An Analysis of User-oriented Behaviour-based Malware Variants Detection System

S. Varshavardhini, R.A. Karthika, K.M. Monica

Abstract--- A virtualized infrastructure (VI) is implemented through one or more virtual machines that depend on built-in software defined multiple instances of hosting hardware. This infrastructure model added an advantage of gathering different computing resources and on-demand resource scaling that facilitated extensive deployment of VI to cloud computing services. BDSA finds the potential attack and protect in VI opposing vulnerabilities. HDFS is used to store the backend information. Security analytical algorithm applied on logs captured at various points within network to identify the attach existence. Ref[1]Graph based event correlation and Map Reduce Parser methodologies are used to identify the attack paths through the network logs obtained.Ref[1] Two step machine learning is used to determine the attack presence, attack's conditional probability based on attributes is calculated through logistic regression and existence of attack on network is calculated through belief propagation. This has steered way for cyber attackers to launch attacks for illegal access on virtualized infrastructures.

Keywords--- Virtualized Infrastructure, Malware Detection, BDSA Approach, HDFS.

1. INTRODUCTION

Malware or Malicious Software has been specifically designed to pull or twist out of shape or give misleading or false account of impression of data that is in the mobile or computer operations. These malware operations are gaining importance in the globalized economy wherein data is very valuable at all levels. Therefore security of data in computer and mobile is preciously important and to be secured. With the Onset of usage of mobile for various applications like mobile banking, internet banking, ecommerce the security of personal data is very important. Big Data analytics detect attacks with the help of huge amount of logs gathered from different network points which can't be possibly discovered through signature or rule based methods. Techniques that are commonly used for security analytics are graph based event correlation and clustering. Assembling of data items in unbalanced data sets as graph based on their similarities and features is clustering. To delete unknown attacks security analytics, clustering works on a systematic way which summarizes the data of same characteristics. [1]Map reduce model is applied to the grouped clusters to detect the possible attacks in groups by allowing to carry out efficiently the detection. [1]Graph based event correlation

---
Revised Version Manuscript Received on 30 May, 2018.
S. Varshavardhini*, PG Graduate, Department of Computer Science and Engineering, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai. (e-mail: varshu28@gmail.com)
R.A. Karthika, Associate Professor, Department of Computer Science and Engineering, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai.
K.M. Monica, Assistant Professor, Department of Computer Science and Engineering, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai.

represents the events from the logs obtained in the graphical mode thus overcoming the boundary. Therefore BDSA approach is in advantage of processing HDFS and real time ability whereas map reduces addresses challenges in security analytics in velocity and vulnerability.

2. RELATED WORKS

Anjum Khairi, Sadia azhar, Muhammad Waseem at [2], proposed which a major research is done for more than ten years now. In convergence of the technologies that exists recur sly would interconnect the objects physically. Though IOT is rapidly developing there are lots of questions and doubts regarding the security and privacy for its sustainability .This paper provides a well-defined security architecture with confidentiality and ensures the privacy of user's and their security which could be adopted by masses widely by analyzing the security issues and challenges.

Clemens Kolbitsch, Paolo Milani Comparetti, Christopher Kruegel, Engin Kirda, Xiaoyong Zhou, and XiaoFeng Wang at [4] proposed malware detection to work out and to replace or as an alternative to the traditional anti-virus software. This approach analyzes malware program is in a controlled environment which is characterized by its behavior as a model. These models show the information flows between system calls necessary to protect against malware which cannot be easily overcome or evade by simple intelligible or using variety of protection techniques. So they extracted responsible program slices for the flows of information as such and execute those slices to match their models against the behavior during runtime of an unknown program. This experiment shows that approach is effectively detects running malicious code for the end user at linear overhead.

Faheem Ullaha, b, Muhammad Ali Babara, b. at [5] proposed about an important of research and practice which is a demanding area of research and practice aimed to protect networks computers and data from the access by unauthorized security analysis and event data by using various tools of data and latest technologies. The report of this paper is systematically reviewed to aim at identifying the reported attributinal quality and architectural techniques for Big Data Cyber security Analytic systems. Method: The method applied here is review of related literature systematically and has reviewed 74 studies primarily which were well defined belong to the criterion selected.

