Malware Detection and Prevention on Cloud

Pawan Jaybhaye[#1], Dr. Bandu B. Meshram[*2]

[#1,*2] *Veermata Jijabai Technological Institute(VJTI), Student, Department of Computer Engineering*

Abstract *— Security on cloud is one of the concerns of this growing technology in today's internet world. More and more organizations, enterprises and institutions are moving towards cloud technology because of affordable, efficient, and computing on demand like features. In recent years malware has became one of the most dangerous threat to the cloud service providers and their users and that threat is growing day by day. In this paper we propose a malware detection and prevention system on cloud based on signatures md5, sha1 etc and patterns of various families of existing malware. In this paper we will use cloud services for deployment of cloud based antivirus and hash signatures and patterns will be used in intrusion detection system. We also propose a system for real time analysis on cloud for malware detection and prevention.*

Keywords *— Cloud, Malware, Hash, Real Time Analysis, Signatures, MD5, SHA1.*

I. INTRODUCTION

In todays world Cloud computing has became an emerging trend. It is a technology that enables large-scale distributed computing. Many firms are now shifting towards Cloud computing because it provides an environment which is ready to use and due to this cloud technology has become the fastest growing technology in the IT world. It has various sources in order to enable resource sharing in terms of scalability, well managed computing services that are delivered based on demand over the network. The main advantage of cloud computing is the users does not have to set up their own infrastructure which is very costly instead they uses ready to go infrastructure.

Now for past few decades, the computational world has gone from centralized to distributed systems and with the new inventions in the technology everything is going towards virtual technology and cloud computing is the enabler of the virtual technology. There are many definitions of cloud computing given by various people and organizations and all are different because cloud computing is such vast technology which is very hard to define. One of the major institute NIST defines cloud computing as pack of five essential features, four deployment models and three service models. The five essential features are resource pooling, broad network access, rapid elasticity and scalability, metered services, on demand service. The three service models are Software as a Service, Platform as a Service, Infrastructure as a service, the four deployment models are private cloud, public cloud, hybrid cloud and community cloud, [4] As the cloud technology grows the security concerns of cloud also grows and at this point of time when cyber attacks and the malwares has grown exponentially it is very important to protect cloud from various malware attacks.

Malware is nothing but a malicious software. It is the type of software that is knowingly designed with a harmful intent in mind. There are different types of malwares such as Worms, Trojan horses, Viruses, Backdoors, Spyware, Ransomware, Rootkits, botnet in addition to other types of software with unwanted behavior.

With the rapid development of the Internet, malware became one of the major cyber threats nowadays. Any program that performs malicious actions, including information stealing, spying etc. is a malware. Few of the definitions of malware family is given below:

Virus: It is the type of program that replicate itself on the host machine and connecting to documents which becomes their carriers.

Worm: It is a program that has similar characteristics as of viruses but instead of affecting the host machine they affect the network.

Trojan horses: It is very unique type of malware as at first glance it seems to be the useful code, but it also contains the harmful code which is hidden and runs itself when that program is executed.

Backdoors: These are the loopholes that are exploited in the program by the attacker and then the attacker creates these loopholes to get or steal information from the victim.

Adware: It is not the harmful code but it slows down the working of host machine by continuously showing the adds which leads to the harmful pages or sources.

Rootkits: These are very advanced malwares as they directly deal at kernel level which is very dangerous as that may crash the entire machine. They causes harm to the infrastructure of the victim directly. [18]

Ransomware: It is a new type of malicious code which encrypts all the user data and then demands ransom to decrypt the useful data of user.

In cloud computing we have three modules and our Malware detection and Prevention System (MDPS) works on these modules VM, Database and Networks for various attacks.

A) Malware Injection:- In these attack mainly servers are targeted and malicious code is dumped on them so when any data is send from that server it

automatically has malicious code embedded in it which will affect all the clients of that server

B) MITM:- These types of attacks are mainly used for stealing data from the user.

