An Expert Anti-Malware Detection System

Richa Bhatnagar, Mariya Khurshid Ansari, Sakshi Bhatnagar, Harshbardhan Barik

Abstract: ***The malware expert system is an enhance approach for analyzing malware and other kinds of software. So, it is necessary to develop an effective malware expert system that can analyze, detect, classify and remove the malware codes. This system is necessary because it removes the errors done by human intervention in determining whether the files to be scanned contain any malicious data or not. There are various diverse approaches that were previously used to find and eradicate the malicious codes. But there were some loop- holes in existing strategies like the systems detect false positive malwares. The objective of malware detection expert system is to evaluate sample as malware or non-malware.***

Keywords: ***Anomaly, Adware, False Positive, False Negative, Hit Ratio, PUI (Program Under Inspection), Spyware, Trojan, Virus, Worms.***

I. INTRODUTION

Malware is a term that is poised of two words i.e. malicious plus software. Before Internet access became common, viruses' extent on personal computers by polluting the executable boot sectors of floppy disks. But today, malicious codes are most frequently written for the Windows OS, although a few are also written for Linux and Unix systems. These code today works in the same basic way as 1988's Internet Worm: they scan the network and use vulnerable computers to replicate. Malware is a malicious code or malicious software that is being designed to gain access to some computer system, to disrupt the various operations executing on a computer system or to perform some unwanted actions. Malware is of many types that are executed to perform unauthorized access to the computer resources or any sensitive information.

Malware are of many types like Virus, Trojan- horse etc. used for crime, industrial sabotage, vandalism etc. "Panda Security" has discovered about 73,000 new types of threats being released every day that was 26 percent raise during the year's first quarter compared with the same period in 2010.

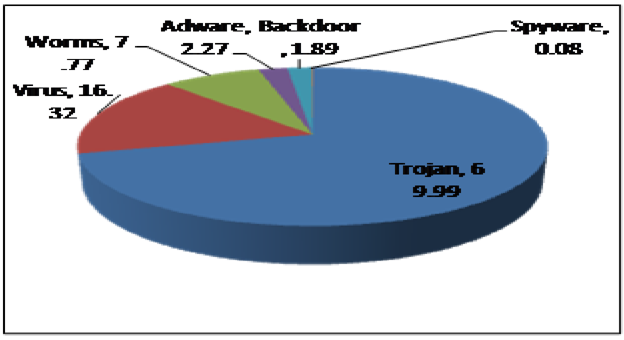


Figure.1. Distribution of various types of malware

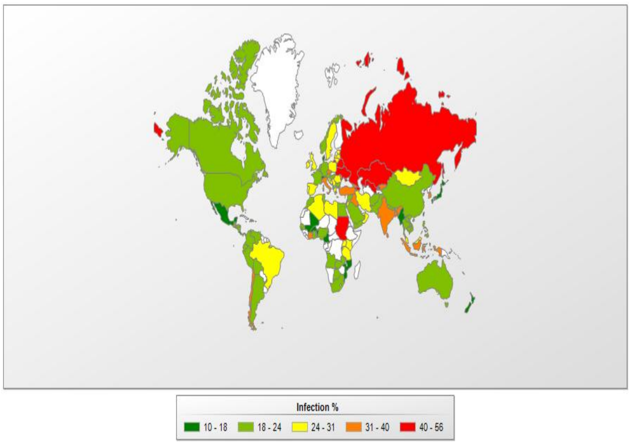


Figure. 2. Map of infection risk while surfing the internet

A. Types of Malware:

There are various types of malware generated and created to disrupt the computer system or its resources. Some of the malware are:

(i) Virus: Virus is a type of malware that promulgate by injecting itself into another program called host program. Virus can affect your system in a severe way. The viruses are attached to executable files that means a virus can exist in a system but are not active until and unless the host program is executed. Once the virus-infected file is executed the virus becomes active effecting the files and folders. A virus might damage or obliterate data on your computer system, or even expunge everything on your hard disk.

(ii) Worm: Worms are quite similar to viruses but the difference is that worms are self-replicating. Worms uses the computer network to transfer the malicious code to another system. Worms are the stand-alone software program that does not require any host program or human intervention to replicate itself to disrupt the data or information.

(iii) Trojan: A Trojan horse is another type of malware that masquerades or behaves as a helpful program with the purpose of conceding a attacker unauthorized access to a computer. Trojan does not inject themselves in host computer like viruses. Trojan horses are able to copy themselves, steal the confidential or sensitive information, or damage their host computer systems. Trojans enter the computer systems through drive-by downloads or installing online games, songs or movies or by internet driven applications in order to reach target computers.

