Survey on Malware Detection Methods And Malware Detection Technologies

Dayakar Suddala[1], Dr.K.Butchi Raju[2]
[1]Dept of CSE
[2]Associate Professor,Dept of CSE
[1, 2]GokarajuRangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India

Abstract- *The usage of internet and cloud computing is widely increasing day-to-day. The cloud services are widely needed in every sector of the business and education. It is creating a huge need of virtual infrastructure which provides the actual cloud services. As, the usage of cloud services is increasing at a high rate, the security attacks on this cloud infrastructure are also increasing. This paper is presenting the various methods of malware detection methods and multiple malware detection technologies. Cloud services are prominent within the private, public and commercial domains. Many of these services are expected to be always on and have a critical nature; therefore, security and resilience are increasingly important aspects. In order to remain resilient, a cloud needs to possess the ability to react not only to known threats, but also to new challenges that target cloud infrastructures. This paper introduces cloud anomaly detection approach, signature based, rule-based, hypervisor malware detection and many other methods of malware detection. More specifically, we exhibit the applicability of novelty detection under the one-class Support Vector Machine (SVM) formulation at the hypervisor level, through the utilization of features gathered at the system and network levels of a cloud node.*

Keywords- Malware detection, IDS, Computer System Attacks, Network Security.

I. INTRODUCTION

In these technology days, every business and organization is operated using the high-end technologies like cloud computing, virtualization and mobile ad-hoc networks. These and more technologies have been dramatically changing the business and digital world. The usage of high computing services is increasing.Day-by-day, this usage increases the data and sensitive information.Virtual machine plays vital role in cloud computing. In the cloud computing the computing devices may move from one place to another place. This mobility may leads the loos coupling of network connections. This loose coupling may leads the network related malware attacks.This useful and confidential information may be interrupted by the malicious software and suspicious malware. This attack may be destroys the entire business. These attacks and unconditional execution programs need to be detected and prevented in order to protect the business and technology.

There are some methods and technologies that are used to detect the malware and are also used to prevent these attacks. These methods are deferent for deferent type of malware attacks. These attacks mainly caused by viruses, worms, spywares, ad-wares, Trojans and botnets. To detect these suspicious malwares a special and effective methods are needed. Due to the high distraction of the system, those malwares need to be analyzed and completely removed from the system. To do all this prominent effective activity, the detection methods and technologies must be more effective and accurate than the strength of the malware. Mainly these methods are divided into three categories, they are, Static, Dynamic and Hybrid. Static method detects the malware based on a predefined signature. Dynamic approach detects the malware by its current behavior at the network node. Hybrid approach is the combination of both static and dynamic approaches. Malware detection technologies are used to apply these methods practically through some tools. Some of the technologies are Host-based IDS, Network-based IDS, Hybrid IDS, and APIDS. All these methods and technologies are surveyed in-depth in the following sections.

II. LITERATURE SURVEY

A review over the various techniques and technologies which are used for malware detection in cloud computing is presented in this section.

2.1 Malware Detection Techniques:

Thu Yein Win, HuagloryTianfield and Quentin Mair [1] proposed a malware detection technique to detect malware and rootkit is presented. That Takes a system call monitoring and system call hashing together and a support vector machine based external host monitoring system is also used. In monitoring system call all the system calls triggered by the users, are monitored over the parameter before execution. In system call hashing, all the stored monitored system called

copies are checked before installation. Then a support vector machine based system is used which classify all the malware and rootkit attacks in virtualize cloud system. But that technique suffers in accuracy thus a new technique to provide accurate results for intrusion detection.

Michael R. Watson, Noor-ul-hassanShirazipreseted [2] support vector machine-based technique to detect intrusion in the cloud computing architecture. In this a support vector machine based monitoring system is used at hypervisor level to detect malware in cloud computing system. In cloud virtualization there is various type of threats can be found which requires an enhanced functionality to deal with such cloud computing threats. Thus an enhanced mechanism is required to provide an accurate result for this problem.

