Protecting Big Data Environment from Security Attacks

Dr. G. NAGALAKSHMI[1] PROF& HEAD, VASANTHARAJ K[2] ASSOCIATE PROF,

1, 2 DEPARTMENT OF CSE

1,2 SIDDARTHA INSTITUTE OF SCIENCE & TECHNOLOGY,PUTTUR, ANDHRAPRADESH

[1]agnl.lakshmi@gmail.com, [2]vasanthgtec@gmail.com

Abstract:

The digital life of human being has changed one's daily activities which had resulted in a huge volume of data. This data, called Big Data, is used by many organizations to extract valuable information either to take marketing decisions, track specific behaviors or detect threat attacks. The term "big data" tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. The processing of such data is made possible by using multiple techniques, called Big Data Analytics, which allow getting enormous benefits by dealing with any massive volume of unstructured, structured and semi-structured content that is fast changing and impossible to process using conventional database techniques. But while Big Data represents an immense opportunity for many industries and decisions makers, it also represents a big security issues for many users. Big data security is a constant concern because Big Data deployments are valuable targets to would-be intruders. A single ransomware attack might leave your big data deployment subject to ransom demands..This paper shows the security issues in an environment and the detection of the harmful threats

Keywords: Big data, Security, Detection

I.INTRODUCTION

Emerging technologies such as smart grids, Internet of Things (IoT) and clouds generate huge amount of data. Several business models have been developed and innovative applications have proposed for making use of this data for improving the quality of life and providing better services to the customers. For example, business models have been developed for capturing the location and behavior of the users from their mobile devices and using this information for targeted advertisement and smart transportation. Utility providers are capturing power usage of the smart devices in real time to estimate the peak time demand for the generation of power and also offer variable pricing depending on the time of use. Although there are several advantages with such emerging technologies, there are significant challenges for securing such environments.

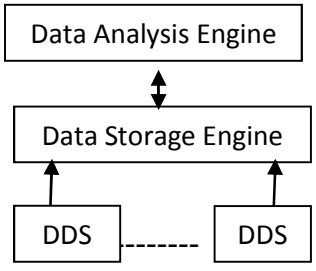

Fig.1 Storage System

Consider the Big data Scenario, a simple big data scenario which consists of capturing structured or unstructured data from several Diverse Data Sources (DDS) such as tiny sensors, servers, laptops, desktops, virtual machines, and smart phones, storing of the data in easily accessible

location (centralized or distributed) and analysis or further processing of data for different applications.

