Cloud Based Malware Detection Technique

Sagar Shaw[1], Manish Kumar Gupta[1], Sanjay Chakraborty[1],

[1]Deaprtment of Computer Science & Engineering,
Institute of Engineering & Management, Kolkata
{shaw.sagar09, gupta.manish414}@gmail.com, sanjay.chakraborty@iemcal.com

Abstract. Security is one of the major concerns in cloud computing now-a-days. Malicious code deployment is the main cause of threat in today's cloud paradigm. Antivirus software unable to detect many modern malware threats which causes serious impacts in basic cloud operations. This paper counsels a new model for malware detection on cloud architecture. This model enables identification of malicious and unwanted software by amalgamation of multiple detection engines. This paper follows DNA Sequence Detection Process, Symbolic Detection Process and Behavioral Detection Process to detect various threats. The Proposed approach (PMDM) can be deployed on a VMM which remains fully transparent to guest VM and to cloud users. However, PMDM prevents the malicious code running in one VM (infected VM) to spread into another non-infected VM with help of hosted VMM. After detecting malicious code by PMDM technique, it warns the other guest VMs about it. In this paper, a prototype of PMDM is partially implemented on one popular open source cloud architecture – Eucalyptus.

Keywords: Malware, Eucalyptus, Antivirus, Security, Cloud Computing, DNA Sequence, Symbolic Detection and Behavioral Detection,Sandbox.

1 Introduction

Detecting malicious file is a complicated work. The big amount of new malware files are growing at a shocking rate. Microsoft receives over 150 thousand new unknown files each day to be analyzed.Antivirus software is one of the most widely used tools for detecting.In this paper, we suggest a new model where a file mainly undergoesthese processes to detect malicious behavior.

1.1 DNA Sequence Detection Process

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule to identify regions of local or global similarity.

1.2 Symbolic Detection Process

In Symbolic detectionprocess we cluster the files and use symbol to detectmalware.

1.3Behavioral Detection Process

Analyzing behavior of the file is one of the best ways to detect malicious file. In Behavioral detectionprocess we use Anubis sandbox to detect new malicious file. ThisProposed Malware Detection Modelis deploying into cloud architecture which gives the resultant as Cloud Deployment Model (CDM) with the help of Eucalyptus.

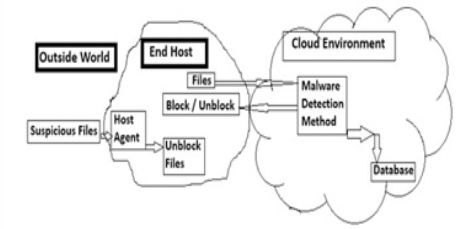
2Background

2.1Cloud Computing

Cloud computing is acommon term. According service model it is divides in 3 types: [4].Software-as-a-service, Platform-as-a-service, Infrastructure-as-a-service.

2.2Security in the Cloud computing

The Security in the Cloud is provided by many companies to detect malware with industry-leading detection rates. “Cipher Cloud” is company which provides service. We are providing a basic infrastructure of Cloud Malware Detection see [Fig1] [3] [9].



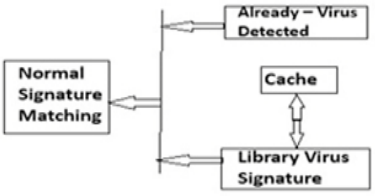


Fig 1. Cloud Malware Detection Technique.**Fig 2**. Traditional Anti-Virus Detection Method

2.3Related Work

The traditional anti- virus software’s can detect only those malware whose signatures are already present in the databases. This approach is based on the anomaly.[Fig 2] [3].

3Proposed System

Paper proposes a new malware detection system built on cloud environment. Initially, we will divide the system architecture into two main sections according to the mechanism of action of each part. First part, explains the Proposed Malware Detection Model (PMDM) and the second part, explains Cloud Deployment Model (CDM).

PART – I

3.1Proposed Malware Detection Model

The proposal is to find the optimal solutions to the problems of anti-viruses and improve performance and find possible alternatives for a better working environment without problems with high efficiency and flexibility.
In this malware detection model, total three process are used to explain the mechanism,

3.1.1 Process1: DNA Sequence Detection Process.

3.1.2 Process 2: Symbolic Detection Process consists of Clustering and Symbolic detection.

3.1.3 Process 3: Behavioral Detection Process using sandbox testing.
All of them are explained below in detail see [Fig 3].

