© 2015 IJSRSET | Volume 1 | Issue 6 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Computer Viruses – Analysis of Detection Techniques and their Limitations

Harjit Singh
Punjabi University Neighbourhood Campus, Dehla Seehan (Sangrur), Punjab, India

ABSTRACT

A Virus is basically developed to disturb the working of a computer and intentionally cause damage to its working. A virus can do so by file corruption, user document damage such as movie files, photographs, text documents, music files and in such a way to make the computer completely useless. A virus is not like a spyware developed for some mysterious behavior, it is developed to arise problems to use the computer and damage something. A computer virus replicates itself similar to a human virus and grows itself with the help of file sharing or email attachments. There are various detection techniques available and used by antivirus software to identify virus infected files and mark those files as infected and also try to clean them if possible. These techniques are successful to detect some specific category of viruses but may fail to detect other category of viruses. Some of the main categories of virus detection techniques are Static Signature Scanning Technique, Generic Signature Scanning Technique, Heuristic Analysis Technique and Integrity Checking Technique. But each of these techniques has its own limitations and the techniques fail to detect viruses with properties beyond the scope of the technique used. So, each antivirus software uses mixed set of techniques to detect virus infections in files. This paper discusses various types of virus threats and analyses various detection techniques in view of their limitations.

Keywords : Virus, Virus Beginning, Virus Targets, Virus Spreading, Virus Detection Techniques, Static Signature Scanning Technique, Generic Signature Scanning Technique, Heuristic Analysis Technique and Integrity Checking Technique.

I. INTRODUCTION

A computer virus is like a biological virus that is not an independent object in itself. It must attach itself to a host which can be some program or document so that it can spread itself. It requires some medium in the form of a document etc. A virus is similar to a computer program and is designed in such a way that it can circulate itself into computer.

It can be so powerful that it can corrupt important files on a system such as user's text documents, music files, movie files, photographs and even executable files making the system unusable. A virus is not like a spyware developed for some mysterious behaviour, it is developed to arise problems to use the computer and damage something. A computer virus replicates itself similar to a human virus and grows itself with the help of shared files.

A virus is very much capable to quickly damage application programs on a computer or damage personal text documents at a slow speed. Even then, it is mostly not capable to propagate itself from computer to computer. Humans help it to spread by sending email attachments, sharing files using floppy or pen drive from one computer to another. If I receive an infected file in my pen drive or through network, it will transmit to my computer and so on.

Even the term "virus" became a common term to call everything malicious on a computer; actually it is a term for a specific type of malware. A digital virus damages the system similar to a biological virus. First it attaches to a host file/document and then tries to propagate through. It is different from other type of computer malware in the sense that it is capable to propagate itself by infecting other documents/files instead of living alone. A virus could be much damaging to delete important files and documents or it can be so simple to display a disturbing message on the screen at some specific time or date.

When a user finds that his/her system is not behaving normally, there are definitely much chances of infection from a virus. To further investigate, we can check some symptoms as follows:

- Applications are hanging.
- Unable to access some files.
- Start up problems while booting up.
- Some keys like caps lock stops responding or take time to respond.
- Documents became bigger in size.
- Display of abnormal messages on screen.
- Oddly looking images or messages on screen.
- Some abnormal sounds.
- Your email contacts are receiving abnormal emails from your email accounts, but you did not send any such mails.

II. WHAT VIRUS DOES TO SPREAD?

Each virus needs the help of some host application to survive in it. We can call it as file infector virus. When we execute such host application, the virus also starts working to do its own business by residing itself in memory. The main motive is to find similar applications where it can spread itself and reside in. When it successfully finds such an application, it attacks such an application and embeds itself into its code as a guest to survive in it (Figure 1).

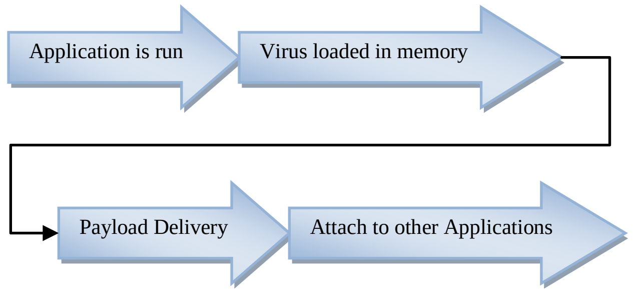

Figure 1: Infection by virus.

Many type of viruses are destructive in nature other than only copying themselves. They can be capable to delete your important documents, they can change the boot sector to make the system non-bootable, abnormal on screen message display, un-pleasing sounds appear, automatically sending multiple emails to others using your email account to spread themselves to others' computers as well. In this way they are carry on copying to other computers.

They are designed in such a way to deliver the payload on target machine when they are run. They are programmed to do so and to copy themselves and damage the files on target systems.

III. THE BEGINNING

During late forties, the computer systems were not so common things and the malicious programs start to birth. Actually, the viruses used the concept from theory of automata presented by John von Neumann. This theory of automata states the concept of replication by a computer program itself. It becomes the basis for the development of computer viruses. The theories were first implemented in the form of computer games at Bell Labs in early fifties. The game was to reproduce programmed organisms where the players compete to get the control of the main computer by such replication. The viruses we observe today works on the similar concept to damage our systems.