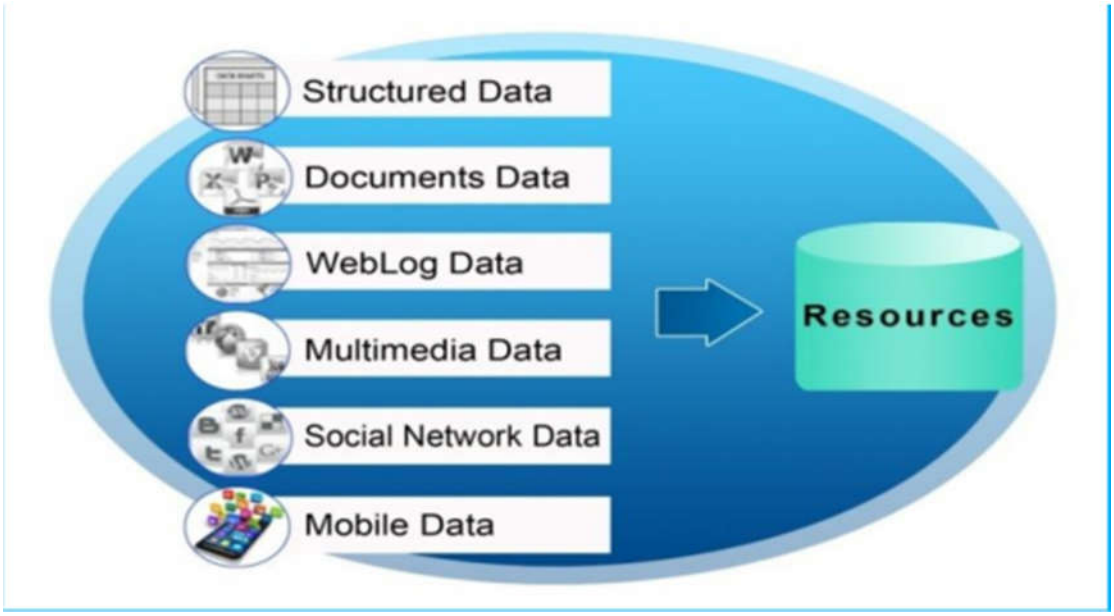

Fig 2 Structure of Big Data

However, since data is captured from untrusted devices, attackers or compromised devices can easily upload malicious data to the storage controller and the attacks can be spread to all other devices that access this malicious data. Also the volume, velocity and variety of the data generated in such environments make it extremely challenging to deal with the attacks in such environment. Hence there is need for techniques for securing such big data environments.

The paper is organized as follows. Section II presents the attacker model and overview of the operation of our model. Section III presents detail discussion on the components and detection of our model. Section IV present the detection of the attack and how it helps to deal with different attacks or how to remove the attacks. Section V concludes.

As shown in Fig 2, our model makes use of Trusted Components (TC) for enforcing service specific security policies on the DDS and also for capturing the data required for security analysis. The TC are placed at different devices in secure locations and the policies enforced in the TC depends on the capabilities of the devices. For example, the TC can be placed in the gateways, access points, base stations and virtual machine monitors.

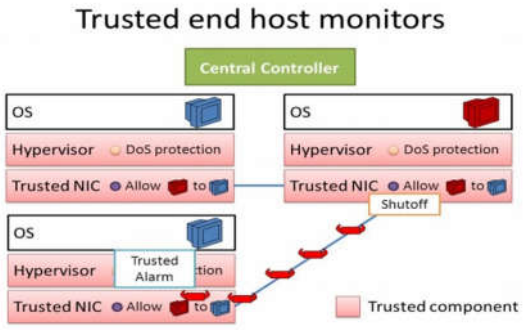

Fig 2 Trusted Components

Also let us consider a simple cloud scenario and discuss an attacker model and operation of our model. For example, in the below Fig 3 case of cloud there can be several millions of devices (volume) that are uploading/downloading data in frequent intervals (velocity) and different types of data (variety) such as tenants using the cloud for different services such as critical infrastructure, health care and utility providers.

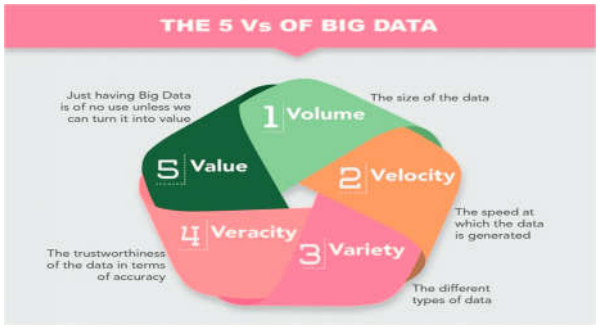

Fig 3. 5Vs of Big data

II LIST OF HARMFUL ATTACKERS

Malware

The word Malware is used for malicious software and is a general term for Virus, worm, trojan, rootkit, spyware and nearly everything which is specifically designed to harm your computer and steal information. Malware does not include buggy software, programs you don't like, software which crashes a lot but software which are specifically created to harm our PC.

Virus

The most common word used for any bad software, however, we use the term "malware" now. A virus is a program which can self-replicate itself after infecting a computer, it attach itself to other programs and get installed while installing the genuine software. After the execution of viral code it may destroy host files and starts infecting files into a PC, from there it creates a replica of itself and travels from PC to PC via external drive, the Internet, and malicious websites.

Worm

Worms are similar to the Virus, but a worm doesn't need a host program to execute itself, worms are the standalone program. It uses a computer network to spread itself, and it relies on network loophole and security failure to travel from one host to another automatically and usually don't requires user intervention. Since worms don't require any initiation they can spread speedily across the network, infecting every PC in their path, Worms are most well know type of malware which affects more computers than a virus can.

