INTRUSION DETECTION SYSTEM FOR CLOUD ENVIRONMENTS BASED SUPERVISED MACHINE LEARNING ALGORITHMS

Mr. M.Krishnamoorthy
Asst Professor
Panimalar Engineering College
Chennai,TamilNadu

Hritish S
Computer Science and Engineering
PanimalarEngineering College,
Chennai,TamilNadu

Gokulakrishnan P
Computer Science and Engineering
PanimalarEngineering College,
Chennai,TamilNadu

Jebinesh E H
Computer Science and Engineering College
Panimalar Engineering College
Chennai,TamilNadu

***Abstract*—There are many risks of network attacks under the cloud environment. Internet Security may be a vital issue and thus , the intrusion detection is one major research problem for business and private networks to resist external attacks. A Network Intrusion Detection System (NIDS) may be a software application that monitors the network or system activities for malicious activities and unauthorized access to devices. The goal of designing a NIDS is to guard data's confidentiality and integrity.Our paper focuses on these issues with the assistance of supervised Machine Learning algorithm to find out the patterns of the attacks classfication , NSL-KDD dataset has been used.**

Keywords—Intrusion, attack, confidentiality, network, unauthorized

I. INTRODUCTION

The appealing features of Cloud computing still fuel its integration in many sectors including industry, governments, education, entertainment, to call few Cloud computing aims to supply convenient, on-demand, network access to a shared pool of configurable computing resources, which may be rapidly provisioned and released with minimal management effort or service provider Interactions.The pay-as-you-go and therefore the on-demand elastic operation Cloud characteristics are changing the enterprise computing model, shifting on-premises infrastructures to office premises data centers, accessed over the web and managed by cloud hosting providers. However, many security issues arise with the transition to the present computing paradigm including intrusions detection.Regardless the important evolution of the knowledge security technologies in recent years, intrusions and attacks still defeat existing intrusion detection systems in Cloud environments . Attackers developed new sophisticated techniques ready to bring down a whole Cloud platform or maybe many within minutes. New records are reached annually by attackers. Intrusion and attack tools became more sophisticated, challenging existing Cloud IDS by large volumes of network traffic data, dynamic and sophisticated behaviors and new sorts of attacks.
Jebinesh E H
Computer Science and Engineering College

It is clear that an IDS for Cloud should analyze large volumes of network traffic data, detect efficient the new attack behaviors and reach high accuracy with low false. However preprocessing, analyzing and detecting intrusions in Cloud environments using traditional techniques became very costly in terms of computation, time and budgets.

II. OBJECTIVE

The Main Objective of the paper is designing a NIDS is to protect data confidentiality and integrity. Our paper focuses on these issues with the help of Machine Learning.

III. TYPE OF ATTACK

A. *Eavesdropping*

In general, the bulk of network communications occurin an unsecured or "clear text" format, which allows an attacker who has gained access to data paths in your network to "listen in" or interpret (read) the traffic..

B. *Data Modification*

After an attacker has read your data, subsequent logical step is to change it. An attacker can modify the info within the packet without the knowledge of the sender or receiver. albeit you are doing not require confidentiality for all communications, you are doing not want any of your messages to be modified in transit. for instance , if you're exchanging purchase requisitions, you are doing not want the things , amounts, or billing information
to be modified.

C. *Identity Spoofing (IP Address Spoofing)*

Most networks and operating systems use the IP address of a computer to spot a legitimate entity. In certain cases, it's possible for an IP address to be falsely assumed— identity spoofing

D. Sniffer Attack

A sniffer is an application or device which will read, monitor, and capture network data exchanges and skim network packets. If the packets aren't encrypted, a sniffer provides a full view of the info inside the packet.

E. Application-Layer Attack

An application-layer attack targets application servers by deliberately causing a fault in during a server's OS or applications.

F. Password-Based Attacks

A common denominator of most OS and network security plans is password-based access control. This suggests your access rights to a computer and network resources are determined by who you're, that is, your user name and your password.

G. Man-in-the-Middle Attack

As the name indicates, a man-in-the-middle attack occurs when someone between you and therefore the person with whom you're communicating is actively monitoring, capturing, and controlling your communication transparently.

IV. THE EXISTING SYSTEM

- A false positive may be a situation where something abnormal (as defined by the IDS) happens, but it's not an intrusion.
- Users will quit monitoring IDS because of noise.
- No Confidentiality. Not real time network data implemented.

V. EXISTING DISADVANTAGES

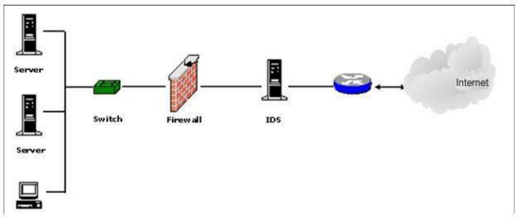
- No Data confidentiality
- No Data integrity
- Less prediction accuracy.
- Not a real time analysis.
- Existing system not used for a long dataset.
- The prediction of identity malicious activity is not very accurate.

