ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse26932020.pdf
https://doi.org/10.30534/ijatcse/2020/26932020

A Virtual Machine Introspection in Cloud Computing for Intrusion Detection

Nida Kousar G[1], Dr. Dayanand Lal N[2], Chaithanya B N[3], Geetha K[4], Manikanta K B[5]
[1-5]Assistant Professor, Department of CSE, GITAM School of Technology, Bengaluru, India
[1]itzme.nida@gmail.com, [2]dayanandlal@gmail.com, [3] chaithanya.bn@gmail.com, [4]geethak382@gmail.com,
[5]11manikantareddy@gmail.com

ABSTRACT

Cloud security is of paramount importance in the new era of computing. Advanced malware can hide their behavior on detection of the presence of a security tool at a tenant virtual machine (TVM).When the client and server exchange messages among each other, there is an activity that can be observed and tracked in detail of the activities that occur in a network that shows the, login and logout durations, the user's behavior etc. There are several types of attacks occurring from the internet. VM Guard applies the software breakpoint injection technique by storing the activities performed by the user on cloud-based ecommerce application and then habitat file is generated. From the habitat file Text Frequency and Inverse Document Frequency of the user actions is performed and then by applying random forest algorithm to classify the users into intruders and non-intruders.

Key words: Virtualization; Security; Intrusion detection, Random Forest.

1. INTRODUCTION

Cloud computing bears us a route through which we can gain admittance to the bundles as utilities, over the internet. It encourages us to make, arrange, and redo programs online. The period Cloud alludes to a network or net. In different expressions, we can say that Cloud is regarded as a blessing in faraway regions. Cloud offer administrations through network. It is a web grounded thoroughly registering wherein the majority of the mutual sources, programming project and realities are provided to the PC frameworks and contraptions on interest. Clients can get right of entry to the facts from everywhere and every time. Cloud computing is an exchange in an angle where figuring is progressed a long way from desktops or even the discrete challenge application server to a cloud desktop. Cloud computing is the fashion of processing wherein noticeably scaled IT-related competencies are given as an administration over the internet to exclusive out of doors customers and are billed by way of usage. Many distributed computing providers have sprung up and there's an excellent development in the use of this management. cloud computing is discovering use in special zones like web hosting, parallel clump coping with, illustrations rendering money related demonstrating, internet creeping, and a few more.

2. RELATED WORK

According to Preeti Mishra [1] et.al describe the security-related topics for the network administrator's arranged security manager or expert to gather the specialized devices expected to fabricate, keep up, dissect, and gain from a honey net inside their association.

Josenilson [2] et.al outlined the security and privacy issues of virtualization. In conventional surroundings along with bodily servers linked by using a tangible transfer, businesses targeted intendancy data and site visitors that proceed between the server's transfers. Regrettably, the statics control isn't usually supplied from an implicit button. Fundamentally, the computerized exchange has joined from the solid button utilizing the real NIC that connects to implicit instruments. The subsequent loss of oversight of the traffic streams between and some of the implicit gadgets at the equivalent solid stage impacts security and execution.

Swati Nirwal [3] et.al explained that System traffic can be viewed as a wealth information stream. As a result of this reason, the information-digging approach is particular for mining stream information. There are two noteworthy issues identified with stream information arrangement. To start with, we practically cannot store and use previous data for training because it requires countless storage and running time. Second, the idea in the data concept is explained For instance, in the perspective on botnets, the bot ace, for the most part, refreshes the bot programming much of the time, the bot programming is updated. Kai Hwang [4] et.al, tries to provide security solutions for the different cloud services and vendors with virtual cloud clusters.