Trojan Horse

Trojan is another kind of malware which looks harmless, but it contains malicious code which creates a backdoor that allows your PC to be controlled remotely. The term Trojan horse came into existence from the Troy Story where the Greeks used a wooden horse to infiltrate Troy.

Spyware

Spyware is another type of malware which collects data from your PC without your knowledge and permission, spyware runs in the background and collects your personal information like your browsing pattern, sites you visit, email, cookies, saved data into browsers, website passwords and even credit card details.

Adware

Adware is a bit different from spyware the primary intent of adware is to show different advertisements, pop-up window, flash ads, links to rogue websites, redirecting to different links, change homepage and default search engine, slows down browsing speed, causes frequent browser crash.

Rootkit

A root kit is a software or set of application typically malicious that enables administrator-level access to a computer or computer network. Rootkit get activated every time you boot into operating system since they activated before an operating system gets completely booted up which makes it very hard to detect by antivirus.

Root kit can get to a computer by a Trojan, suspicious email attachments or by compromised websites after getting user level access to the system either by breaking a password or by exploiting any vulnerability into the system. Once a rootkit is installed it allows the installation of hidden files, hidden user accounts, processes and attackers can mask intrusion and get root access to the system.

A root kit can monitor traffic, keystrokes, can create a backdoor for malicious usages by hackers, and it can remove installed programs and security suits in order to prevent the detection.

Bots

The bot is a short name for "robot" and is an automated process/script which interacts with other computer or network services; bot is a software program which automates different tasks over the Internet using specially written scripts.

Let us consider a generic big data scenario such as public cloud with different tenants (utility, healthcare, finance, governments) making use of IaaS public cloud for hosting their services. The tenants can be running different operating systems (such as Windows, Linux) as shown in Fig 2 and service specific applications in their virtual machines.

Attacks in such environments can lead to catastrophic damages (blackouts in case of attacks on utility services) and in some cases loss of life (eg. doctors unable to access patient's data). There are several challenges to deal with the attacks in such environments. On one hand there are attacks that target specific services (such as Stuxnet for SCADA) of the tenants and on other hand there are some attacks such as botnet and Rootkit that are common for any of the tenants since their applications are running on popular OS such as Windows and Linux. If the attacker can exploit vulnerabilities in such OS then it can compromise different tenant services. Attacks such as botnet or Rootkit are practical in the current state of art.

III.ROOT KIT ATTACKER MODEL

Root kit malware

A root kit is a particularly nasty piece of malware that doesn't behave like your typical virus. Root kits insert themselves into the very heart of the operating system; usually at or below the kernel level. This makes them extremely difficult to detect and sometimes impossible to remove. Specific antivirus programs specialize in the detection and removal of root kits.

A root kit is a collection of computer software, typically malicious, designed to enable access to a computer or areas of its software that is not otherwise allowed (for example, to an unauthorized user) and often masks its existence or the existence of other software. The term root kit is a concatenation of "root" (the traditional name of the privileged account on Unix-like operating systems) and the word "kit" (which refers to the software components that implement the tool). The term "root kit" has negative connotations through its association with malware

Root kit installation can be automated, or an attacker can install it after having obtained root or Administrator access. Obtaining this access is a result of direct attack on a system, i.e. exploiting a known vulnerability (such as privilege escalation) or a password (obtained by cracking or social engineering tactics like "phishing"). Once installed, it becomes possible to hide the intrusion as well as to maintain privileged access. The key is the root or administrator access. Full control over a system means that existing software can be modified, including software that might otherwise be used to detect or circumvent it.

Root kit detection is difficult because a root kit may be able to subvert the software that is intended to find it. Detection methods include using an alternative and trusted operating system, behavioural-based methods, signature scanning, difference scanning, and memory dump analysis. Removal can be complicated or practically impossible, especially in cases where the root kit resides in the kernel; reinstallation of the operating system may be the only available

solution to the problem. When dealing with firmware root kits, removal may require hardware replacement, or specialized equipment.