C) DDoS:- These type of attacks are done to take down the entire network by jamming the network traffic with useless data traffic. [1]

MDPS is a system which detects malware using signatures and other heuristics techniques or rule based or string based pattern matching, Antivirus are also an example of a malware detector, Many malware writers know these detection methods used by the antivirus programs so they keep on developing new ways to evade or bypass these detection techniques my modifying the malware code or put some junk data so the hash of that file changes and then it becomes undetectable. They also use encryption tools or password protected methods to escape from detection. Malware detection techniques generally takes two inputs for detection:

• Malware signature or behavior or rules from the database.

• The program which is to be checked for malicious intent.[18]

MDPS also use the real time malware analysis for higher security on cloud. This real time malware prevention technique is very important to deal with everyday growing malwares as it protects the user from the unknown malwares and attacks which may affect user and protects host machine from getting compromised.

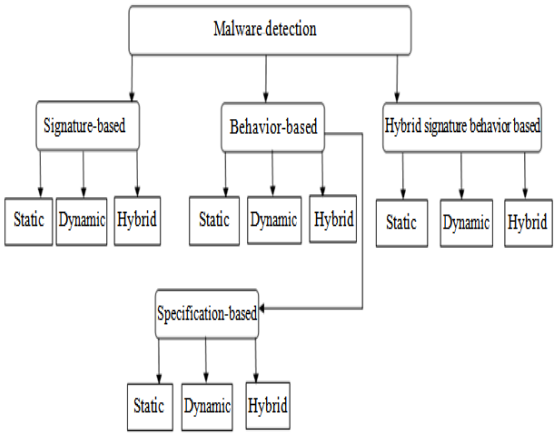

FIGURE 1.1 OVERVIEW OF DETECTION

II. LITERATURE SURVEY

A) General Malware Detection Approach

For the basic malware detection mainly two methods are used. One is shallow analysis and other one is deep analysis. In shallow analysis method the processes parameters are checked on the victim machine. These parameters are checked before and after the execution of malware to determine which processes are initiated by the malware and the effect of them on the machine. These are the features of compromised machine by malware:

• Change or update in Windows Registry Entries or keys

• Unexpected increase in the number of processes executing on the system.

• File Creation / Deletion/Modification

During shallow analysis the above information is used to build a profile of the machine. Various snapshot of the system were taken before and after analysis and then two snapshots are analyzed to determine if any unexpected changes occurred to detect malware. The following system parameters are considered for analysis in shallow analysis:

- CPU usage and speed
- Memory usage
- No of users
- No. of Processes and process state

The other method is deep analysis. In this method hash of the file is calculated, file is checked for strings and malware patterns for the accurate detection of malwares.

B) Signature based detection

Signature based malware detection attempts to define a set of hash file that can be used to decide that a given pattern is that of the malware. Due to this property signature based system can gain very high speed of detection as size is less, accuracy and minimal number of false positives in identifying intrusions on the system. The main drawback of this system is that if there is a small modification in the file the entire signature changes and the signature based detection proves to be inefficient. Therefore, signature based detection sometimes might not be able to detect unknown attacks or might give a false positive results. But considering the above drawback still signature based detection is used because of its ease in maintaining and updating Signature. These signatures are composed by several elements that identify the traffic. Various parts of a signature are the header (e.g. source address, destination address, ports) and its options (e.g. payload, metadata), which are used to determine whether or not the network traffic corresponds to a known signature. In Cloud platforms, malware signature based detection methods are used to detect the samples that are already present in the database. These detection methods can be used at gateway to Cloud to detect external intrusions inside after the firewall of Cloud to detect external/internal intrusions. [3]