Bharath Reddy [5] et.al has explained the denial of services attack. This is one of the predominant attacks, it allows the intruders to access the network services and it prevents the valid users to get entry into the services. To overcome the slippage of Dos attack it became vital to design an

Interruption identification framework. An interruption identification framework (IDS) is a programming program that operates as a community safety mechanism to protect the computer network device from attacks. With an increasing range of records being transmitted step by step from one network to any other, the IDS pick out the intrusions in such huge datasets effectively. Records mining is an efficient tool implemented to define the intrusion detection gadget and save you the large network data from the intruders. Outliers are patterns in information that do not fit a properly-defined belief of normal conduct.

Farzad Sabahi [6] et.al mentioned that inside the ocean of online assaults which might be hampering IT safety the Dos has the most catastrophic impact. This attack had an immense strain on security professionals to locate powerfully Defense solutions. With a variety of tools and codes this attack may be executed. As we dint had any option to Dos attack it wisely dominated the net for a long time. Therefore, it becomes imperative to discover the answer to this assault. Those ongoing assaults are estimated and dissected the use of community traffic video display units. In addition to that, this challenge additionally offers diverse protection techniques. The recognition and moderation instruments planned ideal here are amazing for little network topologies and might be drawn out to similar to huge spaces.

.

Irfan Gul [7] et.al narrates that carrier choice for automated dynamic service composition with patron's necessities orientated carrier selection becomes greater excessive. The prevailing planning and choice algorithms are designed for provider discovery. Similarly, to our understanding, there are only some works that include giving up-consumer requirements into service composition.

Snehal G [8] et.al has explained the kinds of interruption location methods. The interruption recognition is as follows: Trademark Based discernment, Peculiarity Based Discernment Technique, Crossbreed Detection Procedure.

Uttam Kumar [9] et.al describes the summary of intervention cultivation systems. Intervention Detection structures may be applied in the cloud to stumble on diverse assaults. The triumph of this system relies on the techniques used for the infraction detection like designation primarily based overstepping Detection, Anomaly-based Incursion Detection, and artificial Intelligence-based obstruction Detection.
Vaishali B [10] et.al describes the various cloud attacks. The different cloud attacks are as follows: Swaddle Assault, Exploit Inoculation Attack, Facts stealing hassle. P. Mishra [12] listed various intrusion detection techniques in cloud environment. [11,13] used a different approach for intrusion detection by implementing support vector machine technique for detection but the results were not promising.

3. THE PROPOSED METHODOLOGY

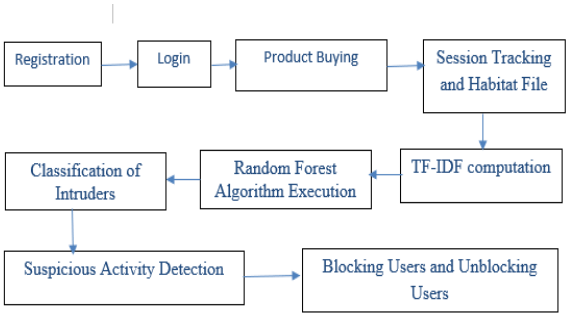

Figure 1: VMGuard architecture

The proposed Vmguard architecture is as mention in the Figure 1. It includes user registration and performing intended activities. These user activities are tracked and then a file known as the habitat file is created based on the activities of the users. The habitat file is used to compute the Text Frequency – Inverse Document Frequency (TF-IDF) is done for each of the unique actions. The weight computation is done and then a random forest algorithm is executed based on the training data and then the prediction is done whether the user is an intruder and non-intruder. Once the user is detected as intruder then the user who will be an intruder can be blocked. The VM guard architecture then brings in an additional innovation of notifying the users of all the suspicious activities across the sessions.

3.1 Registration

This Module is responsible for allowing any external customer to perform the registration by proving the details. Once the registered user can not register again.

3.2 Login

The login Module is responsible for allowing the user to access the user with valid credentials and deny access for users with invalid credentials. The Users are of two kinds one is Admin and the other is Customer. If it is Admin then he/she can see the habitat file for each session of the users using the application which has the session tracking, Find Suspicious pattern using LCS and Detect Dos Attack. If it is a customer then he/she can purchase a product and then get the recommendations.

