# Virus Detection Using Cloud Model

Lateeka
Department of Computer Science
Lovely Professional University
Punjab , India

Karanvir Kaur
Department of Computer Science
Lovely Professional University
Punjab , India

## Abstract

*The Anti-virus scans or the Heuristic approaches are based on the priori knowledge of the virus databases. In other words they implement static analysis for matching the virus strings with the programs code. Then also use of exhaustive search for pattern finding where each files internal structure is listed is really troublesome and a long procedure. In this paper we have discussed the dynamic aspects for the virus detection technique through the use of GA. The generation of new virus signature or pattern based on the priori knowledge of the virus data set which can be used for catching the unknown viruses thus preventing our valuable data from being corrupted. These virus signature are generally found in hexadecimal format. Thus , this proposed system solves the need to detect new launched viruses.*

## 1. Introduction

With the rapid development in the field of computers comes along the certain risks associated with it. These risks can only be reduced if only we have a reliable source of data. In other words we need to secure our data sources and also our systems so that they are not corrupted or modified. But now a days many computer viruses are spread across the network.

A Computer virus here can be defined as a self-replicating program that contains code to replicate itself and can also infect the other programs by modifying them. These viruses can alternately be called as malicious contents. They can be classified on the basis of purpose they serve : Worms , Trojans , Remote Access Tools (RATs) or Distributed Denial of Service (DDoS) Agents. When we talk about any virus program it mainly consists of three parts namely the infection mechanism , the trigger and the payload.

In such cases we have generated their signatures for detecting them via anti-virus scans. However these scans work properly until we the signatures for the known viruses but will fail for those whose signatures are not there in the virus dataset. In other words it lacks the capacity to detect unknown viruses. Other approach used can be Heuristic scanners that can detect viruses based on their structural or behavioral patterns. But again it lacks to detect for the unknown viruses. These virus signatures can be either in hexadecimal format or some binary format. One such example is Dark Avenger which is ugly but innovative memory resident virus that introduced the concept of fast infection. Its signature (Dark Avenger) is ‘2181 ebe7 0072 7b8c c1f9 13cb b44a cd21’.

In this paper we are interested in applying the self-signature generation algorithm for detecting the unknown viruses using cloud as a platform. The rest of the paper is organized as follows. Section 2 discusses the related work , Section 3 discusses about

proposed work , Section 4 tells about the results and Section 5 states the conclusion.

## 2. Related Work

The viruses no matter for what purposes are made by the attacker always causes damage to our system and our data. These viruses can be classified as worms , trojans , backdoors and spyware.

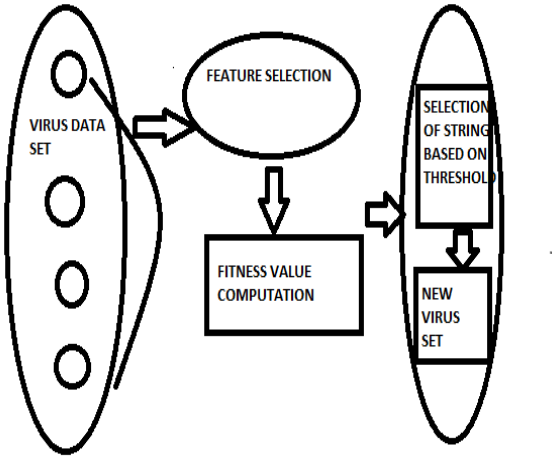
Mohamad Fadli Zolkipli and Aman Jantan[3] dicusses the combination of two malware detection techniques which are Signature-based technique and GA technique. Here they have presented three modules S-based detection , GA detection and S-based generator. S-based detection is a static analysis method where each file's code is matched against certain virus signature. In the S-based generator the signature is created when some malware is found and then it is updated in the virus database.

Lihua Wu and Yu Zhang[2] have introduced Automatic Detection Signature Discovery System (AMSDS) that generates malware signatures for AV cloud desktop. In this the same malware family is taken having some identical raw binary sequences. Then the suspicious sample is taken and its internal structure is listed. Thus some raw data is neglected whereas those segments where hackers may insert malicious content can be added is reserved for signature generation.

We all know that data mining approach is widely used now a days in every field to generate the hidden patterns from the large data sets. This technique can also be used for virus detection as discussed by Olivier Henchiri and Nathalie Japkowicz[1]. They propose an exhaustive search for n bytes of sequences of the files and record their frequencies that meet the virus family with the certain threshold. If the sequence of bytes taken have frequency retained then it is taken as a candidate. Also certain sequences are obtained from the virus family so if the sequences matched per family member (of same virus) is higher then it can be listed into virus feature set.

