IDS FOR CLOUD COMPUTING

Chandrika Mohan
Final Year Student, Department of CS
G.L Bajaj Institute of Technology and Management,
Greater Noida, U.P, INDIA

Abstract - **In this paper I am going to introduce the concept of Cloud Computing which efficiently looks ups and equips the user request. Normally many users login at the same moment due to which the Server might not be able to provide equal performance to all the systems thus by Load Balancing Techniques, and by adding Grid Computing concepts and by clubbing other systems to the server, high performance can be insured. The user would then send the inquiry to the Main Cloud Server, where all the other Cloud Servers (proxy servers) are connected using the concept of clustering. Clustering is done for databases which allow proxy servers to maintain redundancy of data.**

After receiving the request send by the client, main Cloud server would check the exact proxy server to whom the request could be forwarded by a load balancing technique which can undertake the work more deligerently. In the cloud computing, loads of Hackers may attempt to peneterate the data, so a full proof Intrusion Detection system would be of help.

The Intrusion Detection system is commonly of two types Anomaly based detection and Signature based detection.

Anomaly based detection looks for any deviation from the past patterns and accordingly raises an alarm. If the system finds that an Intrusion attack has happened, it would in turn alarm other Cloud Server by raising the alarm, so that data entry is restricted.

Usually an Intrusion detection system raises two types of alarms one is false positive and the other false negative. False positive alarms are raised when system detects intrusion activity when it is accessed by the authenticated user. False negative alarms are raised when an intruder bypasses the security. False negatives are harmful for the system security.

Keywords- **IDS, Cloud Security, Authentication, Authorization, Distributed System, Anomaly Based Detection, PAYL**

I. INTRODUCTION

A. SYSTEM ANALYSIS

In earlier days, the main server could not efficiently handle multiple user requests. It progresses to traffic and congestion. In order to tranquish this problem we can implement the concept called Cloud Computing.
Along with Cloud Computing clubing the concept of clustering is the main idea of this paper. A cluster is composed of one single server along with multiple proxy servers which can be utilized by multiple clients. The aim of the Proxy server is to store data on local disks and then accordingly retains read and write operation specified by a server. The server stores index for all files stored in different proxies. When a client wishes to download some data, it will first send a request to the Server and accordingly the Server would redirect the request to a corresponding proxy which has the required data and more the data will be sent to the client. In this way the data request can be usefully serviced in a prompt manner

Anomaly Based Detection - In this structure we aim to detect computer invasion and misuse by supervising system activity and grade them as either normal or anomalous. This comprises of using a strict mathematical model and identify any alteration from this as an invasion.

The Security execution in this project can be achieved by two phases, namely Behavior Analysis and Knowledge Analysis.

A.1 BEHAVIOR ANALYSIS : By this method, we can recognize anticipated behavior (permissible use) or a grave behavior deviation. The network must be correctly tutored to efficiently detect intrusions. This would be implemented through threshold values of different parameters such as time stamp, frequency to access account, session time, database access time and credit limit.

A.2 KNOWLEDGE ANALYSIS : An expert system, can flag a malicious behavior with a rule. The benefit of using this kind of intrusion detection is that new rules can be added without changing existing. Audit data from both the log system and the communication system is used to essasse the knowledge based system.

II. FEASIBILITY STUDY

A. Operational Feasibility

To develop the project I am using Java as the base platform due to its following features like interoperability, security, portablity and dynamic. Hence it is operational feasibile. It can be enacted in all kinds of environments. The proposed system tends to resolve data pipeline effect which is most frequently found in networks**.**

B. Technical Feasibility

Implementation in LAN environment provides high technical support to the users despite of the operating system on which they are working. The cache data, cache path and hybrid cache are the three main modality which makes the optimal cache in an oderly way even when there is clogging over the network. Java assists both the developer and user by providing connectivity to database and front end.

For the end user perspective who just need to execute the already built code, JRE is enough, they may not require the complete JDK tool kit support while for the developer's perspective the complete JDK tool kit support is required.

The JVM can also be enforced in hardware. Java byte codes visualizes write once, run anywhere possible. Java program can be compiled into byte codes on any platform that has Java compiler.