(iv) Backdoor: Backdoor are the software that allow the hacker/ attacker to access the system without using username, password, or any other method to enter into the system that act as a front door. As the name suggests the software allows accessing the system from backdoor by bypassing the user authentication schemes. Hacker installs the backdoor program that helps them to access the system without entering username and password into the login screen.

Manuscript received on November 25, 2012.

Richa Bhatnagar, M.Tech(Computer Science) from IIMT, Meerut.
Mariya Khursid Ansari, Department of Computer Science MIET, Meerut, India.
Sakshi Bhatnagar, MCA from BIT, Meerut.
Harshbardhan Barik, Department of Computer Science IIMT, Meerut, India.

(v) Spyware: Spyware is a type of malware that is being installed on computer system and fetches information of users without the knowledge of user. Spyware can collect any type of data including user's personal information like bank account number, credit account number, internet surfing habits, user's password etc.

(vi) Adware: Adware are the software that try to sell something to the users which automatically appear as popup windows even if users don't open these or even interested in these. Normally adware enter to the systems in the form of the gambling advertisements or games format/pop-up windows and these advertisements are the part of websites, which you open.

II. MALWARE DETECTOR

A malware detector can be described as an implementation of the techniques used to detect the malware. Malware detector detects the behavior of the files in the system or coming into the system through networking. Malware detector helps in protecting the system from various types of malware by reading and analyzing the malicious behavior. There are two types of input that goes into the Malware Detector. They are as follows:

A. Knowledge of malicious behavior:

The detector has the knowledge about the difference in what is normal behavior and what is malicious behavior.

B. Program under inspection:

The detector has the program that has to be detected under surveillance.

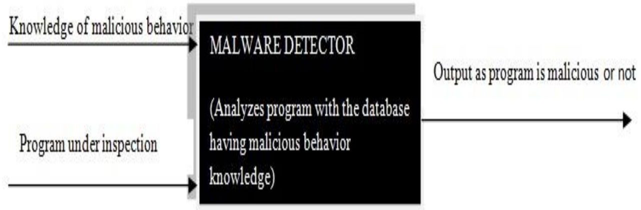

Knowledge of malicious behavior
MALWARE DETECTOR
(Analyzes program with the database having malicious behavior knowledge)
Output as program is malicious or not
Program under inspection

Figure.3. Inputs and output of malware detector

Above figure defines the working of Malware Detector. Malware detector takes two inputs i.e. knowledge of what is considered as malicious behavior and the other input is program under inspection. Once the malware detector has the knowledge of the malicious behavior (normal behavior) and the program under observation it can employ whether the program is malicious free or not. Malware detector is designed to find out the followings:

(i) False positive: False positive means when the detector finds a virus in a non-infected file. This problem occurs when the bit patterns/ signature of file is similar to that of a malware.

(ii) False negative: False negative means when the detector does not find the virus or threat in an infected file. This happens when the virus is new to the detector or the signature is not in the knowledge of detector.

(iii) Hit ratio: Hit ratio means when the detector is able to find the malware in an infected file. This is because the bit patterns or the signature is present in the database of malware detector.

III. RELATED WORK

A. Malware detection techniques:

There are basically 3 types of malware detection techniques:

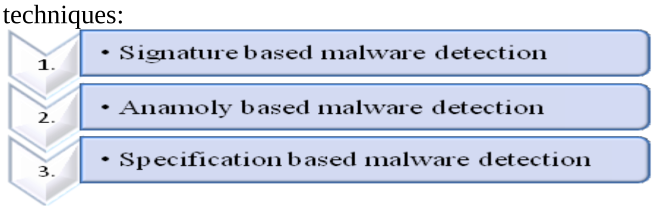

1.
• Signature based malware detection
2.
• Anamoly based malware detection
3.
• Specification based malware detection

Figure.4. Types of malware detection schemes.

(i) Signature based malware detection:

The signature based malware detection is based on the bit patterns a malware contains. Generally, the scanners that use this type of technique to detect the virus or other threats try to look for signatures in a file that is under scanning process. Signatures are basically a sequence of bytes that defines a malware.

In signature based malware detection, the signatures defining malicious code are present in repositories. When the PUI (Program under inspection) is detected then the detector accesses the repository to assess the signature. If the signature matches with the PUI then the program contains the malicious code. In this the human expertise create signature that represent the malicious behavior demonstrated by programs. When a signature is created by the developer, it is being added to the signature repository.