3.3 Product buying

Product Buying is responsible for purchasing the products by providing valid IPIN and Account No. If the credentials are valid and also user has sufficient balance then the product is purchased and then two important information are tracked namely Order Information and Order Details. The Order information can be described as below.

3.4 Session Tracking and Habitat File

Whenever the user clicks on the link or clicks on the button or user navigates from one page to another page each time independent request is made and tracked based on the user id and session id. The habitat file is a set of records that are set of actions performed by the user in each session.

3.5 Suspicious Activity using Least Common Sub- Square (LCS) Algorithm

The LCS algorithm shown in the Figure 2 is responsible for taking a set of patterns and finds the LCS for each of the patterns. The pattern refers to one habitat of the user for the specific session. If the LCS is new as compared to previous activities then the pattern is regarded as an intrusion.

Algorithm:

The steps of the LCS algorithm is as follows:

Step 1: Consider the sequences namely s1 and s2

Step 2: The length of S1 and length of S2 is computed

Step 3: Find the maximum length of S1 and S2

Step 4: Construct a matrix initial with zero's one 1st row and 1st column

Step 5: If the value of the sequence alphabet is not there then maximum on top and left is taken

Step 6: If the value matches then the diagonal value is increment by a value of1.

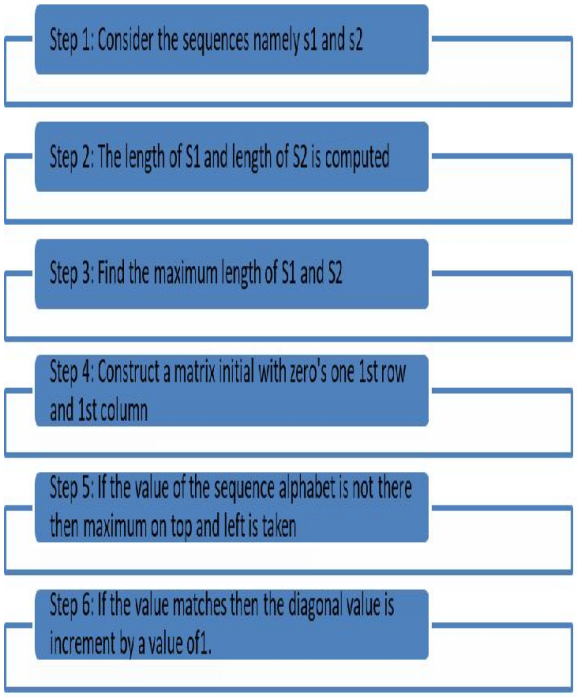

Figure 2: LCS Algorithm

Example: Given two sequences, find the length of the longest subsequence present in both of them. A subsequence is a sequence that appears in the same relative order, but not necessarily contiguous. For example, “abc”, “abg”, “bdf”, “aeg”, ‘”acefg”,.etc are subsequences of “abcdefg”. So a string of Length n has 2^n different possible subsequences.

- The pattern refers to one habitat of the user for a specific session.
- If the LCS is new as compared **to previous activities** then the pattern is **regarded as an intrusion**.

The suspicious activity will be classified into low, medium, and high alerts by using the following steps:

Step 1: The Average pattern count is defined as the number of times LCS value is repeated over the Total number of sessions.

Step 2: The frequency of pattern is considered by the number of times LCS is repeated.

Step 3: The number of a unique pattern is considered by the unique LCS value across the user

Step 4: Consider the total pattern as several unique and non-unique patterns.

Step 5: The Weight of the threshold will be considered with Average pattern count, Frequency of pattern, total pattern, no unique pattern count.

Step 6: The weight is computed for all the session and Maximum weight is computed.

- Maximum weight/3 =Threshold1
- Maximum weight/2=Threshold 2

Step 7: If weight<threshold 1; then it is high suspicious activity.

