Journal of Engineering Sciences
Vol 10, Issue 10, Oct-2019
ISSN NO:0377-9254

A NOVEL BIG DATA-BASED SECURITY ANALYTICS APPROACH TO DETECTING ADVANCED ATTACKS ON VIRTUALIZED INFRASTRUCTURES

[1]A.EMMANUEL RAJU, [2]BUSHRA TAHSEEN
[12]ASSISTANT PROFESSOR
Department Of CSE
DR.K.V.SUBBA REDDY INSTITUTE OF TECHNOLOGY, KURNOOL

ABSTRACT

In this Paper Virtualized infrastructure in cloud computing has become an attractive target for cyber attackers to launch advanced attacks. This paper proposes a novel big data-based security analytics approach to detecting advanced attacks on virtualized infrastructures. Network logs, as well as user application logs collected periodically from the guest virtual machines (VMs), are stored in the Hadoop Distributed File System (HDFS). Then, extraction of attack features is performed through graph based event correlation and Map Reduce parser based identification of potential attack paths. Next, determination of attack presence is performed through two-step machine learning, namely, logistic regression is applied to calculate attack's conditional probabilities with respect to the attributes, and belief propagation is applied to calculate the belief in the existence of an attack based on them. Experiments are conducted to evaluate the proposed approach using well-known malware as well as in comparison with existing security techniques for virtualized infrastructure.

KEYWORDS: Virtualized I nfrastructure, virtualization security, malware detection and security analytics.

I. INTRODUCTION

Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to extract valuable information and knowledge, perform detailed analyses, as well as provide new opportunities and advantages. Anything ranging from customer names and addresses, to products available, to purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is the building block upon which any organization thrives. Now think of the extent of details and the surge of data and information provided nowadays through the advancements in technologies and the internet. With the increase in storage capabilities and methods of data collection, huge amounts of data have become easily available. Every second, more and more data is being created and needs to be stored and analyzed in order to extract value. Furthermore, data has become cheaper to store, so organizations need to get as much value as possible from the huge amounts of stored data. The size, variety, and rapid change of such data require a new type of big data analytics, as well as different storage and analysis methods. Such sheer amounts of big data need to be properly analyzed, and pertaining information should be extracted. The contribution of this paper is to provide an analysis of the available literature on big data analytics. Accordingly, some of the various big data tools, methods, and technologies which can be applied are discussed, and their applications and opportunities provided in several decision domains are portrayed.

The uncontrolled growth of data becomes a burden to some organizations, and they are collecting more data in spite of the rapid growth of their data warehouse. The data are the assets to the company, and the company generates revenue through these data. However, Big Data concerns assigning worth to those unworthy dumped data by Big Data analytics (BDA). The Big Data extracts value from those data and enhances the generation of revenue. Therefore, Big Data analytics is emerging as a dominant player in the field of research in varied areas. The BDA is a process of logical analysis on the very high volume of the dataset. The BDA is deployed in diverse fields. The boundary of the BDA is not limited to Computing but

encompasses areas beyond its discipline. Therefore, BDA in Interdisciplinary research is a golden opportunity for academia, industry people, and practitioners.

Moreover, the Big Data analytics is finding acceptability unimaginably vast area. Therefore, Big Data Security Analytics also emerges. The merging of Big Data analytics and Big Data security gives us many research opportunities.

A virtualized infrastructure consists of virtual machines (VMs) that rely upon the software-defined multi-instance resources of the hosting hardware. The virtual machine monitor, also called hypervisor, sustains, regulates and manages the software-defined multi-instance architecture. The ability to pool different computing resources as well as enable on-demand resource scaling has led to the widespread deployment of virtualized infrastructures as essential provisioning to cloud computing services[1].

II. LITERATURE SURVEY

K. Cabaj, K. Grochowski, and P. Gawkowski, "Practical problems of internet threats analyses" [6]

As the useful multifaceted nature of the malevolent programming expands, their examinations faces new issues. The paper displays these viewpoints with regards to programmed examinations of Internet dangers saw with the Honey Pot innovation. The issues were distinguished in light of the experience picked up from the examinations of adventures and malware utilizing the committed infrastructure sent in the network of the Institute of Computer Science at Warsaw University of Techno logy. They are talked about on the foundation of the genuine instance of an ongoing worm focusing on Network Attached Storage (NAS) gadgets powerlessness. The paper portrays the approach and information examination supporting systems and additionally the idea of general and custom Honey Pots utilized in the exploration.