Fig.5. illustrates the major disadvantage of signature-based methods. Since the set of possible malicious behaviors, U, is infinitely large, there are no known techniques for accurately representing U via signatures. Furthermore, a repository of signatures is a weak approximation to U. Another drawback of signature-based methods is that human involvement/expertise is usually needed to develop the signatures. This not only allows for the introduction of human error, but takes considerably more time than if signature development was completely automated. Given that some malware has the capability to spread extremely fast, the capability to quickly develop an accurate signature becomes paramount [6].

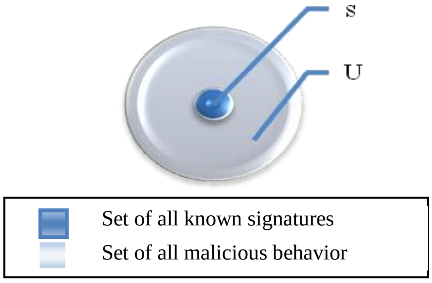


Figure.5. Signature Based Malware Detection

(ii) Anomaly based detection:

In anomaly based detection technique two phases are involved. In training phase the detector learns the normal behavior. It could be the learning of the host computer system or program under inspection or combination of both.

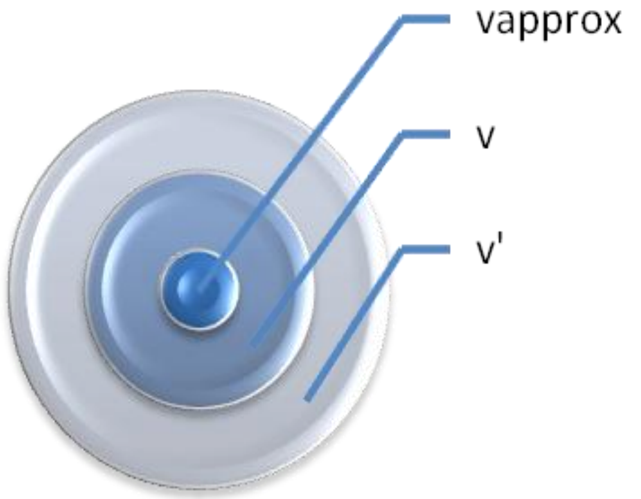


Figure.6. Anomaly Based Malware Detection.

Let A be the set of all behavior, V be the set of valid behavior, Vapprox be the set of approximate to V behavior. V is the valid behavior of the host system or PUI that is derived from a set of non-conflicting requirements. Vapprox is the approxiamte valid behavior might be flagged as malicious under anomaly based detection technique. The approximation to all valid behaviors made by anomaly-based detection methods is shown in the figure as the set Vapprox. Since Vapprox is an approximation,valid behavior might be flagged as malicious under anomaly-based detection methods.

For example, if an exception is never seen during training phase, an exception seen during the monitoring phase would cause an erroneous alarm. The possibility for a system to exhibit previously unseen behavior during the detection phase is not zero. Therefore, the probability of an anomaly-based technique raising a false positive is not zero.

(iii) Specification based technique:

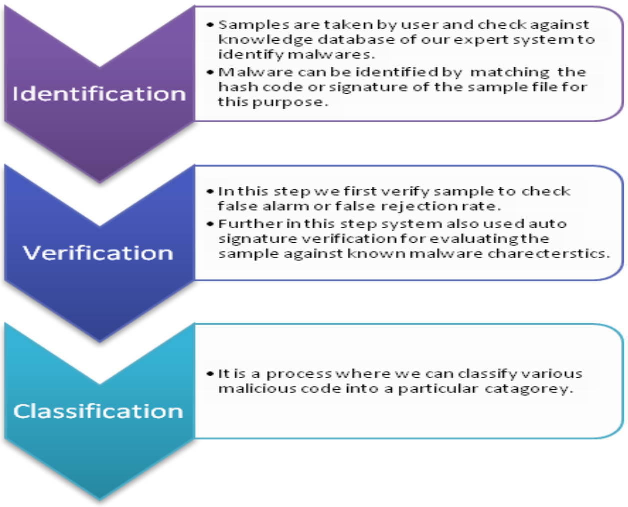
Instead of attempting to approximate the implementation of an application or system, specification-based detection attempts to approximate the requirements for an application or system. In specification-based detection, the training phase is the attainment of some rule set, which specifies all the valid behavior any program can exhibit for the system being protected or the program under inspection.