TECHNOLOGIES USED

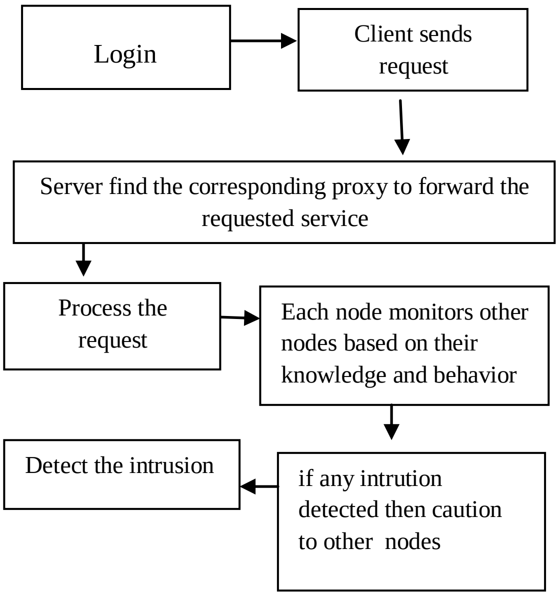
- JAVA JDK 1.8
- Apache Tomcat
- MS SQL Server

III. PROJECT SPECIFICATION

A. PROBLEM DEFINITION

Providing security in the distributed system requires more than user authentication with confidentiality and passwords or digital certificates in data transmission. The Cloud Computing Intrusion Detection System integrates knowledge based and behavior analysis to detect invasions.

B. PROJECT FLOWS


Login
Client sends request
Server find the corresponding proxy to forward the requested service
Process the request
Each node monitors other nodes based on their knowledge and behavior
if any intrution detected then caution to other nodes
Detect the intrusion

C. MODULES

IV.C.1. Client
IV.C.2. Proxy server
IV.C.3 IDS service
IV.C.3.1 Analyzer system

IV.C.3.2. Alert system

IV.C.4. Storage service

IV.C.4.1. Knowledge based service

IV.C.4.2. Behavior based

IV.C.5. Event auditor

D. MODULE DESCRIPTION

D.1. Analyzer System

It uses a history database and some statistics of past patterns to look for any kind of deviation or simply determines the gap between a typical user behavior and the suspected behavior and communes this to the IDS service which further instruct other proxy servers whether an intruder has breached the security or not. The rule analyzer collects audit packages and calculates whether any rule in the database is been broken. It returns the score to the IDS service which further communicates it to other proxy servers. If the probability is high it indicates that an intruder or malicious user has entered the system and alert the other nodes by simply raising the alarm that false positive has occurred. By this the nodes get aware of

intrusive presence and take preventive measures to deal with the intruder.
The authentication technique chosen should not compromise with the privacy and security issues at customer or valid user end, performance and reliability of the system and portability and interoperability should also be taken care of.

D.2. Alert System

Head network administrators blindly trust firewalls to retain a secure environment. However, they still miss some aspects of network security, such as event association and secure the audit trails. This is why intrusion-detection systems come to use.

In this approach a set of distributed systems that spot and alert security administrators about active attacks in real time is used.Whereas, signature analysis generally detect too basic intrusions.

The major objective of IDS is to tally alerts and generate global alerts and discard the false alerts. The context of the work is an intrusion detection environment with several IDS that foster alerts when some unusual event occur. To reach correlation objective, we specify the following three functions:

Alert management function- Alerts are generated, stored and managed in a relational database.

Alert clustering function- Alerts maps the same event of attack are recognized and clubbed to the same cluster.

Alert merging function - Global alert is then initiated for each cluster determined by the clustering function.

D.3. Storage Service

D.3.1. Knowledge-based Service

In the Knowledge-based anomaly based intrusion detection system approach try to seize the claimed behavior from the available system data. Knowledge-based detection techniques apply the knowledge accumulated about system vulnerabilities specific attacks. We audit data from both the log system and the communication system to assese the knowledge based system and create a series of rules in order to illustrate security policies that the IDS should monitor.

It aims to tackle the following situations of password cracking and access violation, Interceptions and Trojan horses which are most frequently associated with TCP/IP stealing and interceptions which often employ additional mechanisms to compromise the operation of the attacked system man in the middle attacks. The accuracy of the knowledge-based detection systems is considered good.

D.3.2. Behavior Based Service

There are two behavior based services, namely User behavior and Node behavior.

User Behavior- For user behavior we have to analyze the user’s behavior. With this method, we accept the anticipated behavior and the behavior deviation. Popular threat detection within network behavior anomaly detection are:

IP spoofing – It is used to gain unauthorized access to computer, whereby the intruder sends messages to a computer with an Internet Protocol address indicating that the message is arriving from a trusted host. While a source address is sent along with every IP packet, it is not genuinely used for routing. The malicious user is never going to see the response packets, but the machine do, but throws them away because they don't compare any requests we have sent.

TCP/UDP Fanout- This happens when single source IP is connected with many ports on a destination IP address. Anything which performs a vulnerability assessment or even a port scan can cause this to trigger. This could include endpoint integrity checks.

DNS poisoning - DNS servers repetedly resolve DNS names. And thereby the DNS server that equips a client request will behave as if it is a client to the next server in the recurrent chain. A malicious user can direct a message demand to the DNS server and a reply to the server supposititious to be from the next server.