Signature Optimizing Pattern Matching: These methods are dependent on the signatures, that are stored in the database. In this method for detecting intrusion, we use string matching algorithm, This is mostly common used in finding the DNA and various sequences of proteins. The signature based detection

is very important part of our malware detection process as it provides basic ground level for virus detection and with already creating a database of viruses the system provides security against almost all the known malware types. If unknown malware samples were detected is is very important to add them to the database. Based on the virus function of self-replication and seasoning, they proposed optimizing policy which mainly focus on Signature database, all the virus that are present use the simple method in which they will scan the victim file system and inject the malicious code into the normal files. Now this leads to the creation of lots of replicas that coexist within one system. So when any virus is detected by signature match, this virus signature is temporarily stored, so the other replicas do not need to match against the other large amount of signatures in the actual signature database. So this pre-comparison method will reduce the signature matching times.

C) **Automatic Signature Extraction**

Traditional signatures extraction methods mostly relied on the manual extraction of a sequence of bytes from a malware by an expert. The byte sequence is generally present in the executable part of the unknown file and it is most unlikely that these type of string bytes found in normal files. This process is done manually and in order to detect those string of bytes which makes signature extraction very tedious and time consuming. Therefore, a system is needed that will automatically extracts signature from a malware sample. In principle, any byte sequence from the executable part of a malware sample can be used as a signature for detecting that very malware. Sometimes the malware signatures that are generated may match some benign contents and as a result causes high false positive rates. So reducing or avoiding false positive rates is a major consideration for extracting malware signatures. Another important factor is time consideration here is the time needed for detecting malware among the network traffic. For reducing the scanning time for malware signatures, they generally need to limit the length of a signature. Also scanning time is reduced by reducing the malware signature and only those that can detect most of the malware samples. In this it is observed that many malware samples consists of same common executable parts. Since the task here is to minimize the total number of signatures and make signatures more efficient, so basically in this researchers have to find a minimum cover set of the signatures that detects all the samples effectively. This will lead to increase in scanning speed, these signatures are more likely to detect the existing variants of malware. But, it is very hard to minimize both the scanning time and the false positive rate and still maintain the efficiency of the program at the same time and automatically determining the optimized set of malware signatures.[5]

D) **Anomaly-based intrusion detection**

Anomaly is also called behavioural based detection of malwares. In this method mainly the events are monitored on regular intervals and these events are captured for analysis. Analysis is done base on the change of behaviour of the machine after the malware is introduced. In this method events before malware injection and the events after are compared and based on that the machine is infected with malware or not is decided and also decide the code is malicious or not. This system has advantage over the signature based approach as it can detect the unknown malware samples The key element for using this approach efficiently is to generate rules in such a way that it can lower the false positive rate for unknown as well as known attacks.

E) **Heuristic Detection**

Fuzzy logic is also a type of signature based detection. It can be used to deal with particular portions of the file in which the probability of finding the malware sample is maximum and then calculate a hash of that particular byte of strings. In this method there can be many signatures that can be generated from the particular file, this is an advanced version to the traditional signature based approach as in that entire file signature is calculated which is not efficient as if any attacker changes the small portion of code or add some junk data then the signature changes and it will bypass the antivirus software. But, in fuzzy logic instead of calculating entire file signature only some portion signature is calculated which will be straight to the point and makes the system more efficient. It also have the capacity to detect the unknown samples of malware.
It increase the scanning speed and reduces the false positive rates. But, signatures database size increases. Fuzzy rules are also used to detect intrusion in real time. Two rule sets are generated and mined for online from training data. Features for comparison are taken from network packet header. This approach is used for large scale network attacks.

F) **Association rule based IDS**

As the virus and malware samples are freely available on the internet the malware attacks have become common now to perform the malware attack someone should not be expert in that field as code is available and after some modification attack can be launched. This scenario has increased the unknown malware attacks on the system. To detect such attacks, signature apriori algorithm can be used, which finds frequent subset that contains some features of original attack of given attack set. In this

approach, signature based algorithm are used to generates signatures for misuse detection. But, the main drawback of the proposed algorithm is its time consumption for generating signatures. This is solved the database scanning time problem examined. In the proposed scanning reduction algorithm to reduce number of database scans for effectively generating signatures from previously known attacks. However, this increases the false positive rate since unwanted patterns are produced.