The main limitation of specification-based detection is that it is often difficult to specify completely and accurately the entire set valid behaviors a system should exhibit.

IV. TAILORED APPROACH

The approach that is suggested follows basically 3-major process for the suggested anti malware expert system model to identify, verify, and classify the malware within a system and then take an appropriate preventive or corrective action to mitigate the risk of malware or trying to reduce their impact.

A. Anti-malware expert system processes:



- Samples are taken by user and check against knowledge database of our expert system to identify malwares.
- Malware can be identified by matching the hash code or signature of the sample file for this purpose.

- In this step we first verify sample to check false alarm or false rejection rate.
- Further in this step system also used auto signature verification for evaluating the sample against known malware charecterstics.

- It is a process where we can classify various malicious code into a particular catagorey.

Figure.7. Anti Malware Expert System Process

(i) Sample Identification: When a sample file is being obtained from the user, it is being checked that the sample contains malicious code or not. The sample is checked against the Knowledge Base of expert system that contains already identified malicious code. Sample identification checks the identification of malware by comparing the hash code or by signature verification or simply by anti-virus scanning system. When a sample is being identified as being malicious by various methods then there is a high probability that the sample is not malicious free. Then the sample is being transferred to next process.

(ii) Auto Signature Verification Function: When the file is being suspected as malicious, there may be a possibility that the sample has produced false alarms. So to verify that the particular sample is actually infected the sample goes through auto-signature verification function. As we know the malicious codes defined in virus or Trojan horse has certain characteristics and those characteristics could be found in any infected file. So, by analyzing those characteristics it is depicted that a sample or Program Under Inspection (PUI) is really infected or not. If the file is infected the sample goes to the next process i.e. Sample Classification and the learning module of expert system make use of extracted signature and learns the new pattern of malicious code if the code is not already detected and the signature is then published into Knowledge Base of malicious signature and Knowledge Base of malicious behavior.

(iii) Sample Classification: Sample classification is a method through which we can classify the various malicious software codes in a category. The classification is defined according to the behavior and signature of malicious codes. The various malicious codes are cataloged so that appropriate actions could be taken.

B. The process flow for our malware detection expert system are taking place as following:

1. User generates, creates or downloads the file into the system.
2. The file with any extension i.e. any document (.doc, .pdf, .txt etc), image (.jpg, .png, .bmp etc) or any other type is scanned to detect any malware present in that particular file.

3. Then the file to be scanned goes through file identification process.
4. The file identification process check the identities of malware by matching the signatures of malware present in file with that of signatures present in database.
5. Sometimes, because of misconception the file identification produces false alarms. So, it is needed to verify that the malware detected in the file is really a malware or not. So, false positive and false negative alarms are checked.
6. To check false positive and false negative alarms the extract signature function comes into play. Here, the signature is being extracted from the file.
7. After the signature has been extracted, the auto signature verification function becomes activated. It works on the principle of artificial intelligence learning capability. It will correctly verify whether the file is really affected or not.
8. If the file is really infected the file undergoes file classification function, where the file being Infected is categorized in one of the category of malware by analyzing the behavior of the infected file.
9. If any new signature and behavior is being found the learning module capture that signature and behavior and store it in Knowledge Base of malicious signatures and the Knowledge Base of behavior for further forensics.
10. When a malware is found, appropriate action is taken and the explanation module explains the type, behavior and features of malware found in file to user. It also depicts the action followed to remove that malware from the file.