IV DETECTION OF ATTACKER MODEL

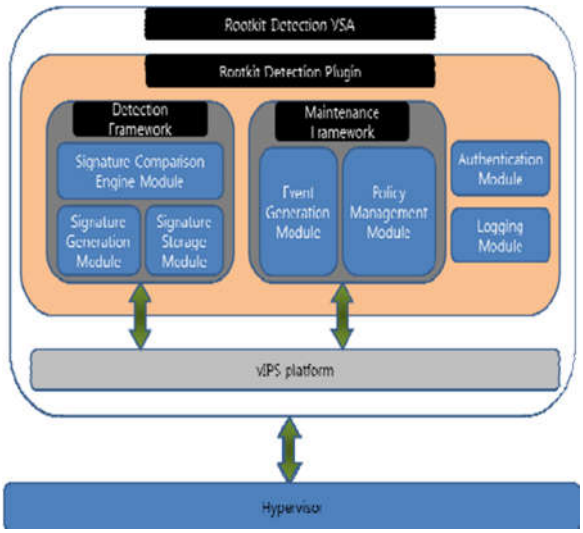

Fig 4 Root kit detection model

In this section, we discuss the key requirements for designing our root kit detection system. These requirements specify the constraints that any root kit detection system must hold to provide optimized performance in cloud computing environments. The root kit detection system needs to be implemented as an agent less security virtual appliance. This structure should be able to access the internal states of VMs through a hypervisor API call or similar libraries, while staying out of the virtual machines in question.

This observation and analysis of the internal states and events of VMs including the contents of virtual CPU, memory, and disk is called VM introspection. Even though VM introspection is the outside observation, it can build an almost same semantic view of system states and events as a semantic view obtained inside VM. Thus, VM introspection is critical to support tamper-resistant, high-fidelity out-of-the-box VM monitoring, resulting in the basis of intrusion detection/prevention. Recent malware is getting gradually more stealthy and elusive. They are trying to detect and compromise even anti- malware software located in the compromised system as well as to hide their own presence from intrusion detection in the system. Many out-of-the-box based approaches have been recently proposed. They place their detection facility in an independent VM without locating an agent inside VM monitored. Thus, they enable the detection facility to be isolated from the monitored VM, making it hard for malware to sense and subvert the detection facility itself. VM introspection is a crucial technique for agent less virtual security appliance. The next upcoming section discusses the necessity of the removal of the root kit, or it can be eliminated.

Remember, the root kit attack to any particular should be a medium one. Very high sensitive malicious software or the environment is affected by Root kit is difficult to detect as well as to eliminate.

A Common Detection of Rootkit:

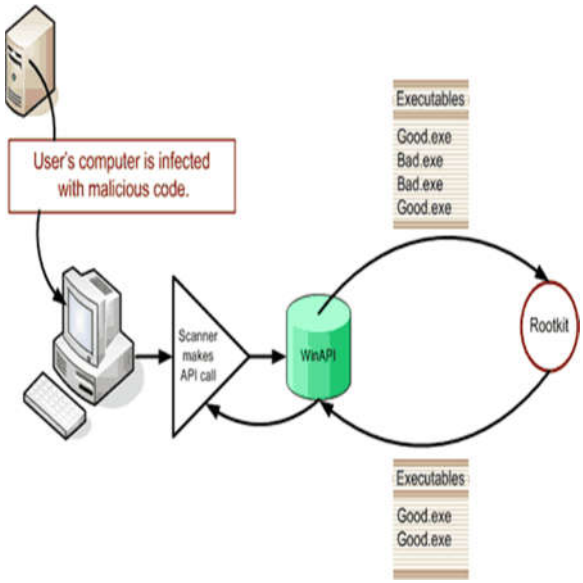

Fig 5. Root kit attacker model

Elimination:

The elimination of a “Root kit " is varied, depending on the type of rootkit. For example; When the infection is directed to the kernel system, then the only option left is to reinstall the infected system. There are cases in which replacing hardware is the only option, given the severity of the attack.