Step 8: If weight is between Threshold 1 and Threshold 2, then it is a medium suspicious activity.

Step 9: If weight >threshold 2; then it is low suspicious activity.

Random Forest Algorithm

Random Forest Algorithm is responsible for executing multiple independent decision trees and then predicts the class as Intruder or Non-Intruder based on training data. Figure 3 depicts the diagrammatic representation of the Algorithm.

The steps of Random forest to detect the DoS attack are as follows:

Step 1: Read all the data sets from the trained data with the following attributes. The Data Sets are collected in the following format and are admin will be able to view the datasets.

Step 2: Measure the count of the number of instances of historic data

Step 3: Divide the entire data set into multiple groups randomly

Step 4: For each of the subset execute the decision tree algorithm

Step 5: After executing the decision tree the output of decision trees which corresponds to maximum class is treated as the class label.

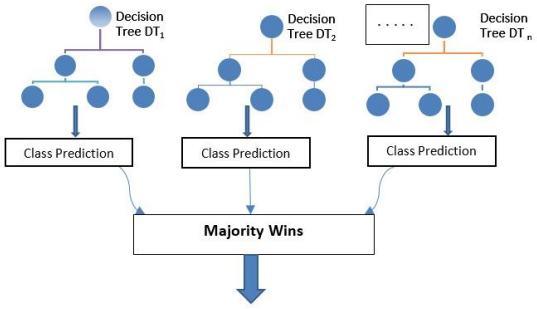

Figure 3: Random Forest Algorithm

Entropy: In statistical mechanics, entropy is an extensive property of a thermodynamic system

Entropy=

$$\sum_{i=1}^{n} stage1 + notstage1(totalstage1 + totalotherstage)Information\ Gain$$

Gain = Information Gain –Entropy

Step 6: After obtaining the gain of all attributes and pick the attribute with the highest information gain as the root node.

Step 7: Again from there create a subset for each value and determine the highest information gain like this repeated process until the stage is obtained

Step 8: The tree is constructed and finally based on input data sets the value is predicted.

3.6 Blocking Intruders

The users which are detected as intruders will be blocked here and they can't harm the system.at a certain period of time we can block the users.

4. RESULTS AND ANALYSIS

The customer can register into the application and after that the customer will be able to purchase different categories of books. We have created a Product Shopping website with around say 5 products where the user will be able to registration, login, product list & product buys. JAVA is used as the programming language with the Eclipse development tool and Heroku cloud for this application.

Every page visit and button click will be captured in the form of a habitat file. The habitat file will have the tracking based on action name, action type, time, and date of the session along with user id as well as the session-id. The admin executes the LCS algorithm and then determine the LCS for each of the session. The admin will be able to run the Intrusion Detection to detect whether the sessions are low, medium, or highly suspicious. If the session's falls under low or medium suspicious then the user will get notified otherwise the user will not be notified for non-suspicious. The Admin will be able to dynamically determine the DOS attacker by executing the random forest algorithm and then the user will be classified as an attacker if the repeated actions are performed based on random forest class label classification. Table 1 shows the sample result of establishing the DoS and Non-DoS attack using Random forest:

Attribute	Class1	Class2
1	4	8
2	10	2
3	5	9

Table 1: Summary of attributes

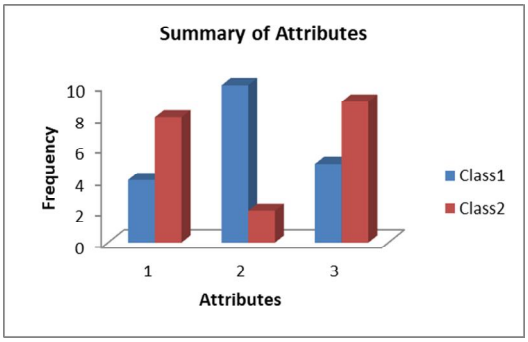

Figure 4: Attack Summary

- Class1 represents the number of times attribute resulted in DDos attack
- Class2 represents the number of times attribute resulted in a non-DDos attack

The Admin will be able to see the classification graphs as shown in figure 4, as well as the DOS, log in terms of which action has caused the DOS attack, and what is the reason i.e. page. Depending on the number of patterns matched, the attribute will be classified into DoS or non-DDos attack. The User who is responsible for the DOS attack will be shown a message while login. Your account is blocked because of the DOS attack. The admin will be able to unblock the user.