Attack on cloud can takes place at various levels. Various types of IDS/IPS used in Cloud computing
There are mainly four types of IDS used in Cloud: Host based intrusion detection system (HIDS), Network based intrusion detection system (NIDS), Hypervisor based intrusion detection system and Distributed intrusion detection system (DIDS).[19]

A number of studies have addressed aspects of cloud security from different viewpoints (e.g. the network, hypervisor, guest VM and Operating System) under various approaches derived either from traditional rule-based Intrusion Detection Systems (IDSs) or statistical anomaly detection models.

G) **Malware & Detection Methods**

Extensive survey has been done in the domain of malware detection systems using both static and dynamic analysis. The main challenges within the development of resilient and secure cloud-oriented mechanisms which is dependent on the adequate identification and detection of malware. Malware detection is important because on cloud to perform various attacks malware acts as an initialization step, it acts as a backdoor for the attacker from where the attacker can perform other attack so it is important to first detect and prevent the malware in order to secure the entire platform. [1]
Despite of intensive research on the malware detection techniques a very little success have been achieved mainly on cloud environments. In particular, aimed to adjust the performance of traditional Intrusion Detection Systems under signature based techniques that employ Deep Packet Inspection on network packets. Various work in studied system-related features on monitored Virtual Machines by employing Virtual Machine Introspection methods in order to detect intrusions or attacks on a given VM's Operating System.
Even after performing such deep research a efficient real time malware detection and prevention technique is not developed. This id due to continuously changing cloud environments, the virtualization technology that is deployed. Some system is developed but it is purely based on signatures, and it may not provide a robust mechanism in detection of real time malware or intrusions as such. Earlier signature-based malware detectors identify malware by scanning unknown binaries for distinguishing byte sequences. Features unique to malware are maintained in a signature database, which is continuously updated as new malware is discovered and analyzed. This is important to provide a secure cloud based environment. This has led to the development of various data mining techniques for malware detection that can automatically infer signatures for previously unknown malware.

III. PROPOSED SYSTEM

A. Generating and collecting the malware signatures

Our proposed system is antivirus for cloud. One of the most important aspect of this system is it is a host based antivirus. For making an antivirus efficient one of the main step is to have a strong database. So in the first step we will create a malware signature database. This database will have signatures in md5 and sha1 format. These are just the hexadecimal strings that are generated after hashing function is applied on the unknown files. These signatures can be downloaded from various resources such as virustotal but in our system we have made our own hash calculator which will calculate the hash in the form of md5, sha1 etc and store it into the database. Now whenever we scan any folder or directory the system will compare them with the existing signatures from the database. If the signature is not found then the file is treated as unknown and then rule based detection is done. If any string matches with the rule then the file is treated as malicious and alert is generated and the signature of that new file is added to the database.

