Journal of Nonlinear Analysis and Optimization
Vol. 14, Issue. 2 : 2023
ISSN : **1906-9685**

IDENTIFICATION OF UNSATURATED ATTACKS IN VIRTUALIZED INFRASTRUCTURES WITH BIG DATA ANALYTICS IN CLOUD COMPUTING

Boddupally Vinod Kumar, K Pranaya Vardhan, Kurceti Subba Rao, Assistant Professor, Dept. of Computer Science Engineering, Brilliant Institute of Engineering and Technology, Hyderabad, Telangana, India
Thipparthy Navya Sree, Student, Dept. of Computer Science Engineering, Brilliant Institute of Engineering and Technology, Hyderabad, Telangana, India

ABSTRACT

Security systems to protect virtualized cloud architecture typically include two types of malware detection and security analysis. Detecting malware typically involves two steps, monitoring the hotspots at various points in the virtualized infrastructure, and then using a regularly updated attack signature database to detect the presence of malware. 'Attack. It allows real-time detection of attacks, the use of special signature databases that are vulnerable to zero- day attacks that do not have attack signatures, and therefore traditional infrastructure. cannot detect complex attacks on virtualized infrastructure. Similarly, security analysis eliminates the need for signature databases using event correlation to detect previously undetected attacks, which are often unmanaged, and the current implementation is scalable in nature. In this article, we recommend BDSA's approach to establish a three-tier system for the continuous detection of future attacks. Initially, network logs from the visiting virtual machine and client application logs are sometimes collected from the visiting virtual machines and stored in HDFS. At this point, the strengths of the attack are removed with a connection scheme and a Map Reduce analyzer. Our BDSA approach uses HDFS distribution management and Spark's map-reduction display capability to address security and speed and volume issues.

Keywords: Cloud virtualized infrastructure, Cloud Computing, Hadoop Distributed File System,big data based security analytics (BDSA).

1. INTRODUCTION

The term "big data" has recently beenused for datasets that have become difficult to usein traditional database management systems. These are sets of data that exceed the capacity of software tools and storage systems commonly used to capture, store, manage, and process data within a reasonable amount of time.

Uncontrolled data growth is a burden for some organizations and they are collecting more data, even if their data warehouse is growing rapidly. Data is an asset of a company and the company generates revenue from it. However, Big Data is concerned about assigning a value to Big Data Analytics (BDA) due to insufficient dumping data. Big Data reflects the value of this data and increases revenue. As a result, Big Data Analytics is becoming a major player in research in various fields. BDA is a logical process ofanalysis for a very large dataset. BDA is used in avariety of adults. The boundary of the BDA is notlimited to IT but extends beyond its discipline. Thus, the BDA academic environment in interdisciplinary research is a great opportunity for people and practitioners in the field.

Cloud computing is a networked environment that focuses on computing andresource sharing. Cloud computing is defined as a set of virtualized IT resources. Cloud service providers typically use virtualization technologies combined with self-service capabilities to compute resources on network infrastructures, especially when the Internet and multiple virtual machines are hosted on the same physical server. Based on virtualization, the example of cloud computing allows for faster workload

http://doi.org/10.36893/JNAO.2023.V14I2.286-292

and scalability with faster delivery of virtual or physical machines. The cloud computing platform supports redundant, self-reinforcing, and highly scalable programming models that allow workloads to recover from inevitable hardware / software failures. Therefore, in the cloud, customers pay only for what they use and do not pay for local resources such as storage or infrastructure. The virtual device alleviates some critical management issues as most maintenance, software update, configuration, and other maintenance tasks are automated and centralized in a data center by a cloud service provider. is responsible. Virtualization is not a new technology and a complete network such as the cloud does not have enough security features.

The hardware for virtualized infrastructure hosted by software consists of virtual machines (VMs) that rely on multi- instance resources. A virtual machine monitor, also known as a hypervisor, manages, controls, and maintains software-defined multi-instance architecture. The ability to combine different IT resources and scale resources on demand has led to the widespread provision of virtualized infrastructure as a necessary cloud service [1].