Xiaoguang Han, Jigang Sun, Wu Qu3, Xuanxia Yao [3] clarified a description over the various malware detection techniques which used for the intrusion detection is presented. In that various machine learning techniques can be used to provide a detection mechanism for cloud computing. In that way it requires an enhanced technique to provide accurate intrusion detection for such techniques.

Angelos K. Marnerides [4] introduces a malware detection technique for virtualize cloud environment is presented. There are various system resilience related risks are occurred in these virtualizations techniques. New technique is required which can deal with the issues related to the risk in the programming. In that technique a NAE (Network Analysis engine) and system analysis engine is used to deal with such issues. But these techniques are not efficient to deal with such issues.

2.1.1 Signature-based detection:

NwokediIdika, Aditya P. Mathur [5] "A Survey of Malware Detection Techniques" are described a system which can detect the malicious software by its model of behavior. The collection of all these models of behavior isspecifies the signature of the malware. This signature should always be able to identify the abnormal behavior of the any malware. They also stated that once a signature has been created, it is added to the signature-based method knowledge. This knowledge is represented as the repository. This repository is pertains to malware detection. They extensionally added that the signature-based approach has the sub-contents as Dynamic Signature-based Detection, Static Signature-based Detection and Hybrid Signature-based Detection. All these methods discussing that there are many approaches to detect the malware in the cloud infrastructure. They proposed a drawback of this method is it cannot detect the zero-day attacks.

Pranit Gaikwad, Prof. Dilip Motwani, Prof. Vinayak Shinde [6] are proposed another way of this method in "Survey on Malware Detection Techniques". They stated that the signature-based approach is also called as a pattern matching mask or finger printing technique. A signature is a set of sequence injected into the computer program by malware program developers, which idiosyncratically identifies a particular malware. For recognizing a malware in the code, the malware detector search for a formerly specified signature in the code. Commercial antivirus scanners look for signatures which are typical sequence of bytes within the malware code to declare asthe computer program scanned is malicious. There are three categories of malwares: basic, polymorphic, metamorphic malwares.

Jyothilandage and Prof.M.P.Wankhade [7] had discussed the malware detection as follows. This type of detection is also called as Misuse detection. It maintained a database of signature and it uses this signature to compare with the malware. These signatures are created by disassembled by the code of malware binary. This disassembled code is analyzed and features are extracted. These features are used to develop a signature of a malware family. A specified library of know code is maintained and updated frequently. So this technique is able to detect the known malware accurately, less amount of resources are required to detect the malware. It mainly focuses on signature of attack. The drawback of this method is it can't detect the new, unknown instance of malware as no signature is available.

This malware detection method is also presented in way by Imtithal A. Saeed, Ali Selamat and Ali M. A. Abuagoub [8] in " A Survey on Malware and Malware Detection Systems" that all scanners use the signature-based malware technique to identify the unusual program code. However there are methods using these techniques: Dynamic method, Static method and Hybrid method. Dynamic method uses run-time information of malware when it is executed in a memory. Static method uses the information which is from the extraction of static malware (when it is in disk). Hybrid method uses both static and dynamic methods.

2.1.2 Heuristic-based detection:

For this method of detection,JyothiLandage and Prof.M.P.Wankhade [7] are described a way in "Malware and Malware detection Techniques:A survey" as it is also called as behavior or anomaly-based technique to detect the malware

based on its behavior. This method analyzes both known and unknown malwares. The behavioral parameter is the combination of factors like source and destination address of the malware, types of attachments and other countable statistical features. This method includes two phases: Training phase and Detection phase. During training phase the behavior of the system is observed in the absence of an attack. A machine learning technique is used to create a profile of that behavior. In detection phase, this profile is compared with the current behavior and the differences flagged as potential attacks. They stated the advantage of this method is, it is capable to detect known and unknown malware attacks as it is focuses on the behavior of the malware. They also stated the disadvantage of this method is, it is need to be update data describing the nature of the malware and statistics in normal profile. It needs more resources like CPU time, memory and disk space.

According to Imtithal A. Saeed, Ali Selamat [8] we can observe, the authors said that this method is used in Artificial Intelligence (AI). Neural networks and Fuzzy logic are also used in malware detection techniques. A machine learning algorithm Genetic algorithm is also used to detect the malware. The main advantage of this technique is the derivation of solutions from multiple directions with no need for prior knowledge about system behavior.