Types of root kits

User mode: These are the low level root kits that are executed together with the other processes. These root kits have the ability to inject anything into memory. That is to say; using these techniques they create new DLL files and dylib files on Windows and on Mac, respectively. They are also capable of intercepting and modifying the behavior of the application programming interface.

We must remember that user-mode applications run in their own memory space, so de facto root kits scale privileges to super user to carry out this allocation in the memory space of each application that runs on each active user of the user.

In addition, root kits need to monitor the system for new applications that are running and anticipate the vulnerability correction patches on the memory space of those programs when they are executed in their entirety.

In kernel mode: These rootkits affect the core part of the operating system.Writing these rootkits is very difficult and their detection too. This, because they run on the same security level of the core space of the operating system. Its effects on system stability and even on antivirus applications make the infected device vulnerable.

Most systems support device drivers in kernel mode, that is; that run in the same kernel. Thus rootkit in kernel mode can modify the data structure of the operating system using the method called: Direct Manipulation of Kernel Objects *(* DKOM: Direct Kernel Object Manipulation *).*

The above also messes up the System Descriptor Table *(* SSDT: System Service Descriptor Table *)* or modifies the doors between user mode and kernel mode to hide.In operating system based on Linux modify table calls to the system. They also create encrypted disks to hide the original copies of the infected items.

Boot Kit: The Boot kit affects the Master Boot Record *(* MBR *)* and the boot volume register *(* VBR: Volume Boot Record *).*This is executed at any time when the system is powered on, so reinstalling the system does not always work because Basic Input and Output System *(* BIOS *)* files may be compromised .

V CONCLUSION

Big Data serves as a good basis for many organizations and governments in different sectors that intend to automatically process and extract valuable insights in order to help decision makes. However, the fact to collect and compute all possible and varied data could lead to many security and privacy violations. In this paper, we have highlighted a set of security and privacy challenges that should be considered by big data tools. But when speaking about Rootkit threat it is always better to say "Prevention is better than cure". Detecting the Rootkit malware is a very big challenge. As a future work, we intend to implement some of the protection techniques may be implemented for eliminating the rootkit in big data environment.

REFERENCES :

[1] Thu Yein Win, Member, IEEE, Huaglory Tianfield, and Quentin Mair, "Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing", In Proceedings of the DOI 10.1109/TBDATA.2017.2715335, IEEE Transactions on Big Data

[2] Youssef Gahi, Mouhcine Guennoun, Hussein T. Mouftah, "Big Data Analytics: Security and Privacy Challenges", In Proceedings of the 2016 IEEE Symposium on Computers and Communication (ISCC).

[3] B. Saraladevia, N. Pazhanirajaa, P. Victer Paula, M.S. Saleem Bashab, P. Dhavachelvanc, "Big Data and Hadoop-A Study in Security Perspective", In Proceedings of the 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15).

[4] Dušan Mondek, Rudolf B. Blažek, and Tomáš Zahradnický, "Security analytics in the Big Data Era", In Proceedings of the 2017 IEEE International Conference on Software Quality, Reliability and Security (Companion Volume)

[5] Udaya Tupakula Vijay Varadharajan, "Securing Big Data Environments from Attacks", In Proceedings of the 2016 IEEE 2nd International Conference on Big Data Security on Cloud,

IEEE International Conference on High Performance and Smart Computing, IEEE International Conference on Intelligent Data and Security

[6] A. Cuzzocrea, I-Y. Song, and K. C. Davis, "Analytics over large-scalemultidimensional data: the big data revolution!,", In Proceedings of theACM 14th international workshop on Data Warehousing and OLAP, pp.101-104. 2011.

[7] Elisa Bertino, "Big Data – Security and Privacy,", In Proceedings of the 2015 IEEE International Conference on Big Data .

[8] Boel Nelson, Tomas Olovsson, "Security and Privacy for Big Data: A Systematic Literature Review,", In Proceedings of the 2016 IEEE International Conference on Big Data .

[9] Duygu Sinanc Terzi, Ramazan Terzi,Serif Sagiroglu, "A Survey on Security and Privacy Issues in Big Data,", The 10th International Conference for Internet Technology and Secured Transactions (ICITST-2015)