X. Wang, Y. Yang, and Y. Zeng "Accurate mobile malware detection and classification in the cloud" [7]

As the dominator of the Smartphone working system advertise, subsequently android has pulled in the consideration of s malware creators and specialist alike. The quantity of kinds of android malware is expanding quickly paying little mind to the significant number of proposed malware examination systems. In this paper, by taking points of interest of low false-positive rate of abuse location and the capacity of oddity discovery to distinguish zero-day malware, we propose a novel mixture identification system in view of another open-source structure Cuckoo Droid, which empowers the utilization of Cuckoo Sandbox's highlights to dissect Android malware through powerful and static investigation. Our proposed system for the most part comprises of two sections: irregularity identification motor performing anomalous applications location through unique investigation; signature discovery motor performing known malware recognition and arrangement with the blend of static and dynamic examination.

M. Watson, A. Marnerides, A. Mauthe, D. Hutchison, "Malware detection in cloud computing infrastructures"[8]

Cloud computing is an increasingly well-known stage for both industry and purchasers. The cloud shows various exceptional security issues, for example, an abnormal state of dissemination and system homogeneity, which require extraordinary thought. In this paper we present a strength design comprising of an accumulation of self-arranging flexibility directors distributed inside the infrastructure of a cloud. All the more particularly we delineate the relevance of our proposed design under the situation of malware discovery. Here depict multi-layered arrangement at the hypervisor level of the cloud hubs and consider how malware location can be distributed to every hub.

III. SYSTEM ANALYSIS

Existing System

Security approaches to protecting virtualized infrastructures generally include two types, namely malware detection and security analytics. Malware detection usually involves two steps, Monitoring hooks are placed at different points within the virtualized infrastructure, and then a regularly-updated attack signature database is used to determine attack

presence. While this allows for a real time detection of attacks, the use of a dedicated signature database makes it vulnerable to zero-day attacks for which it has no attack signatures.

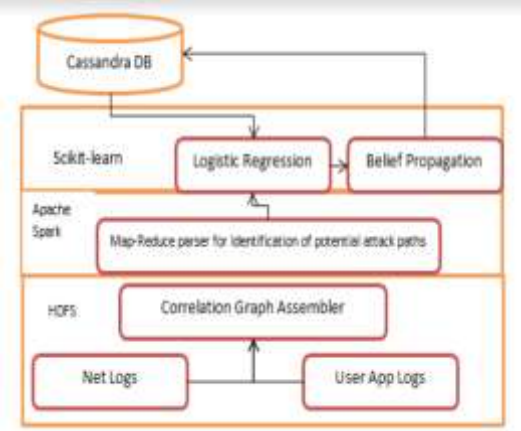
Proposed System

In this project we use a novel big data based security analytics (BDSA) approach to protecting virtualized infrastructures against advanced attacks. By making use of the network logs as well as the user application logs collected from the guest VMs which are stored in a Hadoop Distributed File System (HDFS), our BDSA approach first extracts attack features through graph-based event correlation, a Map Reduce parser based identification of potential attack paths and then ascertains attack presence through two-step machine learning, namely logistic regression and belief propagation.

Advantages:

In our proposed work we are using a novel BDSA(big data based security analytics) approach to protecting virtualized infrastructures against advanced attacks.

Extraction of attack features is performed through graph-based eventcorrelation and Map Reduce parser based identification of potential attack paths.



SYSTEM ARCHITECTURE

IV. IMPLEMENTATION

Big Data Security Analytics

The objective of Big Data security analytics (BDSA)[3] is to provide comprehensive and up-to-date IT activities; thus, security analytics makes timely and data-driven decisions. Thus, the Cloud Security Alliances emphasizes on security data as follows:

• 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 streaming data.

• **Network Traffic:** The BDA supports in detecting and predicting malicious and suspicious sources and destinations, along with abnormal traffic patterns. The network traffic rapidly rises or sharply declines due to some abnormal activities. The BDA discovers those hidden abnormalities.

• **Web Transactions:** The BDA enhances the access control mechanism and also helps in detecting and predicting abnormal user access patterns, particularly in the usage of critical resources or activities.

• **Network Servers:** The server configuration changes suddenly and behaves abnormally. The BDA is useful in the detection and predicting abnormal behavior of the server.

• **Network Source:** The BDA detects and predicts abnormal usage patterns of any machine.

• **User Credentials:** The BDA detects abnormal user behavior. The user sends abnormal access time, amount, and operations. The BDA improves in discovering all the misbehavior of users.