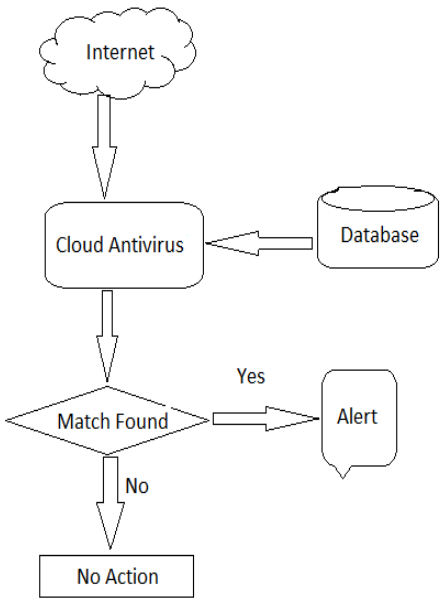

FIGURE 3.1 ARCHITECTURE OF THE SYSTEM

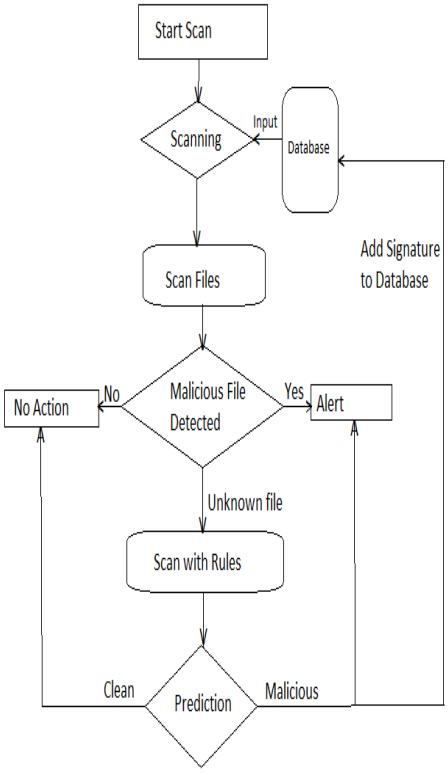

FIGURE 3.2 FLOW CHART OF THE SYSTEM

B. Malware Detection and prevention system

In this phase malware detection is done using our system. In this process folder or directories are scanned using signature base detection as given by the user. The system will convert all the files into their hash functions and then compare them with the signatures present in the database. When the scan is complete the system will show the detected infected files and option will be given to the user to delete the detected file. When user clicks on delete button the file gets deleted from the directory or folder. In cloud based detection when the user goes online and receives a file from any source and downloads it the system then automatically scans for the downloaded file if the signature is matched then alert is generated. If no signature is matched then the file is treated as unknown file and then the specific portions of file string are compared with the rules dataset. If the rules are matched the infected file alert is generated and the signature of the unknown file is added to the database. So next time when that same file is downloaded our system will automatically block it.

C. Real time malware prevention system

As the system is based on cloud so it is important to have the prevention system online. For this purpose we have a real time detection system. This system will continuously scan for the received files by the user and alert user if any malicious file is received and notify user. This an important phase in our proposed system. It is necessary to find the malicious file before entering into the system and infect our system before our antivirus detect it. Prevention system is an active method to prevent malware whereas detection is passive method that is why prevention method is very important in the system. When the user downloads any file the real time analysis system analyses the file automatically and pops up are shown if the file is infected this will help user to take the necessary steps and secure the cloud from any malicious intent of the attacker.

IV. CONCLUSIONS

We have proposed an antimalware system which can detect and prevent new and known malwares. Our system also provides a real time protection against malware on cloud. This real time protection is important to implement intrusion prevention system on cloud. In this paper discussed various cloud security details and malware detection methods on cloud and collection techniques.

With the help of our proposed system we have introduced an idea of combining the malware signatures and rules of pattern and string matching for the detection of new malware. We have also proposed the idea of real time analysis on cloud. This will enhance the intrusion prevention system and protect cloud user from the incoming threats.

REFERENCES

[1] M. R. Watson, N. Shirazi, A. K. Marnerides, A. Mauthe and D. Hutchison, "Malware Detection in Cloud Computing Infrastructures," in IEEE Transactions on Dependable and Secure Computing, vol. 13, no. 2, pp. 192-205, 1 March-April 2016.

[2] S. Das, Y. Liu, W. Zhang and M. Chandramohan, "Semantics-Based Online Malware Detection: Towards Efficient Real-Time Protection Against Malware," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 2, pp. 289-302, Feb. 2016

[3] L. J. Garcia Villalba, A. L. Sandoval Orozco and J. Maestre Vidal, "Malware Detection System by Payload Analysis of Network Traffic," in IEEE Latin America Transactions, vol. 13, no. 3, pp. 850-855, March 2015.