This makes virtualized infrastructure an attractive target for cyber-attackers to launch attacks against illegal access. Use of advanced vulnerabilities, such as software vulnerabilities in the hypervisor source code, VENOM (Virtualized Neglected Operations Manipulation) [2], allows attackers to exit the guest virtual machine and access the hyperspect in the -jacent section. Similarly, attacks such as Heart Bleed and Shellshock, which exploit operating system vulnerabilities, can be used to gain access to guest virtual machine credentials against virtualized infrastructure and the escalation of privileges through distributed denial of service. (DDoS).

More recently, security scanning has used scanning to generate a large number of logs created by different security systems to maintain various points on the network to detect the presence of an attack. In Big Data analysis, the system can detect attacks that are not detected by signature or rule-based detection methods. While security analysis eliminates the need for a signature database using event correlation to detect attacks that were not previously detected, it is often not performed in real time and is inherently evolutionary in implementation. We are now proposing a new method, a big data security analysis approach, to detect complex attacks on virtualized infrastructure. Virtual Infrastructure Guest Virtual Machines (VMs) are hosted on Hadoop Distributed File System (HDFS). The virtual machine is used to collect the network log and the user application log. The BDSA Approach Schedule first captures attack characteristics using Event Correlation, using graphs, minimizes potential attack paths with the analyzer, then guarantees attack using two-stage machine learning, logistic regression, and confidence building.

2. BACKGROUND

Virtualization Components:

Virtualization is one of the mostimportant aspects of cloud computing. Virtualization is a technology that helps IT organizations optimizes the performance of their applications at the lowest cost, but it also presents their share of application delivery challenges that can cause security problems. The current interest in virtualization is related to virtual servers as virtualization servers can generate significant cost savings. The term virtual machine as a physical computer refers to the operating system and software computer running programs. The virtual machine operating system is called the guest operating system. In addition, there is a management layer called Virtual Machine Monitor or Manager (VMM) that creates and monitors all virtual machines in a virtual environment. Hypervisor is one of many virtualization methods that allow guests to run multiple operating systems on a host computer simultaneously, called hosts, called hardware virtualization. He is appointed because he is conceptually superior to a manager. The hypervisor provides a virtual operating system for guest operating systems and monitors the implementation of the guest operating system (guest operating system). Multiple instances of different operating systems can share virtualized hardware resources. A hypervisor installation is installed on the server hardware to run guest operating systems [3].

Virtualization Approaches:

In a traditional environment, where physical servers are connected to a physical switch, IT organizations can obtain detailed management information about the traffic between servers from that switch. Unfortunately, the level of information management is not usually provided by the virtual switch. Usually, the virtual switch is connected to the virtual machines from the physical switch through the physical network adapter. The lack of monitoring of traffic flows between and between physicalmachines at the same physical level affects safetyand performance analysis. There are many common approaches to virtualization, with differences in managing each virtual machine.

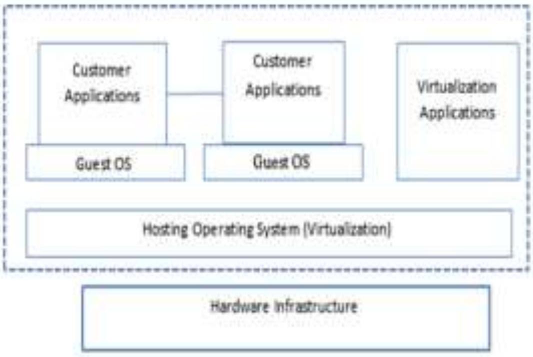

Figure1.Operating System based virtualization

In this process (Figure 1), virtualizationis provided by a host operating system that supports multiple discrete and virtualized guest operating systems on the same physical server, all of which have the same functionality in the same system kernel. Operation with special control over physical infrastructure. The host operating system can view and control virtual machines [4]. This approach is simple, but when an attackermanages host operating system scripts, all guest operating systems have control over the host operating system on that kernel. As a result, the attacker controls all existing virtual machines or will be installed in the future.