Results: The findings that has come out is twofold: (i) 12 most often reported quality attributes were identified and their significance is justified for Big Data Cyber security Analytic Systems; and (ii) 17 architectural techniques to address the quality attributes which were commonly associated to Big Data Cybersecurity Analytic systems were identified.

Safaa Salam Hatem, Dr.Maged H.Wafy, Dr. Mahmoud M. El-Khouly at [3] proposed antivirus software and stopping malicious that are no need. Thelong established host antivirus is question for its long term effect. The increasing complexity has resulted invulnerabilities exploited by malwareis not being fully detected or fails to detected by antivirus software. To end hosts based and malware practices and detection model. This model enables to identify all malware and the software which were not authorized. The benefits that could be advantageous are inclusion of detecting exactly the malware and enhance literally rhetorical in finding deploy ability in a better manner is possible in this approach. Detecting malware usage especially in cloud computing needs less weight storage in a cross manner and service network. The unique feature in this model suggested is combination of techniques in detection, analyzing signatures that are static and importantly detecting dynamic analysis. This model can detect 35% better detection coverage than the latest challenges when compared to a single antivirus engine and there for 98% detection is possible in the cloud environment.

Gonc¸alves1, Bota2, Miguel Correia at [7] proposed that it is very difficult to manage the complex network infrastructures. The misbehaviourial ways that are unpredictable in there infrastructure which contains a large number of devices. A lot of this kind of device that have logs with innumerable information about the infrastructure security, their reliability and its performance. It is an important however to extract the information available from the data. They have presented a new approach for assessing the security usages of the logs which are taken from original telecommunication network. Machine learning and data mining techniques are used to analyze information and by discovering the misbehaving hosts semi-automatically without instructing the system about how the hosts misbehave.

3. PROPOSED SYSTEM

The key advantage of proposed system, the implementation of efficient big Data based Security Analytics approach to identify the vulnerabilities in network attacks. The proposed system is based on BDSA approach to identify the potential attacks in VI and protect VI against the vulnerabilities. The BDSA approach primarily determine the attacks through graph-based event correlation and in next step, it identifies the potential vulnerabilities of attack and discovers attack presence by logistic regression and belief propagation machine learning process.

4. METHODOLOGY

4.1 Virtualized infrastructures

Virtualization means creating something, alike rather than actual version of a particular thing it may also be virtual computer storage device or resources of computer network. VI gathers all the login activities of the guest VMs and performs all kinds of attacks such as Privilege issues to Distributed denial of service. Over the time data collected from virtualized infrastructures should be detected and be capable of determining potential attack as well as protect all possible malware issues.

4.2 Distributed files

Hadoop Distributed File system acts as a backend to store all the intra networks login and user application login which are gathered from various virtual guest infrastructure and machines.

4.3 Event management

Security analytics eliminates the use of signature database to recognize hidden attacks by using event correlation. It may not be performed in actual practice which will be naturally non-scalable. Clustering finds out the possible attacks present in the means of grouping more common attack in the same character and it is limited in determining correlation accurately sometimes that happen in between the events. To overcome the limitation on virtualizing the happening from these logs that are collected in sequence as graph by using graph based event correlation.

