Cloud VM Security with Machine learning based Agent

Dr.K.Venkataramana, Prof.M.Padmavathamma

Abstract—Cloud computing is an effective way of utilizing and managing IT resources and services at low cost, which is available but with few challenges to security which is to be addressed succinctly. Since virtualization forms the base of cloud computing one should ensure security of virtual-machine images that encapsulate each application of the cloud. VM Security should be introspected on process that runs in it, should ensure isolation and should not effect the process running in others VM's and also in the same VM which may be accessing internet or web service. So to avoid this security path every VM will have an INTELLIGENT AGENT as a process or a thread that does not allow the malwares (computer viruses, worms, Trojan horses, spyware, dishonest adware, scareware, crimeware, most rootkits, and other malicious and unwanted software or program.) to run on the virtual environment of a cloud by comparing the features of the applications that are ready to run on VM with the built-in application features stored in the repository of VM.

Index Terms— cloud computing, Service models, Virtual machine,security,Agent,Machine learning ,Intelligent Agent.

1. INTRODUCTION

Innovation is regular activity in competitive world of Information Technology which created technologies like grid computing, distributed computing, Cloud computing etc., Cloud computing is a TCP/IP based computer technology which is integration fast microprocessor, huge memory, high-speed network and reliable system architecture. Cloud computing, is a new paradigm of distributed computing, introduces many new ideas, concepts, technologies and architectural styles into enterprise service-oriented computing. A cloud is pay-per-resources business model may be thought of a large pool of resources unified through virtualization and are charged as per usage and can be managed dynamically scale up or down to match the load. Cloud computing relies on technologies like Web Services, Virtualization, Utility computing promotes a model for providing IT capacities over the Internet as services and on a lease based and on-demand style.

Cloud computing is typically a three layered stack, with each layer providing its own services and is utilized by upper layer, as illustrated in Figure 1. The Cloud Infrastructure Service or the Infrastructure as a Service (IaaS) provides IT infrastructures as a service over computer networks. The Cloud Platform Service or the Platform as a Service (PaaS) delivers computing platforms as a service to sustain the cloud applications. The Cloud Application Services or Software as a Service (SaaS) delivers software as a service over the network, allowing users to use applications without having to install and run software on their own computers. Cloud deployment models include public, private, community, and hybrid clouds. Public clouds are external or publicly available cloud environments that are accessible to multiple tenants, whereas private clouds are typically personalized environments with dedicated virtualized resources for particular organizations. Similarly, community clouds are designed for particular groups of customers.

With Cloud computing there is a substantial reduction in Operational and Capital expenditures for organizations, also services in cloud provide users with scalable resources in the pay-as-you-go fashion at relatively low costs. For example, Amazon's EC2 sells 1.0-GHz x86 ISA 'slices' for $0.10

- Dr.K.Venkataramana,Assoc.Prof,
  Dept of MCA ,KMMIPS,Tirupati
  ramanakv4@gmail.com

- Prof.M.Padmavathamma,Dept of
  Comp.Science,S.V.University,Tirupati,
  prof.padma@yahoo.com

per hour, and a new 'slice', or instance, can be added in 2 to 5 minutes. Amazon's S3 charges $0.12 to $0.15 per giga byte month, with additional bandwidth charges of $0.10 to $0.15 per gigabyte to move data into and out of Amazon Web Services over the Internet. Comparing with building and managing their own infrastructures, users are able to save their investments significantly by migrating businesses to a cloud. With the increasing development of cost-effective cloud computing technologies, in future more organizations move their businesses to cloud for better service at cheaper costs.

Many companies, such as Amazon, Google, Microsoft and so on, moving their business in developing Cloud Computing systems and enhancing its services providing to a larger amount of users. The successes of the above companies, say Google, Amazon and so on, are great examples and encourage an amount of other companies to step into the Cloud, such as Media Temple, Mosso, Joyent, Flexicale, and so on. Lots of services, such DaaS, SaaS, PaaS, IaaS, etc. are provided to users on pay-per-usage basis. Major Cloud Service competitors in the market are Google Apps Engine[1] or Microsoft Azure Platform is a PaaS, while Google Docs[2] is a SaaS, and DropBox [3] is an IaaS .