Big Data is difficult to deal with the traditional techniques, so new methods are provided to deal with big data, such as data mining. Also, there are numerous machine learning algorithms to analyze a system. However, machine learning is engaged with security for better prevention, and protection of malpractices. The prime network security methods are as follows:

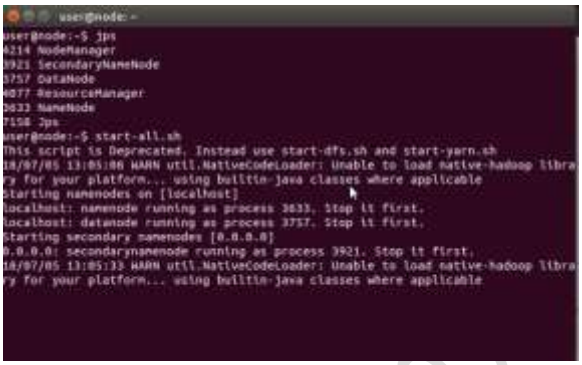
1) Misuse Detection: resource misuse detection becomes prominent in the security system, for example, distributed denial of service. The misuses do not directly harm the system; however, it slows down the system.

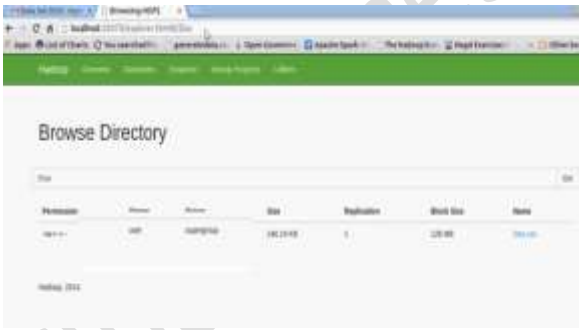
2) Anomaly Detection: Anomaly detections are vital issues nowadays. The digging for anomaly information from a large set of the dataset is the crucial challenge in the security system.

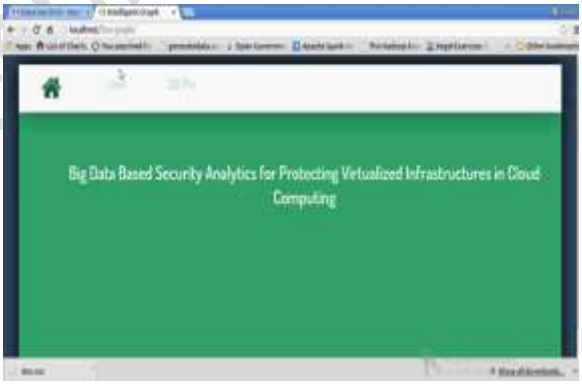
Decision-maker thinks systematically about the objectives and preferences, the structure and uncertainty in the problem, and model them quantitatively and other essential aspects of the problem and their interrelationships.

From above proposed architecture figure 2.4, the process has been started from different advanced attacks it carries out Graph-Based Event Correlation. Which collect the event information periodically from the guest virtual machines. Log information's are obtained from two sources such as network and user applications and stored in HDFS.Correlation graphs are made based on logs information using Correlation Graph Assembler . furthermore it carries out the Identification of Potential Attack Paths. A MapReduce model is used to parse the correlation Graphs and identify the potential attack paths.

RESULTS







. CONCLUSION

This project concludes with a novel Big Data-based security analytics system to detect improved attacks on virtualized infrastructures. Virtualized

infrastructures can use virtual machines stored on the Hadoop Distributed File System (HDFS). Our BDSA method an ll be an advantage for distributed HDFS processing and the real-time ability of Map-Reduce's Spark to address security-related speed and volume issues. N order to solve the problem of sincerity caused by zero-day attacks, our BDSA approach responds to this problem by using a mechanism that s processed for the processing of logistic regression classifiers.

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. https://doi.org/10.1155/2018/9410278.

[3] P. K. Chouhan, M. Hagan, G. McWilliams, and S. Sezer, "Network based malware detection within virtualised environments," in Euro-Par 2014: Parallel Processing Workshops. Porto, Portugal: Springer, 2014, pp. 335–346.

[4] 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.

[5] A. Fattori, A. Lanzi, D. Balzarotti, and E. Kirda, "Hypervisorbased malware protection with accessminer," Computers & Security, vol. 52, pp. 33–50, 2015.

[6] 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.

[7] X. Wang, Y. Yang, and Y. Zeng, "Accurate mobile malware detection and classification in the cloud," SpringerPlus, vol. 4, no. 1, pp. 1–23, 2015.

[8] 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.

[9] 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.

[10] P. K. Chouhan, M. Hagan, G. McWilliams, and S. Sezer, "Network based malware detection within virtualised environments," in Euro-Par 2014: Parallel Processing Workshops. Porto, Portugal: Springer, 2014, pp. 335–346.

[11] H. Kim, I. Kim, and T.-M. Chung, "Abnormal behavior detection technique based on big data," in Frontier and Innovation in Future Computing and Communications. Springer, 2014, pp. 553–563.