## 3. Proposed Work

In this paper we are proposing the use of self-signature generation algorithm using cloud as a platform. This proposed work is different from the work proposed by Mohamad Fadli Zolkipli and Aman Jantan in the way that here already known virus data sets are taken for which certain fitness value is computed. Based on this fitness value certain virus string are chosen irrespective of the class they belong to in order to generate the new set of viruses which are further used for matching each file's content in our system.




### 3.1 Self-Signature Generation Algorithm

This algorithm takes GA as a base where strings of data is taken as chromosomes. Then certain chromosomes are selected from the large set based on some value. And the

new chromosomes are produced from the existing ones.

The steps followed in this algorithm are :

1.Create an initial population of a set of viruses V.

2.Computation of fitness value for each virus set

2.1Certain substring are taken from the virus strings (hexadecimal format).

2.2For each substring created for each virus string we convert the hexadecimal value to the decimal value.

2.3Lastly certain computation is done on the above values in order to generate their fitness value.

3.After computing fitness values crossover operation is performed for the certain candidates selected.

4.The mutation operation is performed after the crossover operation to get new signatures.

5.After the steps 1-4 are done successfully we have a database of new virus set that can be used for unknown virus matching.

## 4. Results

In this paper two modules have been proposed : Firstly , Building the virus database module of the previous known viruses and Secondly , the Scanning module. The scanning module contains the new virus dataset generated and is used for scanning.


**Figure 1 - Scanning Module**

The above figure-1 shows the scanning module and the below figure-2 shows the virus database module.

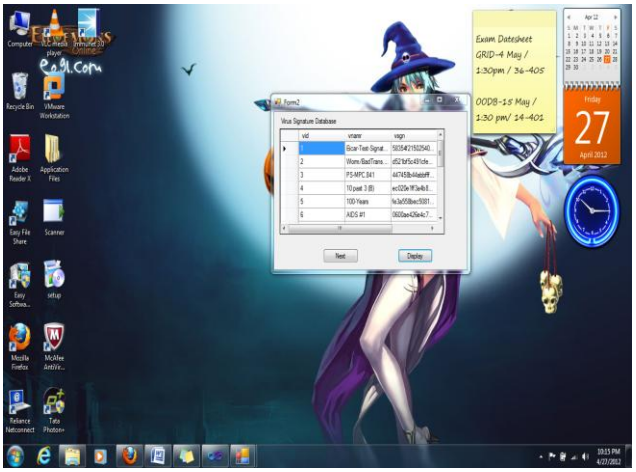


**Figure 2 - Database Module**

The figure-3 below shows the scan performed on the system and the threats detected.

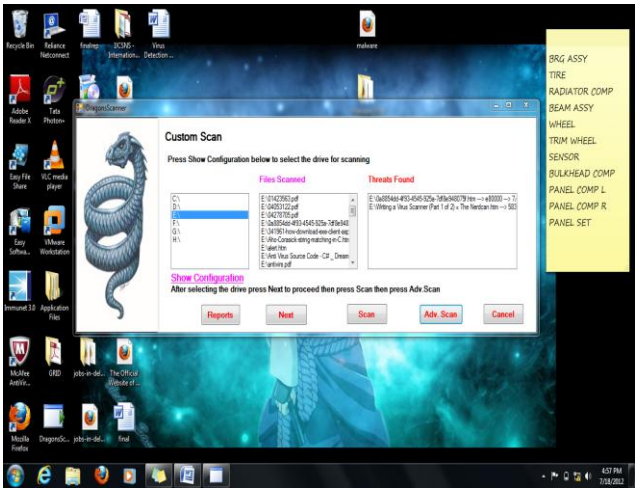



**Figure 3 - Threats Detected**

## 5. Conclusion

So in this work we have applied the use of previous knowledge that is use of virus dataset collected over time. And through the use of genetic concept have developed the new dataset that can be used for the virus detection of the unknown viruses.

## 6. References

[1] Olivier Henchiri and Nathalie Japkowicz "A Feature Selection and Evaluation Scheme for Computer Virus Detection" , 2006.

[2] Lihua Wu and Yu Zhang , "Automatic Detection Model of Malware Signature for Anti-virus Cloud Computing" , 2011.

[3] Mohamad Fadli Zolkipli and Aman Jantan , "A Framework for Malware Detection Using Combination Technique and Signature Generation" , 2010.

[4]Raviraj Choudhary and Ravi Saharan , "Feature detection Approach From Viruses Through Mining" , Volume 2 No,6 , June 2011

[5]Meng-songg Zou , Lan-sheng Han , Qi-wen Liu and Ming Liu ," Behavior based Malicious Executables Detection by Multi-Class SVM" , 2009.

[6]Soumya Banerjee , Nashwa El-Bendary , Aboul Ella Hassanien and Tai-hoon Kim , "A Modified Pheromone Dominant Ant Colony Algorithm for Computer Virus Detection" , 2011.

[7]Bo Li and Eul Gyu Im , "A Signature matching optimization policy for anti-virus programs" , 2011.