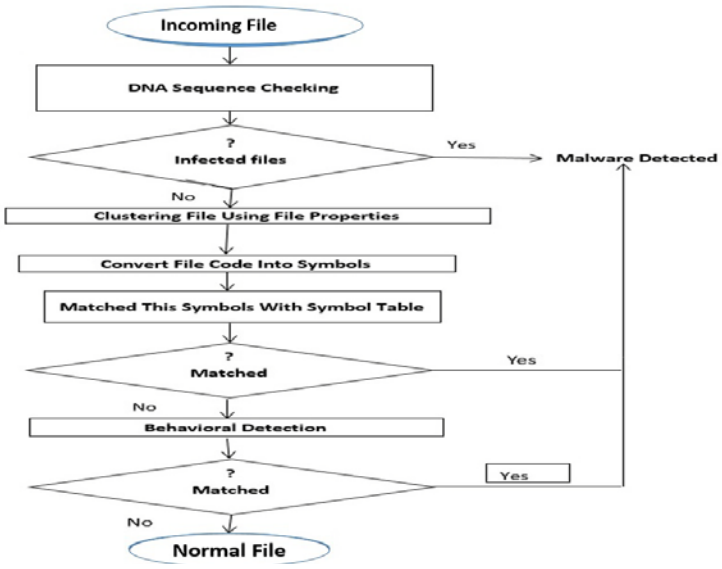


Fig 3. Flow Chart for Detection Method

In Process 1 we go for DNA Sequence checking, the initial step is the extraction of DNA sequence from a file is done by converting the file into its binary form and change each two corresponding bits into a DNA sequence character by using [Table I] [2]. The conversion is completely reversible.

TABLE I. DNA SEQUENCE MAPPING TABLE

Binary Bits	DNA Character	Binary Bits	DNA Character
00	T	10	C
01	G	11	A

After that Malware_Sequence_Database is created using these steps shown in Fig 4.

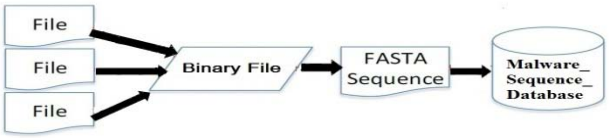


Fig 4.Creating BLAST Database.

The various input files are converted into binary filesthen converted into FASTA sequenceand merged to create aMalware_Sequence_Database.

```
>c: /user/name.txt
GTAGGGCCCGTTTGGCCAAAAATTTTTTTT
```

Fig 5.Example of DNA Sequence.

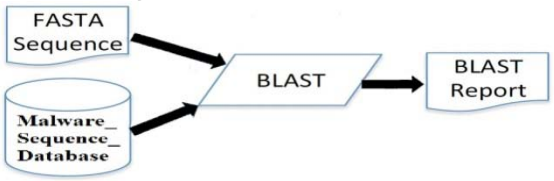


Fig 6.Comparing FASTA Sequence with Malware_Sequence_Database.

Now when new files come, it first convert to the DNA or fasta sequence then check with the Malware_Sequence_Database using Blast online software. The result of this comparison is a BLAST report, determine that the file is malicious or not see [Fig 6].

In Process 2 those files are undetected in process 1 are comes where we first cluster the files according to their file format, by checking the file format see [Fig 7]

Fig 7. Clustering File Formats

Then we use symbolic detection techniquein which files are converting to symbol using symbol database table. Then after we match symbol file with the existing symbol database which contain symbol of conventional malware signatures, see [Table II].

TABLE II. SYMBOL DATABASE TABLE

SL No.	String	Symbol	SL No.	String	Symbol
1.	:A	◄	**2.**	explorer	‼
3.	Start	↕	**4.**	shutdown	¿
5.	goto	§	**6.**	%random%	‰

If the file symbols are not matched with the existing symbol database then the file may be a new malicious file. Otherwise the files are detected in process 1 and 2 and blocked for the third process.In Process 3 the files which are pass through the second process are only go for the third process. In this process we detect malicious files using a virtual machine that extensively used for this type of analysis bytesting and running the file into a sandbox gives an optimal result to detect malware. For this purpose we use Anubis sandbox [7] which is free available. Anubis interact with file using API call and check the behavior of the file to identify whether it contain malware or not.