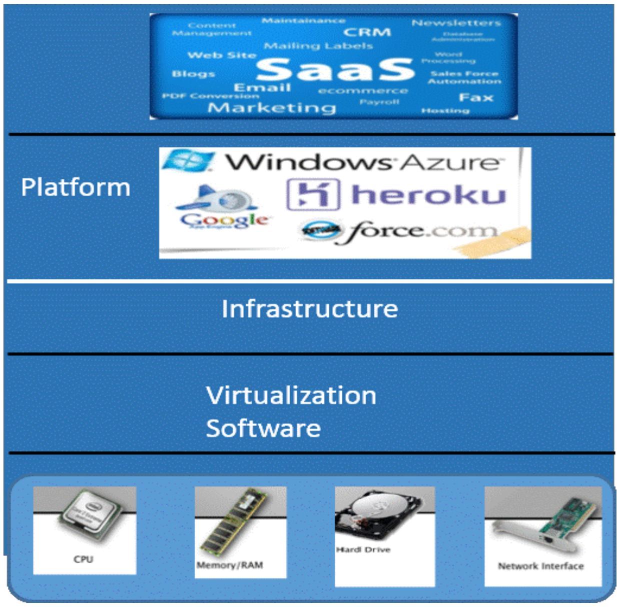

Fig-1 Cloud Model

For adoption of cloud there should various issues to be addressed one among them is the security and privacy to the data stored in cloud. The inherent challenge is to how to ensure data privacy, as data being present in unencrypted form on a machine owned operated by a different organization from the data owner. There are threats of unauthorized uses of the data by service providers and of theft of data from machines in the cloud. Fears of leakage of sensitive data or loss of privacy are a significant barrier to the adoption of cloud services. Security concerns exist for both Service Oriented Computing and cloud computing. In SOC, services underlying a composite service may originate from different providers across multiple organizations. It's difficult to ensure an end-to-end security solution, as it would require all service providers to guarantee the same level of security guarantee. This is compounded when services reside in cloud computing environments and if the customer uses services at various layers. The underlying infrastructure of the service providers reside with other third-party providers, and as such, there are significant negotiations required between end users, cloud service consumers, and cloud providers to define a certain level of security. Since cloud provides services at SaaS, IaaS and PaaS levels, ensuring security is a difficult task, so in this paper agent based security model is proposed.

2. RELATED WORK

In cloud different services will be running at all the layers each should be protected against malicious software's like Malwares, Trojans, virus or rootkits etc., which have different characteristics to be detected and removed either by using a protective services or agents.

2.1 Malwares

Malware, is a malicious software, consists of programming (code, scripts, active content, and other software) designed to disrupt or deny operation, gather information that leads to loss of privacy or exploitation, gain unauthorized access to system resources, and other abusive behavior. The expression is a general term used by computer professionals to mean a variety of forms of hostile, intrusive, or annoying software or program code. Software is considered to be malware based on the perceived intent of the creator rather than any particular features [7].

Malware includes:
2.1 Computer viruses & Worms
2.2 Trojan horses
2.3 Spyware, Dishonest adware
2.4 Rootkits and Data stealing malware

2.1 Computer viruses & Worms

A computer worm is a self-replicating malware computer program, which uses a computer network to send copies of itself to other nodes (computers on the network) and it may do so without any user intervention. This is due to security shortcomings on the target computer. Unlike a computer virus, it does not need to attach itself to an existing program. Worms almost always cause at least some harm to the network, even if only by consuming bandwidth, whereas viruses almost always corrupt or modify files on a targeted computer. Worms sometimes infect these file types (EXE, COM, DLL) Some Worms are coded by Machine Language (Low Level Language).

2.2 Trojan Horses

A Trojan horse, or Trojan, is a destructive program that masquerades as an application. The software initially appears to perform a desirable function for the user prior to installation and/or execution, but (perhaps in addition to the expected function) steals information or harms the system. Unlike viruses or worms, Trojan horses do not replicate themselves, but they can be just as destructive. The term is derived from the Greek myth of the Trojan War, in which the Greeks gave a giant wooden horse to their foes, the Trojans, ostensibly as a peace offering. However, after the Trojans dragged the horse inside their city walls, the Greek soldiers sneaked out of the horse's hollow belly to open the city gates and allowed their compatriots to pour in, capturing Troy.