4.4 Detect attacks

From the security framework Graph-based event correlation approach available to build efficiently to identify attacks within complex infrastructures. A temporary graph node is generated from the events gathered from different network sources to derive various event correlations for detecting threats. Prolonged or continuous period is needed for the collection of data qualitatively which requires a motive or awareness or be awake of the threats that have happened over a period of time within the same network. It is hard and may not be possible to concentrate on the instant events and immediate prompt actions uncompromisingly with in the network.
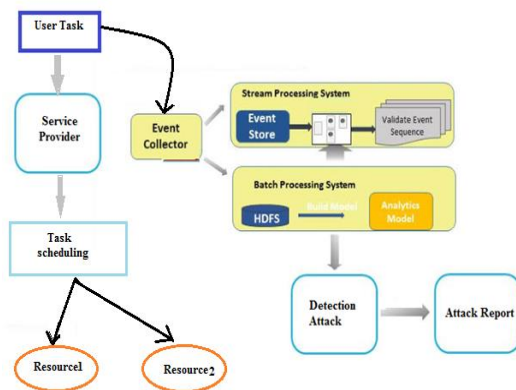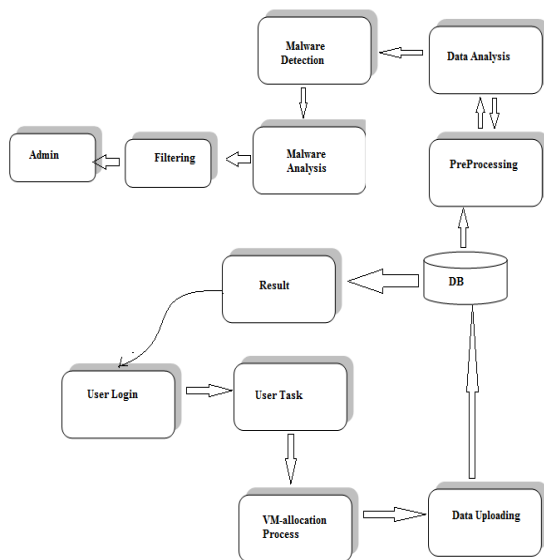

**Figure 1: Architecture Diagram**

Figure 2: Flow Diagram

5. RESULTS &DISCUSSION

Window displays the list of incoming client connections in Server console. Agent Window displays the file information such as Source of the File, Download details and Status of file action then it displays the IP address and Type of Attack initiated by the File. The activities performed in the respective Virtual Machines and number of times file accessed so far. Interface to Upload and Download the files to/from the server and display number of times file accessed so far.
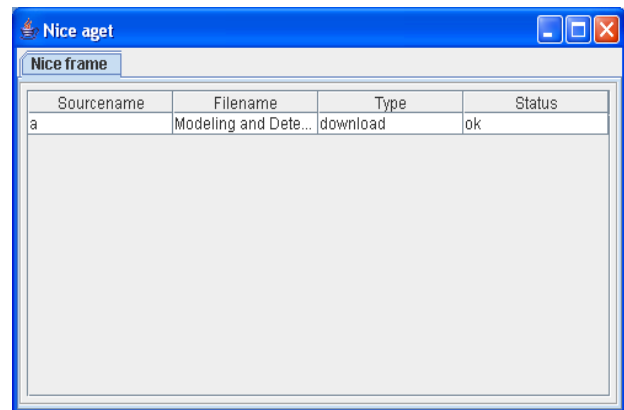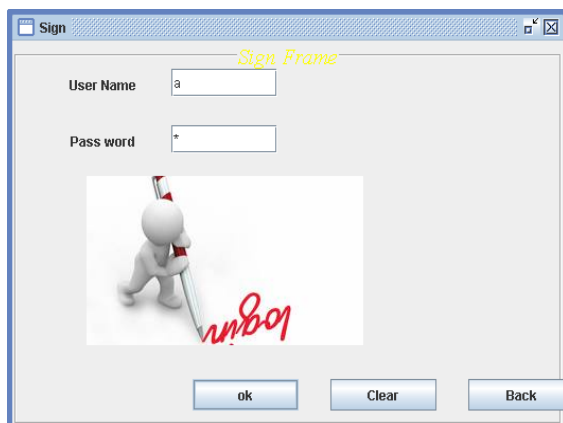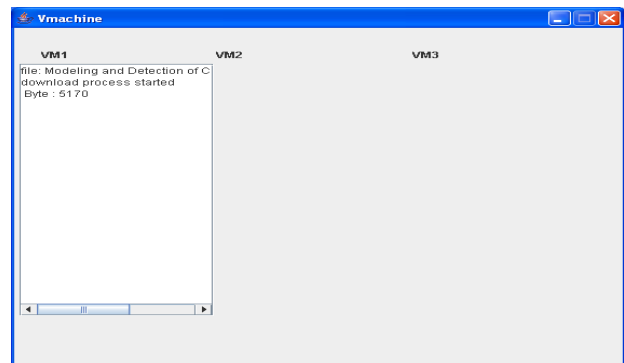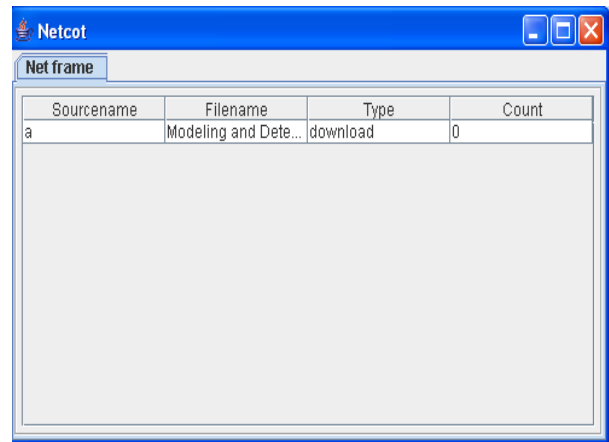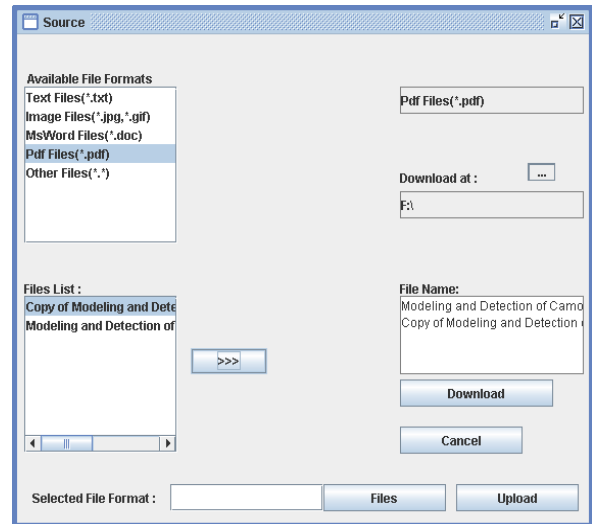
Nodename	IP	Attacktype
a	claro4	dos attack
a	claro4	dos attack
a	claro4	dos attack