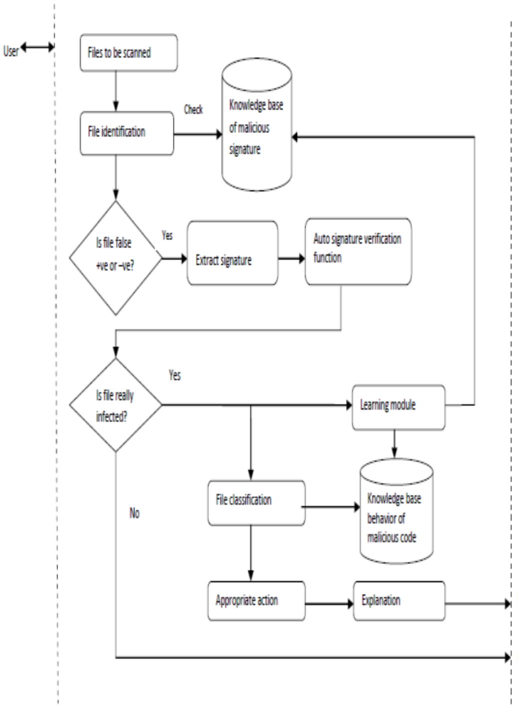


Figure.8. Flowchart for malware detection expert system

V. CONCLUSION

The suggested anti- malware detection system it is not claimed that the assurance is there for 100% security against malware, but we can fight more efficiently and in effective manner against various type of malware. As it is known that malware do not require any human intervention, so it is required to build such an anti malware detection system which is expert enough to handle almost all type of known malwares. In the suggested approach it is tried to build an anti malware detection expert system which can satisfy the above statement. Till now there is only proposed model but in future very soon we are trying to implement this and create an effective working anti-malware expert system.

REFRENCES

1. Mihai Christodorescu and Somesh Jha," Testing Malware Detectors", in Proc. ISSTA'04, July 11 - 14, 2004.pages 33-44, Boston, MA USA, ACM Press.
2. J.Xu, A.H.Sung, P.Chavez, S.Mukkamala," Polymorphic Malicious Executable Scanner by API Sequence Analysis", Proceeding of 4th IEEE Symposium of International Conference on Hybrid Intelligent Systems (HIS '04).
3. Arun Lakhotia , Aditya Kapoor , Eric Uday, "Are Metamorphic Viruses Really Invincible Part 2" , Virus Bulletin ,January 2005.
4. Abhishek Karnik, Suchandra Goswami and Ratan Guha," Detecting Obfuscated Viruses Using Cosine Similarity Analysis", in The Proceeding of IEEE Symposium First International Conference on Modelling & Simulation (ASM '07).
5. Heng Yin, Dawn Song, Manuel Egele, Christopher Krugel, and Engin Kirda, "Panorama: Capturing System – wide Information Flow for Malware Detection and Analysis", in Proc CCS'07, October 29 November 2, 2007, Alexandria, Virginia, USA,ACM Press.
6. Gerard Wagener, Radu State, Alexandre Dulaunoy," Malware Behavior Analysis", Springer-Verlag, France 2007.
7. Greoigre Jacob, Herve Debar, Eric Fillol,"Behavioral detection of malware: from a survey towards an established taxonomy", Springer-Verlag, France 2008.
8. http://download.microsoft.com/download/a/b/e/abefdf1c-96bd-40d6-a138-e320b6b25bd3 /understandingantimalwaretechnologies.pdf.
9. http://www.security.iitk.ac.in/contents/events/workshops/iitkhack09/papers/vinod.pdf.
10. http://www.fortiguard.com/sites/default/files/DetectingMalware Threats.pdf.

AUTHORS PROFILE

Richa Bhatnagar: Born in 1988 in Meerut District (Uttar Pradesh). Phone Number +919456019431 and E-mail Id: bhatnagar.richa1@gmail.com. She completed her B.Tech. (InformationTechnology) from Uttar Pradesh Technical University and currently pursuing M.Tech(Computer Science) from Mahamaya Technical University. She had 2 year Teaching Experience from Meerut Institute of Engineering & Technology, Meerut and Meerut International Institute of Technology, Meerut. Her major field of study areas are Information Security, DBMS, and Software Testing.

Mariya Khurshid: Born in 1989 in Kanpur District (Uttar Pradesh). Phone Number +918791277359 and E-mail Id : khurshid.mariya@gmail.com. She completed her B.Tech. (Computer Science And Engineering) from Uttar Pradesh Technical University and currently pursuing M.Tech. She had 2.0 year Teaching Experience from Meerut Institute of Engineering & Technology, Meerut. College Of Engineering and Rural Technology, Meerut. Her major field of study areas are Data Mining and Warehousing.

Sakshi Bhatnagar: Born in 1989 in Meerut District (Uttar Pradesh). Phone Number +919456019430 and E-mail Id: sakshibhatnagar0929@gmail.com. She completed her BCA from CCS University and currently pursuing MCA from Mahamaya Technical University. Her major field of study areas are Information Security, DBMS, and Networking.

Harshbardhan Barik: Born in 1983 in Raurkela District (Orissa). Phone Number +919675453914 and E-mail Id: harsa03_iacr@rediffmail.com. He completed his B.Tech. (Computer Science and Engineering) from Bijupatanayak Technical University and M.Tech (InformationTechnology) from GGSIPU, Delhi. He had 7.5 years teaching Experience from IACR, Rayagada(Orissa) and IIMT Engineering College, Meerut . His major field of study areas are DAA, OOS, Theory Of Computation and Speech Processing.