2.3 Spyware, dishonest adware

Spyware is a type of malware that can be installed on computers, and which collects small pieces of information about users without their knowledge. The presence of spyware is typically hidden from the user, and can be difficult to detect. Typically, spyware is secretly installed on the user's personal computer. Sometimes, however, spywares such as key loggers are installed by the owner of a shared, corporate, or public computer on purpose in order to secretly monitor other users.

Adware, or advertising-supported software, is any software package which automatically plays, displays, or downloads advertisements to a computer. These advertisements can be in the form of a pop-up. They may also be in the user interface of the software or on a screen presented to the user during the installation process. The object of the Adware is to generate revenue for its author. Adware, by itself, is harmless; however, some adware may come with integrated spyware such as key loggers and other privacy-invasive software.

2.4 Rootkits and Data stealing malware

A rootkit is software that enables continued privileged access to a computer while actively hiding its presence from administrators by subverting standard operating system functionality or other applications. The term rootkit is a concatenation of "root" (the traditional name of the privileged account on Unix operating systems) and the word "kit" (which refers to the software components that implement the tool). The term "rootkit" has negative connotations through its association with malware.

Data-stealing malware is a web threat that divest victims of personal and proprietary information with the intent of monetizing stolen data through direct use or underground distribution.

Characteristics of data-stealing malware

1. The malware is typically stored in a cache that is routinely flushed
2. The malware may be installed via a drive-by-download process

2 It is difficult for antivirus software to detect final payload attributes due to the combination(s) of malware components
3 The malware uses multiple file encryption levels
4 Thwarts Intrusion Detection Systems (IDS) after successful installation
5 There are no perceivable network anomalies
6 The malware hides in web traffic
7 The malware is stealthier in terms of traffic and resource use
8 Thwarts disk encryption
9 Data is stolen during decryption and display
10 The malware can record keystrokes, passwords, and screenshots

Cyber-attacks targeted at virtualization infrastructure underlying cloud computing services includes malwares, spyware or rootkits has become sophisticated which can withstand the cryptographic alternatives. In [6] discussed about a novel malware and rookit detection system which protects the guests against different attacks. It combines system call monitoring and system call hashing on the guest kernel together with Support Vector Machines (SVM)-based external monitoring on the host.

In paper [8], has discussed various security issues at various service models like Data security, Network security, Data locality, Data integrity, Data segregation, Data access, Authentication and authorization. Cloud computing has significant implications for the privacy of personal information as well as for the confidentiality of business and governmental information. The confidentiality of sensitive data must be protected from mixing with network traffic with other cloud hosts. If the data is shared between multiple users or clouds , the CSP must ensure data integrity and consistency. The CSP must also protect all of its cloud service consumers from malicious activities or data modification [9-10].

3. INTELLIGENT AGENT

Intelligent agents[4,5] is defined as self-designed software programs that are capable of learning and communicating with various sources of information in a distributed system to carry out task at regular schedule. These information sources can be very diverse: real-time information systems, databases containing historical information, simulation frameworks, human operators, external agent systems, etc. Typically, an agent program, using parameters you have provided, searches all or some part of the Internet, gathers information you're interested in, and presents it to you on a daily or other periodic basis. An agent is sometimes called a bot (short for robot).

The working of intelligent agent is given as:

(1) Initialize the environment and a creation of agents,

(2) Each agent observes or learns its environment, and

(3) Each agent takes an action based on the current observations and classify the process based on the learned data and current data as a malicious or not.