[4] Fischer A. et al. (2015) CloudIDEA: A Malware Defense Architecture for Cloud Data Centers. In: Debruyne C. et al. (eds) On the Move to Meaningful Internet Systems: OTM 2015 Conferences. OTM 2015. Lecture Notes in Computer Science, vol 9415. Springer, Cham

[5] K. R. Choo, O. F. Rana and M. Rajarajan, "Cloud Security Engineering: Theory, Practice and Future Research," in IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 372-374, 1 July-Sept. 2017.

[6] N. Paladi, C. Gehrmann and A. Michalas, "Providing User Security Guarantees in Public Infrastructure Clouds," in IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 405-419, 1 July-Sept. 2017.

[7] Fischer A. et al. (2015) CloudIDEA: A Malware Defense Architecture for Cloud Data Centers. In: Debruyne C. et al. (eds) On the Move to Meaningful Internet Systems: OTM 2015 Conferences. OTM 2015. Lecture Notes in Computer Science, vol 9415. Springer, Cham

[8] S. Lins, S. Schneider and A. Sunyaev, "Trust is Good, Control is Better: Creating Secure Clouds by Continuous

Auditing," in IEEE Transactions on Cloud Computing, vol. 6, no. 3, pp. 890-903, 1 July-Sept. 2018.

[9] W. Sha, Y. Zhu, M. Chen and T. Huang, "Statistical Learning for Anomaly Detection in Cloud Server Systems: A Multi-Order Markov Chain Framework," in IEEE Transactions on Cloud Computing, vol. 6, no. 2, pp. 401-413, 1 April-June 2018.

[10] K. R. Choo, O. F. Rana and M. Rajarajan, "Cloud Security Engineering: Theory, Practice and Future Research," in IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 372-374, 1 July-Sept. 2017.

[11] N. Paladi, C. Gehrmann and A. Michalas, "Providing User Security Guarantees in Public Infrastructure Clouds," in IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 405-419, 1 July-Sept. 2017.

[12] P. D. Ezhilchelvan and I. Mitrani, "Evaluating the Probability of Malicious Co-Residency in Public Clouds," in IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 420-427, 1 July-Sept. 2017.

[13] D. Gonzales, J. M. Kaplan, E. Saltzman, Z. Winkelman and D. Woods, "Cloud-Trust—a Security Assessment Model for Infrastructure as a Service (IaaS) Clouds," in IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 523-536, 1 July-Sept. 2017

[14] Y. Lin, C. Lee, Y. Wu, P. Ho, F. Wang and Y. Tsai, "Active versus Passive Malware Collection," in Computer, vol. 47, no. 4, pp. 59-65, Apr. 2014.

[15] G. Zhao, K. Xu, L. Xu and B. Wu, "Detecting APT Malware Infections Based on Malicious DNS and Traffic Analysis," in IEEE Access, vol. 3, pp. 1132-1142, 2015. doi: 10.1109/ACCESS.2015.2458581

[16] T. Y. Win, H. Tianfield and Q. Mair, "Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing," in IEEE Transactions on Big Data, vol. 4, no. 1, pp. 11-25, 1 March 2018.

[17] A. Mahboubi, S. Camtepe and H. Morarji, "A Study on Formal Methods to Generalize Heterogeneous Mobile Malware Propagation and Their Impacts," in IEEE Access, vol. 5, pp. 27740-27756, 2017.

[18] Ammar Ahmed E. Elhadi, Mohd Aizaini Maarof and Ahmed Hamza Osman "Malware Detection Based on Hybrid Signature Behaviour Application Programming Interface Call Graph" in American Journal of Applied Sciences 9 (3): 283-288, 2012 ISSN 1546-9239

[19] Chirag Modi, Dhiren Patel, Bhavesh Borisaniya, Hiren Patel, Avi Patel, Muttukrishnan Rajarajan, "A survey of intrusion detection techniques in Cloud" in Journal of Network and Computer Applications, 2012