Sourcename	Filename	Type	Status
a	Modeling and Dete...	download	ok
a	JA6001.PDF	download	ok
a	JA6017.PDF	download	ok
a	JA6001.PDF	download	dos attack
a	JA6001.PDF	download	dos attack
a	JA6001.PDF	download	dos attack

6. CONCLUSION

Big Data based Security Analytics (BDSA) protects virtual infrastructure which are from progressive headway attacks. Hadoop Distributed file System is used to store all the network logs and logs of user application which were gathered from other virtual machines. By using BDSA concept invasion features are immediately removed through graph-based event correlation, then by means of Map Reduce Parser identify potential attack paths to remove. Later by making use of logistic regression and belief propagation fix all attacks present by increasing number of guest VM's.

7. FUTURE ENHANCEMENTS

If an unauthorized user seeking to open the file of some specific user, they must send request to user of which they need access. It can be seen in user's window. If the request is accepted then unauthorized user can view the files only through seeking permission. Or else the file will be secured from the requester. Another enhancement is blocking the IP of the attacker.

REFERENCES

1. Huaglory Tianfield, and Quentin Mair, "Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing", Volume 4-NO.1, January-March 2018
2. Anjum Khairi, "A Critical Analysis on the Security Concerns of Internet of Things (IoT)", International Journal of Computer Applications , Volume 111 - No. 7, February 2015
3. Dr.Maged H, "Malware Detection in Cloud Computing" International Journal of Advanced Computer Science and Applications, Vol. 5, No. 4, 2014
4. Paolo Milani Comparetti, Christopher Kruegel, Engin Kirda, "Effective and Efficient Malware Detection at the End Host"
5. Faheem Ullah ,"Architectural Tactics for Big Data Cyber security Analytic Systems"
6. Christian Wressnegger, "Content-based Anomaly Detection for Industrial Control Systems"
7. Daniel Gonc¸alves, Jo˜ao Bota, Miguel Correia, "Big Data Analytics for Detecting Host Misbehaviour in Large Logs"
8. L. Aniello, A. Bondavalli, A. Ceccarelli, C. Ciccotelli, M. Cinque, F. Frattini, Luca Invernizzi, "Detecting Malware Distribution in Large-Scale Networks" "Big Data in Critical Infrastructures Security Monitoring: Challenges and Opportunities".