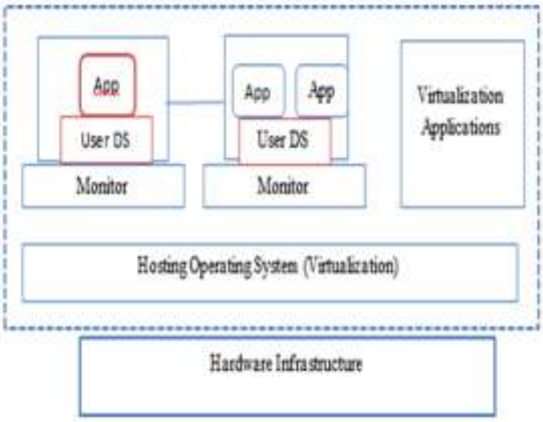
App
User DS
Monitor
App
App
User DS
Monitor
Virtualization Applications
Hosting Operating System (Virtualization)
Hardware Infrastructure

Figure 2 Application based virtualization

An application-based virtualization is hosted on top of the hosting operating system(Figure 2). This virtualization application then emulates each VM containing its own guest operating system and related applications. This virtualization architecture is not commonly used in commercial environments. Security issues of this approach are similar to Operating system-based.

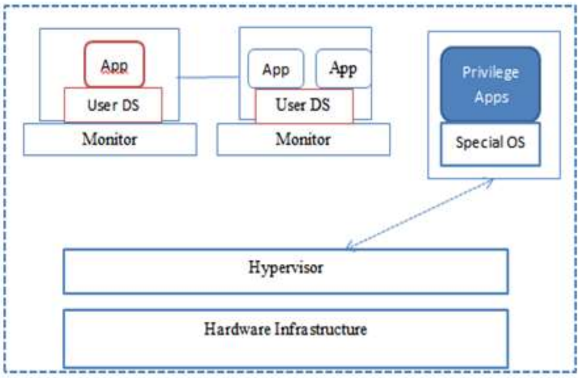

Figure 3 Hypervisor based virtualization

A hypervisor is available during machine startup to control system resource sharing across multiple virtual machines. Some of these virtual machines are for partitions that host a virtualization platform

http://doi.org/10.36893/JNAO.2023.V14I2.286-292

and host virtual machines. In this architecture, privileged partitions display and control virtual machines. This approach creates a highly controllable environment and can use additional security tools such as intrusion detection systems [2]. However, it is vulnerable because the hypervisor has one point of failure. If the hypervisor crashes or the attacker takes control, all virtual machines are under the attacker's control. However, controlling the hypervisor from the virtual machine level is difficult, but not impossible. Given this functionality, this layer has opted to implement the security architecture offered.

Malware detection in a virtualizedinfrastructure

Malware is any executable file designed to compromise the integrity of a running system. There are two popular approaches for detecting malware in cloud computing, namely collaboration within and outside the VM and hypervisor-assisted malware detection [4].

3. SYSTEM STUDY

Existing System

Security policies for protecting virtualized infrastructure typically include two types of malware detection and security analysis. Malware detection usually consists of two steps, monitoring hooks at various points in the virtualized infrastructure and then using a regularly updated attack signature database to detect the presence of malwareAttack. It allows attacks to be detected in real-time, and the use of a dedicated signature database is vulnerable to zero-day attacks that do not have attack signatures.

Limitations of the existing system

- Virtualized infrastructure cannot detect complex attacks.
- Analytics security analytics eliminates the need for a signature database using event correlation to detect previously undetected attacks, which are often unmanaged and scalable in their current implementations.
- The centralized implementation process delayed the management of the calculations.

4. PROPOSED SYSTEM

In this project, we are using a new approach to Big Data Security Analysis (BDSA) to protect virtualized infrastructures against complex attacks. Using network logs and user application logs collected from guest virtual machines stored in Hadoop Distributed File System (HDFS), our BDSA approach first captures attack functionality based on eventcorrelation. graphics, minimizing the map. Identification based on a potential attack path analyzer and subsequent two-stage machine learning confirms the existence of an attack, namely logistic regression and dissemination of views [7].

Advantages:

- Based In our work we use a new BDSA [5] approach (Big Data-based Security Analytics) to protect virtualized infrastructures against complex attack.
- Character capture is accomplished by graphically correlating events and minimizing detection based on a map attack path analyser.