Pranit Gaikwad, Prof. Dilip Motwani ,Prof.Vinayak Shinde [6] described this method is also known as proactive technique. Instead of searching of particular signature, the malware detector searches for the instructions that are not present in the application program. It leads to an advantage that is easy to detect the new variant of malware that has not yet been discovered. They also proposed the various types of Heuristic methods. File-based heuristic method analyzes the contents of the file, purpose of file and location of the file, etc. If the file orprogram contains commands to delete or damage other file, than it is painstaking as malicious. Weight-based heuristic analysis technique, each application is weighted according to the danger it may possess. Application program is considered as malicious when weighted value exceeds the predefined threshold value. Rule-based heuristic analysisextracts the rules defining the application. These rules are then matched with thepreviously defines rules. If the rules are mismatched, then the application contains malware. Generic signature analysis detects the variants of the malware. A variant of malware means, the malware are different in behaviour, but belong to same family like “identical twins”. The previously defined antivirus definitions will be used by this technique to discover variants of malware.

NwokediIdika, Aditya P. Mathur [5] explained this heuristic-based method as it is proposed as an Anomaly based method. It usually occurs in two phases they are Training phase and Detection phase. In training phase, the detector tries to learn from the normal behaviour of the malware. In the detection phase, the repository of the behaviour is used to identify the malware. The key advantage of this method is it can detect the zero-day attacks. The two fundamental limitations of this techniqueis its high false alarm rate and the complexity involved in determining what features should be learned in the training phase.

2.1.3 Specification-based Detection:

NwokediIdika, Aditya P. Mathur [5]proposed Specification-based detection is a type of anomaly-based detection. It always tries to address typical high false alarm rate associated with most anomaly-based detection techniques. Instead of attempting to approximate the implementation of an application or system, specification-based detection attempts to approximate the requirements for an application or system. In specification-based detection, the training phase is the attainment of some rule set, which specifies all the valid behaviour any program can exhibit for the system being protected or the program under inspection.The main disadvantages of specification-based detection is that it is often difficult to specify completely and accurately the entire set of valid behaviour a system should exhibit. This method further divided into three types of methods: Dynamic method, Static method and Hybrid method.

JyothiLandage and prof.M.P.Wankhade [7] stated that this method is derivative of the behaviour-bade detection method. This method relies on program specification that describes the intended behaviour of the security critical program. It involves monitoring program executions and detecting deviation of their behaviour from the specification, rather than detecting the occurrence of specific attack patterns. Unlike the behaviour-based method, this method is based on manually developed specifications that capture legitimate system behaviour. In this paper they stated the advantage of this method is, it can detect the known and unknown instances of malware. The level of false positive is low, but the level of false negative is high. The main drawback of this method is, it is not as effective as behaviour-based detection method in detecting new attacks.

Vinod P., V.Laxmi, M.S.Gaur [9] have described specification-based detection method as itis derived from behavior-based method. This method is approximates the requirements of application or system. This method has a training phase which attempts to learn the all valid behavior of

a program which needs to be inspected. The limitation of this method it is very difficult to accurately specify the program. They also proposed a tool called Panorama which captures the system wide information flow of the program under the inspection of the system. It also checks the behavior against a valid set of rule to detect malicious activity.

Elizabeth Hansson [10] proposed an approach to demonstrate this specification-based malware detection method. This author is described it as a model of the protocol in order to detect attacks based on protocol specification. RFC's and other descriptions of protocols provide these specifications. This approach has a challenge that is how to define the set of rules that describe correct operation of the protocols to efficiently detect attacks. This author proposed an achievement of this method with extended finite state machine.