PART – II

3.2Cloud Deployment Model

The Proposed Malware Detection Model (PMDM) discuss in Part-I is deploy into cloud architecture i.e. Cloud Deployment Model (CDM) byusing a free open source computer software Eucalyptus. The purpose of this CDM is to implement PMDM in real cloud environment. In this experiment we determine the fitness of the proposal into the Eucalyptus architecture. PMDM is valuate against known and unknown malicious attacks. Here the PMDM system is partially implemented in the Eucalyptus architecture due to the architectural (infrastructure) limitations. The PMDM is mainly used to detect the malicious code based on the above discussed processes. And keep them as a set of warnings in a dedicated thread storage pool and block the malicious file to enter into a Guest VM or Guest Operation System see [Fig 8] [4] [5] [6].

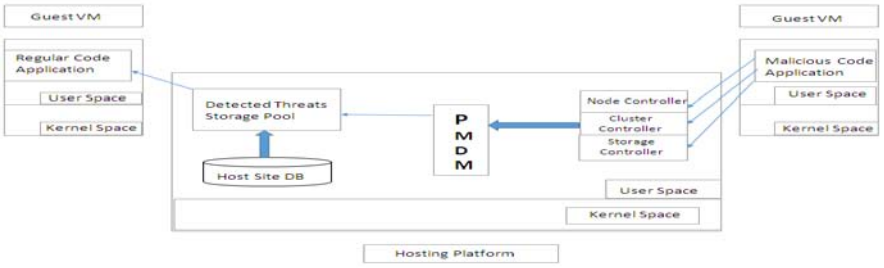


Fig 8. Proposed Malware Detection Method (PMDM) Embedded In Cloud Deployment Model (CDM)

4Result Analysis

4.1Result of DNA Sequuence Process.

Several experiments were designed to evaluate the usefulness of this process including document gathering, modification of DNA sequences, database creation, and software identification.

4.1.1Document Gathering

Total 1020 files were collected, types and counts of file were recorded in [Table III].

TABLE III. FILE COUNTS

186 Text files	152 Image Files	194 HTML Files
169 Java Files	99 Binary Files	

4.1.2Modify DNA Sequence

Converting the file into its binary form and change each two corresponding bits into a DNA sequence using by a java program character see [Table I].

4.1.3Database and software

The result which comes out from Process 1 is obtained by using BLAST software. We analyze the above files and observe the result one of the result output is given below: Groovemonitor.exe malware on BLAST and found a 100% identities matched i.e. it is a malware, see results in [Fig 9].

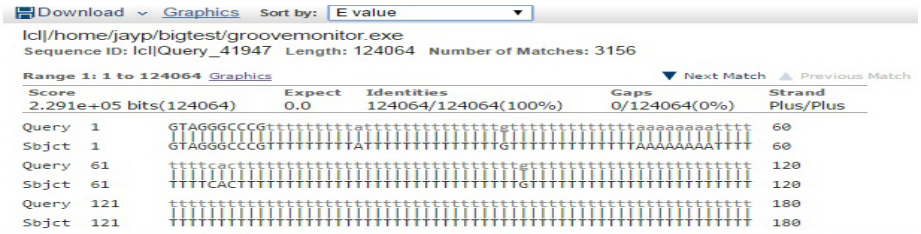


Fig 9. Descriptive Analysis Result of groovemonitor.exe

Same Analysis process is applied on an image file which is malware free and it result is below 70% as expected .From the above various files results the groovemonitor.exe file is blocked and the Image File1.bin and other files are pass for the second process.

4.2Result of Symbolic Detection Process

For symbolic detection there is also several operations were designed to evaluate the usefulness of this process including file clustering, converting characters into symbols and matching symbol with symbol table database.

4.2.1File Clustering.

In file clustering different types were cluster according to their file format see [Fig 7].

4.2.2Converting File Characters into Symbols

The conversion of file characters into symbol is done by using [Table II].

4.2.1Matching Symbol with Symbol Table Database.

The matching of symbols with symbol table is done by using the following pseudo code.

```
Take symbolfile name as Input
Scan symbolfile
While not EOF do
        If symbolfile match with virus symbol Then
                Print "File contain virus"
        Else
                Print "File does contain virus"
        Endif
End While
```

Fig 10.Pseudocode for Matching Symbol with Symbol Table Database.