5. CONCLUSION

The application to allow users to purchase products. The application will track each button click as well as the navigation pattern of the user to generate the habitat file. Our approach will classify the users into the intruder and non-intruder categories and block suspicious activities.

REFERENCES

1. Preeti Mishra, Vijay Varadharajan, Emmanuel S. Pilli, Senior and Uday Tupakula, "VMGuard: A VMI-based Security Architecture for Intrusion Detection in Cloud

Environment" IEEE transactions on cloud computing 2018
https://doi.org/10.1109/TCC.2018.2829202

2. Josenilson Dias Araújo and Zair Abdelouahab, "Virtualization in Intrusion Detection Systems: A Study on Different Approaches for Cloud Computing Environments..."IJCSNS International Journal of Computer Science and network security, VOL.12 No.11, Nov 2012.
3. Swati Narwhal, Komal Pujari, Sakshi Rawat, Preeti Nagrath. Virtual Machine Introspection: Security Architecture Overview "International Journal of Innovative Research in Science, Engineering and Technology Vol. 6, Issue2, February 2017
4. Kai wang and Sameer Kulkarni, Yue Hu, "Cloud Security with Virtualized Defense and Reputation-based Trust Management", Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009
https://doi.org/10.1109/DASC.2009.149
5. Bharath Reddy S., Malathi D. and Shijoe Jose "An intrusion detection and prevention system in cloud computing a technical review" Iran Journal of Engineering and Applied Science VOL. 12, NO. 12, ISSN 1819-6608s, JUNE 2017
6. Farzad Sabahi, "Secure Virtualization for Cloud Environment Using Hypervisor-based technology", International Journal of Machine Learning and Computing, Vol. 2, No. 1, February 2012.
https://doi.org/10.7763/IJMLC.2012.V2.87
7. Irfan Gul, M. Hussain "Distributed Cloud Intrusion Detection Model", International Journal of Advanced Science and Technology Vol. 34, September 2011.
8. Snehal G. Kene and Deepti P. Theng"A Review on Intrusion Detection Techniques for Cloud Computing and Security Challenges" IEEE sponsored 2nd international conference on electronics and communication systems (ICECS '2015) ©2015IEEE
9. Uttam Kumar and Bhavesh N. Gohil." A Survey on Intrusion Detection Systems for Cloud Computing Environment "International Journal of Computer Applications (0975 – 8887) Volume 109 – No. 1, Jan 2015
https://doi.org/10.5120/19150-0573
10. Mrs. Vishali B. Kosamkar Intrusion Detection System in Cloud Computing: An Overview International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 4 Issue: 1 IJRITCC, January 2016 B. D. Payne, "Simplifying virtual machine introspection using life," Albuquerque, New Mexico,Tech. Rep., 2012
11. D. Shona and A.Shobana "Fast and Effective Network Intrusion Detection Technique Using Hybrid Revised Algorithms" IJETER Volume 4, Issue 11, November,2016
12. P. Mishra et al., "Intrusion detection techniques in a cloud environment: A survey," Journal of Network and Computer Applications, Elsevier, vol. 77, pp. 18–47, 2017.
https://doi.org/10.1016/j.jnca.2016.10.015
13. N. Chandra Shekhar Reddy et al "An Emperical Study on Support Vector Machines for Intrusion Detection" IJEAT Volume 7, No. 10 October 2019.
https://doi.org/10.30534/ijeter/2019/037102019