5. PROPOSED ARCHITECTURE

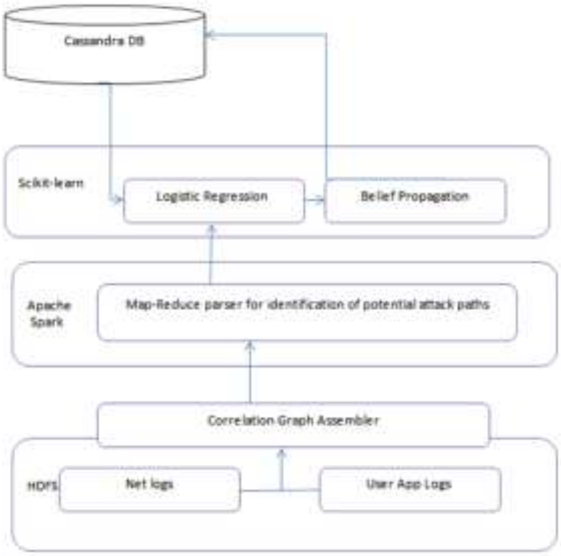

Figure4. Conceptual framework of the proposedbig data-based security analytics (BDSA) approach

6. SYSTEM MODEL

Big Data Security Analytics

The mission of Big Data Security Analytics (BDSA) [3] is to provide complete and up-to-date IT operations; therefore, security analysis can be timely and data-driven. Therefore, the Cloud Security Alliance focuses on the following security data:

- ➢ Acquiring the massive amount of data from several sources and external sources such as vulnerability databases.
- ➢ Performing more in-depth analytics on the data.
- ➢ Providing a consolidated view of security- related information.
- ➢ Achieving real-time analysis of streamingdata.

How the BDA helps in networksecurity and outline as follows:

- **Network Traffic:**The BDA supports the identification and prediction of malicious and suspicious resources and objectives, as well as emergency C traffic patterns. Unusual activity can quickly increase or decrease your network traffic. The BDA detects hidden anomalies.
- **Web Transactions:**The BDA improves the access control mechanism and assists in defining and evaluating emergency user access patterns, particularly in resource use or critical operations [9].
- **Network Servers:**The server configuration suddenly changes and runs abnormally. BDA is useful for detecting and evaluating abnormal server behavior.
- **Network Source:**The BDA identifies any device and so-called abnormal usage patterns.
- **User Credentials:**The BDA detects unusualconsumer behavior. User sends access time, amount, and emergency actions. The BDA improves the detection of abuse by all consumers [6].

Big data Traditional methods are difficult to manage and new big data processing methodssuch as data mining will be proposed. In addition,there are many machine learning algorithms for analyzing the system. However, machine learning is about security for better prevention and prevention of abuse. Key network security practices include:

Misuse Detection: Identifying abuse of resources, such as widespread denial of service, isimportant in a security system. Abuse does not directly harm the system; However, it slows down the system.

Anomaly Detection: An important problem nowadays is the identification of anomalies. A big problem with a security system is findingdisorganized information from many datasets.

The decision-maker systematically reflects the goals and priorities, the structure and the uncertainty of the problem, and quantifies them and other important aspects of the problem and how they interact. Starting from the architecture shown in Figure 4 above, this process starts with a variety of advanced attacks that perform event correlation based on a graph [10]. Does the guest periodically collect virtual machine event information? Log information is obtained from two sources, such as network and user applications, and stored in HDFS. Correlation diagrams are based on log information using the correlation graph builder and identify potential attack paths. Map Reduce Model Used to Analyse Correlation Graphs and Identify Possible Attack Paths

.

7. ALGORITHM USED

Algorithm : Security analytics inBDSA

1: *Initialize* : Obtain benign and malicious parameters of the attack feature from cassandra DB

2 : Train classifiers for monitored features using logisticRegression

3: *whileTruedo*

4 : Collect network and user application logs from guestVMs 5 : filter worklog entries using the guestVMs IPaddresses