Files match are found in the symbol table, as example here symbolvsample.txt file and it is blocked after that and other files that not matched are pass for the third process.

Fig 11. Symbolic Detection Result of symbolvsample.txt

4.3Result of Behavioral Detection Process

The files which pass the second process are only entertain in this process like from the above result all the files are entertained excluding groovemonitor.exe and symbolvsample.txt files. All the files are the input to the Anubis sandbox after that Anubis will check the files using API call and also check the behavior of the file to identify whether it contain malware or not.

4.4Comparison Between Traditional Malware Detection Vs Proposed Malware Detection.

Table IV. COMPARISON BETWEEN TRADITIONAL MALWARE DETECTION VS PROPOSED MALWARE DETECTION.

Features	Traditional Malware Detection Method	Proposed Malware Detection Method
Effected by security attacks	More	Less
Process time	More	Less
Implementation	Expensive	Cheap
For large number of files	Decrease Performance	Increase Performance
Attackers Intrusion	Easy	Difficult
Cloud solution	May or may not be available	Available

4.5Advantages.

4.5.1Advantage of DNA Sequencing.

Using DNA Sequencing detection method we can detect malware without opening the file, it will save time because we need not to see wholefile content to detect malware.

4.5.2Advantage of Smbolic Detection Process.

We cluster the file which gives a benefit of Post infection protection i.e. we actually know which portion of file content we have to see exactly for malware detection. Then matching the converted file with symbol database, will increase time efficiency as we not required to see whole malware signature matching only a small part of traditional malware signature symbol is sufficient to detect malware [10] [11].

4.5.3Advantage of Behavioral Detection Process.

InBehavioral malware detection solve the problem of cannot cope with malware variants i.e. malware can change their code and compiler setting to bypass the detection or Zero day protection problem i.e. once a new malware is produced its signature or symbol is unknown, so by testing the malicious file into a sandbox we can say that it is malicious or not.The overall advantage of the system is that it increases the efficiency and effectiveness for detection of malwares [10] [11]

5 Conclusion and Future Work

To conclude, it proposes an effective and advanced cloud security method that can detect the different malware attacks during cloud communication. It is also partially implemented on a popular architecture of Eucalyptus with modified. This paper discusses basic techniques in brief needed for the development of Proposed Malware Detection Model (PMDM) as it is actually cheap, requires less processing time and provides good performance for large numbers of files compare to other traditional malware detection systems. It is totally transparent to the user. We used the both optimal traditional detection methods and modern era methods to detect malwares in this paper. The proposal of this work is to find the best solutions to the problems of anti-malwares and improve performance and find possible alternatives for a better working environment without problems with high efficiency and flexibility. In future, we will see an increase in the dependence of cloud computing. The advantages of this approach include an increase in the number of clients that can be served for every physical server.

References

1. Dahl G.E., Stokes J.W. et al., "Large-scale malware classification using random projections and neural networks", IEEE International Conferenceon Acoustics, Speech and Signal Processing, pp. 3422 - 3426(2013).
2. Pedersen j., Bastola D., et al., "BLAST Your Way through Malware Malware Analysis Assisted by Bioinformatics Tools", International Conference on Security and Management (2012).
3. Hatem S.S., wafy M.H., et al., "Malware Detection in Cloud Computing", International Journal of Advanced Computer Science and Applications (IJACSA),Vol 5, Science and Information (2014).
4. Marinescu D.C., "Cloud Computing: Theory and Practice", MK Publication (2013).
5. Johnson D, Murari K. et al., "Eucalyptus Beginner's Guide- UEC Edition", v1.0 (2010).
6. Parmar H., Champaneria T., "Comparative Study of Open Nebula, Eucalyptus, Open Stack and Cloud Stack", International Journal of Advanced Research in Computer Science and Software Engineering ,Vol.4,No. 2, pp-714-721 (2014).
7. Mandl T., Bayer U.et al., "ANUBIS ANalyzing Unknown BInarieS The automatic Way", VIRUS Bulletin Conference, v 1.0.02(2009).
8. Oberheide J., Cooke E. et al., "CloudAV: N-Version Antivirus in the Network Cloud", 17th conference on Security symposium, pp- 91-106 (2008).
9. Graham M., “Behaviour of Botnets and Other Malware in Virtual Environments”, The Open Web Application Security Project (2014).
10. https://www.youtube.com/watch?v=fV5kED7nryw.