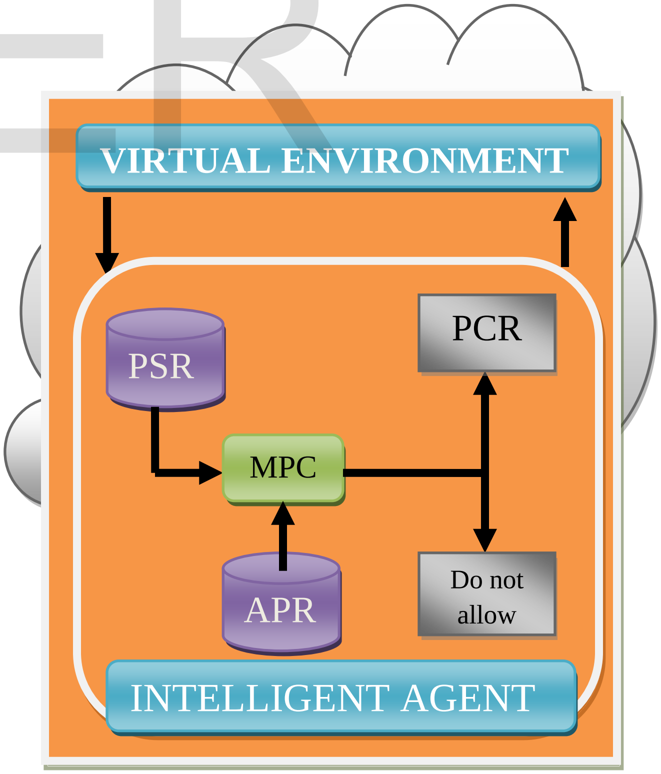

Fig: 2 Intelligent Agent Architecture

3.1 IA Architecture.

The above architecture defines the functionality of an intelligent agent in the Virtual Environment. The agent has basically two databases, one register, and a process to compare the malware properties.

3.1.1 PSR

The PSR (Process State Repository) stores the active processes that are to run on the Virtual Machine. These processes are stored in the sequence and based on the priority, each process is sent to the MPC.

3.1.2 MPC

MPC (Malware Property Comparator) is a machine learning process that compares the properties of each active process received from the PSR with the properties of the application that are already stored in the APR. Machine Learning (ML) process will store and learn the malware process properties from APR. The data stored by MPC will be treated as experience data by which ML process will use to classify the process.

3.1.3 APR

The Application Properties Repository (APR) consists of the properties of a unique process that runs on a Virtual Machine. Among them the Malware Properties are also stored. The MPC will classify malware with ML by using the data from the APR and compares with the properties of the each active process. If the malware properties match with the active process is not allowed to run on the Virtual Machine. In the either case, the active process is stored in the PCR.

3.1.4 PCR

Platform Configuration Register (PCR) receives the secured active processes and stores them in the sequence. Then the each process in the PCR is sent to the VM for the safety run in the user's environment if it is classified as a non-infective process or a secure process.

3.2 Security Analysis

In this model of security, machine learning mechanism is used to identify the process as a malware or any other malicious process which may pose problems to virtual machine. So the process should be terminated, so to identify it as malicious or not, machine learning algorithm Naïve Bayes classification is applied as the malwares are of various types and have more attributes to classify them we apply Naïve Bayes classification technique to find and thus agent will report the process to VM Manager and VM will be secured. The data set stored in PSR and APR are used. IA will be created by a secure process in each VM at regular intervals and exist for certain period of time to check and verify the current status of VM and its process by running Machine Language algorithm and terminates the process which are malicious or harmful to the VM. IA will not only secures IaaS but also PaaS by allowing IA to be created in any platform to secure the applications running. Since PaaS runs in IaaS, secure features of IaaS can be used to protect PaaS and SaaS.

4. Conclusion

In the above proposed model we have tried to show malware and their characteristics, which can access shared memory of other users virtual machines causing data leakage. So by using IA which is a part of VM will avoid such programs to run in VM and thus provides security to VM in cloud (Iaas). The results of the above model will be published in subsequent papers.

REFERENCES

[1]. http://www.google.com/apps
[2]. http://docs.google.com
[3]. http://www.dropbox.com
[4] Murch R, Johnson T. Intelligent Software Agents. NewJersey, USA: Prentice Hall PTR, 1998.
[5] Jennings N R, Wooldridge . Applying agent technology.Applied Artificial Intelligence, 1995, 13(6): 357-370.
[6] Thu Yein Win; Huaglory Tianfield; Quentin MairDetection of Malware andKernel-Level Rootkits in Cloud Computing Environments, 2015 IEEE 2nd International Conference on Cyber

Security and Cloud Computing Year: 2015

[7]. K. Venkataramana, A Smart Agent Based Security for VM in Cloud, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 5, Issue 9, September 2015

[8] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing",Journal of Network and Computer Applications (2011),pp. 1-11. doi: 10.1016/j.jnca.2010.07.006

[9] Federated identity management",[Online] [Available] http://en.wikipedia.org/wiki/Federated_ide tity_management

[10] Xiao Zhang; Hong-tao Du; Jian-quan Chen; Yi Lin; Lei-jie Zeng,"Ensure Data Security in Cloud Storage",IEEE