6 : from correlated_Log

7 : Use Correlated_log to form a correlation graph g

8 : input G into Map □Reduce parser to identify potetial attack paths

□*attack _ paths* □,*Which is a subset of al* lg *raphpaths*

9 : *foreach* monitored fearture *attack _ pathin* □*attack _ path* □

10 : i □0

virtualized cloud infrastructures. Visitor virtual mission (VM) replicated network logs and client application logs are stored in Hadoop Distributed File System (HDFS). Our BDSA approach establishes a three-tier system to continuously determine traction power. Initially, network logs from the visiting virtual machine and client application logs are sometimes collected from the visiting virtual machines and stored in HDFS. At this point, the strengths of the attack are removed

11: *foreach* monitored fearturet

do

feature

inattack _ path with a connection scheme and a Map Reduce analyzer. Our BDSA approach utilizes HDFS

12 : calculate $P^{port\_change}, P^{Unknown\_Exet}, P^{in\_connect}, and P^{out\_connect}$

13 : *Pass* $P^{port\_change}, P^{Unknown\_Exet},$

$P^{in\_connect}, and P^{out\_connect}$ int *ostep.4ofa* lg *orithm*1. 14 : *Enddo*

15 : Enddo

16 : End

This process begins by extracting attack properties, which perform graph-based event correlation. Does the guest collect information from the virtual machine from time to time? Log data is sourced from two sources, such as network and user applications, and stored in HDFS. Correlation graphs are created based on log information using the correlation graph editor; It also manages the identification of potential attack paths. Diagram and identify possible attack paths, that is, the most common graphical paths, taking into account the IP addresses of guest virtual machines [8]. It is based on the observation that a compromised guest virtual machine generates more traffic when it attempts to communicate with an attacker. The Map Reduce model is used to analyse correlation graphs and identify possible attack paths.

CONCLUSION

This article recommends extensive security investigations to differentiate drive in distribution management and the ability to display map reduction in Spark to address speed and volume issues in security investigations. Other ways to recommend future data science work is to help security analysts and security tools make better use of security data, discover hidden patterns, and better understand system behavior.

REFERENCES

1. Win, T. Y., Tianfield, H., &Mair, Q. (2018). Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing. IEEE Transactions on Big Data, 4(1), 11–25. doi:10.1109/tbdata.2017.2715335.
2. Shuhui Zhang, XiangxuMeng, Lianhai Wang, LijuanXu, and Xiaohui Han, “Secure Virtualization Environment Based on Advanced Memory Introspection,” Security and Communication Networks, vol. 2018, Article ID 9410278, 16 pages, 2018.
3. https://doi.org/10.1155/2018/9410278.
4. P. K. Chouhan, M. Hagan, G. McWilliams, and S. Sezer, “Network based malware detection within virtualized environments,” in Euro-Par 2014: Parallel Processing Workshops. Porto,Portugal: Springer, 2014, pp. 335–346.
5. M. Watson, A. Marnerides, A. Mauthe, D. Hutchison et al., “Malware detection in cloud computing infrastructures,” IEEE Transactions onDependable and Secure Computing, pp. 192 –205, 2015.
6. Fattori, A. Lanzi, D. Balzarotti, and E. Kirda, “Hypervisor based malware protection with access miner,” Computers & Security, vol. 52, pp. 33–50, 2015.
7. K. Cabaj, K. Grochowski, and P. Gawkowski,“Practical problems of internet threats analyses,” in Theory and Engineering of Complex Systems and Dependability. Springer, 2015, pp. 87–96.
8. X. Wang, Y. Yang, and Y. Zeng, “Accurate mobile malware detection and classification inthe cloud,” Springer Plus, vol. 4, no. 1, pp. 1–23, 2015.
9. M. Watson, A. Marnerides, A. Mauthe, D. Hutchison et al., “Malware detection in cloud computing infrastructures,”IEEE Transactions on Dependable and Secure Computing, pp. 192 –205, 2015.
10. L. Chen, T. Li, M. Abdulhayoglu, and Y. Ye, “Intelligent malware detection based on file relation graphs,” in Semantic Computing (ICSC), 2015 IEEE International Conference on. Anaheim, California, USA: IEEE, 2015, pp. 85– 92.
11. P. K. Chouhan, M. Hagan, G. McWilliams, and S. Sezer, “Network based malware detection within virtualized environments,” in Euro-Par 2014: Parallel Processing Workshops. Porto,Portugal: Springer, 2014, pp. 335–346.