IP Fanout -This happens when one IP is directing traffic to a large number of other IPs within a specified time window.

Duplicate IP- This happens when two network packets are analyzed that have different source MAC addresses, but a unique source IP address. We can report this for innocuous reasons when traffic is sampled on both sides of a router.

Node Behavior - It measures the basic standard or level of CPU utilization and activities concerned with file and disk. It can thus detect the anomalies without having to understand the subtle cause behind the anomalies.

Statistical Based Intrusion Detection System(SBID)
The behavior-based analysis of traffic patterns and malicious activities can be lethal.It requires no prior knowledge of an attack, as is generally the case with rule-based Intrusion Detection and Prevention System.

Whereas, the SBID systems can detect these attacks. Warnings are based on actual usage patterns, statistical systems can habituate with behaviors and compose their own rule usage behaviour. Statistical-Based systems depend on the Baye's Theorem that identifies anomalous packets on the network. Bayes' theorem is a theorem that helps in computing conditional or posterior probability of each of a set of possible parameters for a given observed outcome from knowledge of the prior probability of each cause and the class conditional probability of the outcome of each case.

IV. IMPLEMENTATION

The implementation of Anomaly Based IDS is by Statistical-based systems. Statistics-based systems proceeds on a different path for detecting the presence of intruder. The concept of the Statistics-based system is simple it computes the general network activity and then all the traffic that falls outside the scope of general activity is marked anomalous. SBID systems attempt to learn about network traffic patterns on a particular network. This process of traffic analysis continues till the SBID system is active. By analyzing network traffic and processing the data with statistical algorithms, SBID systems search for anomalies in the established normal network traffic patterns. All packets are assigned a anomaly score on the basis of degree of irregularity for the specific event. If the anomaly score is more than a threshold which is set by the user, the IDS will initiate an alert. When an unknown or rare (anomalous) event occurs, the SBID analyze it and apparently generates an alert. The gains of the statistical-based approach are threefold. We have information regarding previously hidden attacks, it doesn't need frequent signature updates, and a method to find port scans which stretch substancial time frames.

V. ENHANCEMENTS

I would further work on integrating Payload-based Network Detection is the system. Firstly, compute during a training phase a profile byte standard deviation and frequency distribution of the application payload flowing to a single host and port. We then use squared Mahalanobis distance during the identification phase to measure alikeness of new data against the pre-computed profile. The Anomaly Payload has the capability of providing 100% correctness along with 0.1% false positive rate for port no. 80 traffic.

Simplified Mahalanobis Distance- Mahalanobis distance is a commonly used distance metric which compares two statistical distributions. It is a very impactful method which is used to calculate the alikeness between new, unrecognized payload sample and the previously measured model. I figured out the difference between the byte distributions of the newly measured payload against the figuration from the model measured. The greater the value of the distance score, more likely the payload is deviating. Commonly used equation for Mahalanobis distance is:

$$D^2(x,y) = (x-y)^T C^{-1}(x-y)$$

Where x and y are two vectors. x is the feature vector of new measurement, and y is the intermediate feature vector which is evaluated from the training examples.

C^{-1} is inverse covariance matrix and $C_{ij} = \mathrm{Cov}(y_i, y_j)$, are the i^{th} and j^{th} elements of the training vector. It not only takes into account the average value, but also its covariance and variance of the variables to be evaluated.

VII. REFERENCES

[1] Vaid, C. and Verma H.K. : Anomaly Based IDS implementation in Cloud environment using BOAT algorithm IEEE 2014.

[2] Junho Hong, Chen-Ching Liu and Govindarasu M. -Integrated Anomaly Detection Cyber Security of the Substations IEEE 2014.

[3] Ke Wang, Salvatore J. Stolfo, Computer Science Department, Columbia University - Anomalous Payload-based Network Intrusion Detection

[4] S. Axelsson, Research in Intrusion-Detection Systems: A Survey, tech. report TR-98-17, Dept. Computer Eng., Chalmers Univ. of Technology, 1999.

[5] A. Schulter et al., ―Intrusion Detection for Computational Grids,‖ Proc. 2nd Int'l Conf. New Technologies, Mobility, and Security, IEEE Press, 2008, pp. 1–5..

[6] P.F. da Silva and C.B. Westphall, ―Improvements in the Model for Interoperability of Intrusion Detection Responses Compatible with the IDWG Model,‖ Int'l J. Network Management, vol. 17, no. 4, 2007, pp. 287–294

[7] P. Luo, F. Lin, Y. Xiong, Y. Zhao, and Z. Shi, TowardsP. Luo, F. Lin, Y. Xiong, Y. Zhao, and Z. Shi, Towards combining web classification and web information extraction: a case study, In KDD'09:Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009.