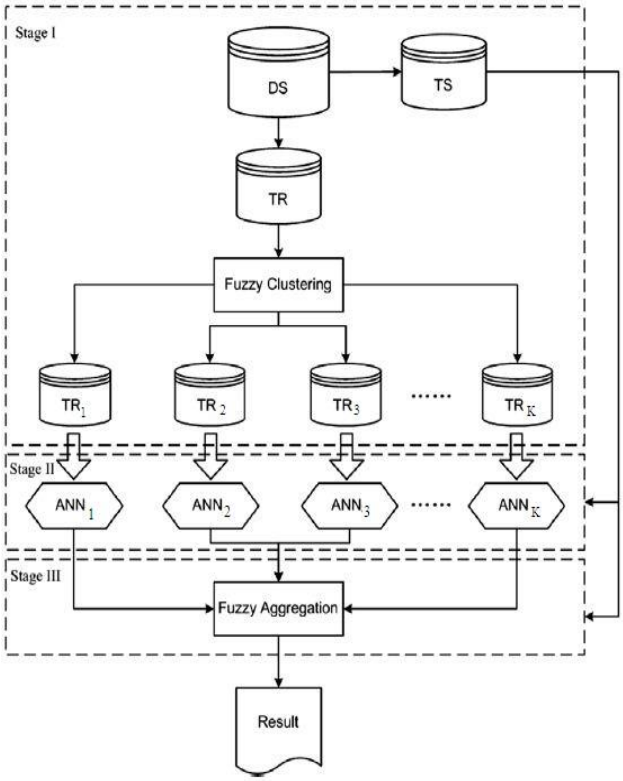

Fig. 4 Framework of FC-ANN

C. Restore Point

When network outages occur however, caused by either hardware failure or human error, organizations need to be able to react quickly to bring services back online. It is often at this point that the organization learns they do not have an accurate or current backup of the firewall, switch or router configuration for example which can result in hours or sometimes days being spent rebuilding a device which can lead to costly downtime. Hence we proposed the concept of Restore point with IDS which can automate the backup of network devices like data, and recover or rollback to a previous state without fuss, within minutes. It will reduce any organization operating costs by simplifying administration and automating often complex and manual tasks across a multi-vendor network estate. A single Restore point appliance can centrally manage the disaster recovery and compliance needs of large enterprises or service providers with thousands of devices. Our automated compliance monitoring also enables any organizations to reduce the time and skills required to analyze each of their network vendor configurations because restore point provides a single and centralized view which can be especially useful during external audits. The creation and utilization of network restore points is provided by a schema associated with a structured data set that can be conveniently backed up by periodically taking snapshots of the structured data to establish a series of restore points that can be used in the event that the primary structured data becomes lost or corrupted. In general, the snapshots are only taken after

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 4, April 2013

the structured data set has undergone a change in content, although they may be taken at other times as well. The snapshot may be taken by a network restore point engine that can be accessed by a user over the Internet. In addition, the snapshot may be stored on an Internet-based storage medium.

If IDS fails to detect the new attacks which can result in system down, in such cases requirements regarding configuration management, backup, change detection and security is needed which is maintain by the proposed restore point engine. A network restoration arrangement for backing up at least underlying data is associated with a structured data set, comprising: (1) an Internet-based primary storage medium configured to store a plurality of structured data sets each associated with a user. (2) a network restore point engine for establishing restore points of at least underlying data associated with at least one user-defined structured data set that is eligible for backup. (3) a secondary storage medium on which the restore points are stored by the network restore point engine. The network restoration arrangement is the Internet-based primary storage medium which is further configured to store change data that reflects an accumulation of changes that have been made to each of the plurality of structured data sets since establishment of a previous restore point.

III. CONCLUSIONS

Cloud computing is a "network of networks" over the internet, therefore chances of intrusion is more with the erudition of intruders attacks. Prevention of security breaches completely using the existing security technologies is unrealistic. As a result, intrusion detection is an important component in network security. IDS offers the potential advantages of reducing the manpower needed in monitoring, increasing detection efficiency, providing data that would otherwise not be available, helping the information security community learn about new vulnerabilities and providing legal evidence. In this paper, we propose a new intrusion detection approach, called FC-ANN, based on ANN and fuzzy clustering. Through fuzzy clustering technique, the heterogeneous training set is divided to several homogenous subsets. Thus complexity of each sub training set is reduced and consequently the detection performance is increased.

REFERENCES

[1] S. Sontakke, " Intrusion Detection System for Cloud Computing," International Journal of Scientific & Technology Research Volume 1, Issue 4 (page no. 67-71), May 2012.

[2] L. Huang, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering," School of Management, Fudan University, Shanghai 200433, PR China, Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong.

[3] A. Haeberlen," An Efficient Intrusion Detection Model Based on Fast Inductive Learning," Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007. Tavel, P. 2007.

[4] S. Mukkamala, "Designing Intrusion Detection Systems: Architectures, Challenges and Perspectives," Department of Computer Science, Oklahoma State University, USA.

[5] M. Hussain, "Distributed cloud intrusion detection model," International Journal of Advanced Science and Technology Vol. 34 (page no. 71-82), September, 2011..

[6] J. Anderson, "An introduction to neural networks," Cambridge, MIT press,1995.

[7] A. Kumar, " Security Patterns for Intrusion detection Systems," 1st LACCEI International Symposium on software Architecture and Patterns (LACCEI-ISAP-MiniPLoP'2012), July 23-27, 2012.