VI. MACHINE LEARNING IN IDS

The pattern of the normal activities and malicious activities can be learned and distinguished.Machine learning can result in higher detection rates, lower false alarm rates.

On dynamic data, the model can be updated and maintained.

VII. BLOCK DIAGRAM OF IDS



VIII. PROPOSED SYSTEM

- This research focuses on solving the problems in intrusion detection communities which will help the administrator to form pre-processing, classification, labeling of knowledge and to mitigate the result of Distributed Denial of Service Attacks.
- Since, the network administrator feels difficult to pre-process the info . thanks to the overwhelming growth of attacks which makes the task hard, attacks are often identified only after it happens. to beat this example , frequent updating of profiles is required .
- Reduced workload of administrators increases the detection of attacks.

Advantages

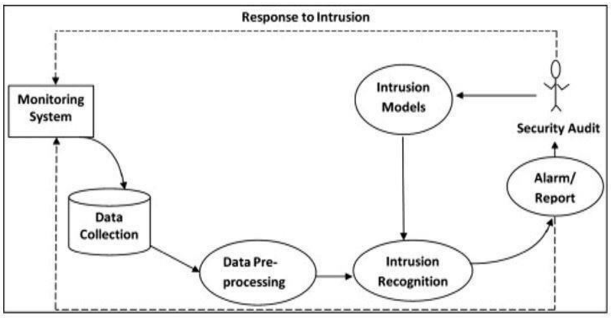
- When designing a IDS, the mission is to guard the data's

Confidentiality – read

Integrity – read/write

- Real time data analytics.
- High accuracy.

IX. ARCHITECTURE DIAGRAM



X. MODULES

- Data collection.
- Data preprocessing.
- Machine learning algorithm
- Training and testing

A.MODULES DESCRIPTION

1. DATA COLLECTION

Real time data collected from Twitter ,kaggle, UCI , Data.gov,NSL-KDD dataset

Collection of knowledge is one among most vital the keythe foremostthe main and most important tasks of any machine learning papers. Because the input we feed to the algorithms is data. So, the algorithms efficiency and accuracy depends upon the correctness and quality of knowledge collected. therefore the data are going to be the output.

2. Data preprocessing

As you'll see, the dataset contains nominal values also and to coach a model we'd like all numerical values.

Here is that the transformation table that we used. Dataset is extremely large and there's an outsized variation between values, Data Normalization is additionally required for better performance..

3. Machine learning algorithm

The next step is algorithms are applied to data and results are noted and observed. the choice tree and random forest algorithm applied for accuracy at each stage.

4. Training and Testing

Finally after processing of knowledge and training the very next task is clearly testing. this is often where performance of the algorithm, quality of knowledge , and required output all appears out. From the large data set collected 80 percent of the info is employed for training and 20 percent of the info is reserved for testing. Training as discussed before is that the process of creating the machine to find out and giving it the potential to form further predictions supported the training it took.

XI. CONCLUSION

- Algorithms based on Machine Learning were implemented successfully showing different accuracies.
- The NSL-KDD dataset was preprocessed using mean normalization method.
- Linear regression, surprisingly, proved to be very effective in detecting network attacks with a high accuracy.
- Neural Networks were implemented with one hidden layer one time and two hidden layers another time. With two hidden layers, it proved to be the best among all the approaches above.
- But there is always a trade-off between the accuracy and time an algorithm takes. Neural Network took the most time to get trained while K-Means Clustering took the lowest amount of time.

XII. REFERENCE

- [1]A Dynamical Growing Self-Organizing Tree (DGSOT) for Hierarchical Clustering Gene Expression Profiles," Feng Luo, Latifur Khan , Farokh Bastani, I-Ling Yen and J. Zhou, the Bioinformatics Journal, Oxford University Press, UK, 20 16, 2605-2617.
- [2]"Automatic Image Annotation and Retrieval using Weighted Feature Selection" Lei Wang and Latifur Khan to appear in a special issue in Multimedia Tools and Applications, Kluwer Publisher.
- [3]"Hierarchical Clustering for Complex Data" Latifur Khan and Feng Luo, to appear in International Journal on Artificial Intelligence Tools, World Scientific publishers.
- [4]"A New Intrusion Detection System using Support Vector Machines and Hierarchical Clustering" Latifur Khan, Mamoun Awad, and Bhavani Thuraisingham, to appear in VLDB Journal: The International Journal on Very Large Databases, ACM/Springer-Verlag Publishing.
- [5]R. Lippman J. Haines, D. Fried., J. Korba, and K. Das, "The 1999 DARPA off-line intrusion detection evaluation" , Computer Networks, 34, pp. 579-595, 2000.