2.1.4 Other Approaches

Hypervisor malware detection method

Thu Yein Win, HuagloryTianfield and Quentin Mair [11] described this method as it uses the underlying hypervisor to detect malware in guest VM's. It is also designed to detect the botnets in the guest VM's. This method installs the network sniffers on hypervisor to monitor external traffic as well as inter-VM traffic. It is also proposed using guest application and network flow characteristics. This scheme initially usesLibVMI to extracts the key process features from the processes running within the VMs and then it usestcpdumptogether with the CoralReef network packetanalysis toolfrom CAIDA (Centre for Applied Internet Data Analysis) to extract network flow features. This method is also used in Access-Miner.

In-VM based Monitoring method

Thu Yein Win described this method as in-VM method consists of an agent running within the guest VM [11]. A remote scrutiny servermonitors the behaviour of the VM's behaviour. When a potential malware execution is detected the in-VM agent sends suspicious executable to the scrutiny server. A signature based database is then used to verify malware presence and then informs to the in-VM agent for the results.

Erick Bauman [12] presented a mathematical representation of this in-VM based detection method. A security monitoring system can be defined as follows

$$M(S, P) \square \{True, False\}$$

Here M denotes the security enforcing mechanism, S denotes the current system, Pdefines the Predefined policy. From the above equation we can derive that if the current state S satisfies the security policy P, thenit is in a secure state (True). If M is an online mechanism, it can continue its execution. Otherwise, it is insecure (False). If an attack is detected, M can halt the execution and then it reports the attack presence. In this paper they also stated that the crucial step of a security mechanism is collecting the state informationS from predefined sensors, such as those embedded in the running OS or processes. Then monitoring with a well-defined security policy P is the final state in this crucial step. In-VM based monitoring resides within the OS. This makes the in-VM based method advantageous in Rich abstractions and Fast speed. Disadvantages of this method aregenerating the false state (S), Tamper with the security policy (P) and Tamper with the enforcing mechanism (M).

Out-VM based Monitoring method:

Erick Bauman [12] proposed a method which overcomes the limitation of the in-VM method. The monitoring system moves outside of the system in out-VM method.This is one of the prominent malware detection methods. This process is also used by CUCKOODROID. This technology detects the malware in android mobile devices [11]. It consists of an in-device agent which scans executable on the device and sends any suspicious executable to the remote scrutiny server. This scrutiny server runs hybrid of anomaly-based and signature-based malware detectors. This procedure first extracts the features static and dynamic analysis on malware apps. The obtained features are then used to train one-class SVM classifier for anomaly-based detection. On some specific platforms this CUCKOODROID achieved 98.8%of accuracy.

2.2 Malware Detection Technologies:

The malware detection technologies are the applications or software tools which are used to detect and analyze the malware presence. These technologies are widely used in this generation of technology. Due to the rapid increasing in the usage of technology, the malware attack presence is also increasing proportionally. The detection and prevention of this malware is essential. The following section in this paper introduces some of familiar malware detection technologies.

Imtithal A. Saeed, Ali Selamat and Ali M. A. Abuagoub[8] are presented some malware detection technologies. Host-based intrusion detection system is capable of monitoring dynamic behavior of the state of specific

compute system to see if there any external or internal activities that defraud the system. This type of detection system is also called as "in-the-box" detection system because they reside in the same host that they are monitoring.

Ali M. A. Abuagoub [8] introduces another Intrusion Detection System that is based on Network. It is referred as Network-based IDS. Thiskind of systems generally used to sniff all the packets on the network nodes for analysis. A single sniffer is placed in the network to monitor the traffic in the segment. In the distributed-network, there are multiple modules placed in each node to monitor the traffic in those nodes. This type of systems are also called as "out-of-the-box" detection systems because of they reside out of the host that they are monitoring.

W.Jin and D.Zhao [13] introduced a system which is the mixture of both host-based and host-based network capabilities.They are called "Hybrid-intrusion detection systems". This type of network detection system has multiple sub-systems that are located at different nodes in the network to monitor and gather the data from these nodes. This gathered data will be sent to the main system for analysis and classification. This author presented the drawbacks of both the systems. Host-based system protects effectively internal system but it is susceptible to external attack. Network-based can prevent external attack but it can't protect inside host.

Ye, D and Ou, C.-M.along with [14] and [15] C.R. Ou are clearly explains the Agent-based intrusion and malwaredetection system. These technologies depend up on characteristics of agent technology such as autonomy, decentralization, scalability, platform independency and mobility. Host-based IDS cannot detect outside attacks but it is effective internally. The distributed IDS do not take care of internal attacks, but it is effective externally. Agent-based system is designed such that it combines the characteristics of both Host-based system and Distributed system.

NwokediIdika, Aditya P. Mathur[5] proposed another IDS system, that is Application protocol based Intrusion detection systems (APIDS). This system focuses on monitoring and analysis on its specific protocol. By this system, the protocol will be analyzed as its behavior dynamically. The exact location of this APIDS in the network is between the webserver and database management system and monitors the SQL protocol specific to the middleware business logic when it interacts with the database.

III. CONCLUSION

By examining the various papers related to the malware detection methods and malware detection technologies, it is observable that there are many methods had been developed. Since there is a necessity of improving the security against the malware attacks, we believe that there should be some methods and technologies developed to meet the new generation's information security standards. In this way we say that, this analysis will help the researchers to develop a better and efficient methods and technologies to handle the latest and up-coming suspicious malware attacks.

REFERENCES

[1] Thu Yein Win, HuagloryTianfield, Quentin Mair "Detection of Malware and Kernel-level Rootkits in Cloud Computing Environments" IEEE, 2015.
[2] Michael R. Watson, Noor-ul-hassanShirazi, Angelos K. Marnerides, Andreas Mauthe and David Hutchison "Malware Detection in Cloud Computing Infrastructures" IEEE, 2015.
[3] Xiaoguang Han, Jigang Sun, Wu Qu3, Xuanxia Yao "Distributed Malware Detection based on Binary File Features in Cloud Computing Environment"IEEE, 2014.
[4] Angelos K. Marnerides, Michael R. Watson, NoorulhassanShirazi, Andreas Mauthe, and David Hutchison "Malware Analysis in Cloud Computing: Network and System Characteristics" 2013.
[5] NwokediIdika, Aditya P. Mathur "A Survey of Malware Detection Techniques" 2007.
[6] PranitGaikwad, Prof.DilipMotwani ,Prof.VinayakShinde "Survey on Malware Detection Techniques", IJMTER Volume-02 jan-2015.
[7] JyothiLandage and prof.M.P.Wankhade"Malware and Malware detection Techniques: A survey". IJERT, volume-2 issue 12,December -2013
[8] Imtithal A. Saeed, Ali Selamat, Ali M. A. Abuagoub "A Survey on Malware and Malware Detection Systems" IJCA, 2013.
[9] Vinod P., V.Laxmi, M.S.Gaur. "Survey on Malware Detection Methods",
[10] Elisabeth Hansson"Specification-Based Intrusion Detection for Mobile Ad Hoc Networks", Sweden
[11] Thu Yein Win, Member, IEEE, HuagloryTianfield, and Quentin Mair, "Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing" IEEE-2016.
[12] Erick Bauman, GbadeboAyoade and Zhiqian Lin "A Survey on Hypervisor-Based Monitoring: Approaches, Applications, and Evolutions", University of Texas at Dallas. ACM Computing Surveys, Vol. 48, No. 1, Article 10,Publicationdate: August 2015
[13] Gou, X., W. Jin, and D. Zhao "Multi-agent system for Worm Detection and Containment in Metropolitan Area Networks". Journal of Electronics (China), 2006. 23(2): p. 259-265.
[14] Ye, D.," An Agent-Based Framework for Distributed Intrusion Detections." 2009.
[15] Ou, C.-M. and C.R. Ou, "Agent-Based immunity for computer virus: abstraction from dendritic cell algorithm

with danger theory, in Proceedings of the 5th international conference on Advances in Grid and Pervasive Computing". 2010, Springer-Verlag: Hualien, Taiwan. p. 670-678.

[16] Safaa Salam Hatem, Maged H. wafy, Mahmoud M. El-Khouly "Malware Detection in Cloud Computing" IJACSA, 2014.

[17] Garfinkel, T. and M. Rosenblum, "A virtual machine introspection based architecture for intrusion detection". 2003: p. 191--206.