The viruses were widely started being developed in early eighties when the computers became the common gadgets. It was the era of floppy diskettes where these diskettes performed the major role in spreading those malicious codes from one computer to another with the help of human users. Humans played the major role in that era when networks were not common.

An important incident of virus infection was in early eighties when Apple II was infected through a floppy diskette. It became popular with the name Elk Cloner but was not harmful to damage any files on the target. It was designed to display a message on the screen that may be disturbing and make the users believe that something serious has occurred on their system. It makes the users afraid of something that is not visible to them but doing some abnormal tasks with their system. The message was shown as a warning message that something is being done on your system though a cloner. That is why it was named Elk Cloner. The warning was to control the computers all diskettes and even to embed into the computer chips. Although it did not do any such damage in real time. The spread of this virus challenged the computer professionals and they start thinking about to do something to stop these malicious codes from further infections. In this routine, a computer professional named Len Adleman presented a demo of an experimental virus on a computer system. The demonstration was to get the deep understanding of the malicious codes that may be harmful the computer systems and installed application programs including the operating systems. In mid-eighties, Microsoft disk operating systems became a very popular personal computer operating system. A virus named Brain Virus was spread to attack ms-dos by infecting its files. It is basically known as the first file infector that attached ms-dos. In the same year a Trojan horse was also spread to infect personal computers which include ms-dos and other similar operating systems. In early nineties, among the globally spread viruses, one is named Michelangelo. It was globally spread on the computers in the whole world. It covered large number of computers worldwide by infecting the operating system files and application programs. In late nineties, emails became popular with the popularity of internet. To spread through emails over the internet, a virus named Melissa was spread. It was capable to replicate itself through emails sent over the internet. It was a combination of worm and macro-virus and was specially designed to embed in emails and replicate like a worm. In the end of nineties, a very much popular virus named love bug was spread to shut-down most the worlds email systems and affected the corporate email accounts.

IV. THE VIRUS TARGETS

A computer virus is nothing different from a computer program that we develop in various programming languages. The only difference is that normally we perform simple calculations and productive tasks by developing such computer programs. But when a program is designed to do some harmful work, it is a virus. These are specially designed to attach themselves to other productive application programs but do something that is unexpected from a productive program. There are exist other type of such malicious program that do not attach themselves to other productive programs. They do not embed. They may be Trojan horses or computer worms. Virus behaviour is very much different from these malicious programs. Technically, they cannot be called viruses. We can examine where the virus attach itself i.e. what may be the target of a virus.

A virus can target the boot sector of a hard disk drive or a floppy diskette. By doing so, it loads itself into the memory by executing with operating system files when the systems boots up. In case of a hard disk drive, we call this sector as master boot record but in case of a floppy diskette we call it the boot sector. Since it loads itself into memory on booting up, so it remains alive in memory and when some other floppy diskette or hard disk drive is attached to the computer, it finds a new target for it to replicate. It copies itself to that attached floppy diskette or hard disk drive and resides in the appropriate sector i.e. the master boot record or the boot sector. Such viruses were very common in the nineties because in that ear floppy diskette was very much common to transfer files from computer to computer. So, floppy diskette was the major medium for the spread of this virus that targets the master boot record or boot sector to reside. The virus that targets the boot sector or master boot record, works by overwriting the booting program with the booting code that is already infected. The virus is so intelligent that it moves the original booting code from the boot sector to some other sector on the floppy diskette or hard disk drive and then marks it as bad sector to make that sector inaccessible to the operating system. Due the boot sector as the target, it becomes the most dangerous type of virus and difficult to detect taking the whole control of the target computer system by loading itself in the memory.

Another target of a virus can be some executable application program or file. Very common virus programs targets such executable files and hides themselves in such files. We know that executable files differ from one operating system to another, the type of viruses also differ. A virus designed to attack a Windows computer will be different from a virus that is designed to attack Linux computer. The target application program can be a video game, some utility program, some word processing application, some presentation program or a music/movie player. Such viruses designed to attach Windows computer system mostly attack files such as .exe, .com, .bat, .sys etc. Other working of such viruses is similar i.e. when we execute an infected application or tool, the virus program loads itself into the main memory of the computer even much before the main application loads itself. The virus loads in separate memory space than the space occupied by the application program. It makes the virus capable to remain alive in memory even if we close the application program. So even after exiting the host application, the virus will be able to infect other applications that will be launched after that. It makes the virus very dangerous. The working of these viruses is similar to the viruses that target the boot sectors of floppy diskettes and hard disk drives. In this sense, they overwrite the original program's loading statements with their own loading statements. The original program's loading statements are moved to other section of the target file. In this way it increases the size of the target executable files, which is a symptom to investigate that the files get infected and should be repaired or replaced with original ones. Another way of working of such virus programs is to rename the original file with some other extension e.g. .com files are renamed with .exe extension. After that new files with same name and .com extension are created by the virus. The virus hides the new files by manipulating the attribute settings on operating system. The order of file execution of files on ms-dos operating system is .com files and then .exe files. So by executing infection .com files before .exe executable files the virus loads itself into the main memory and remains alive there to fine its new target application programs for replication. In the beginning when internet was not so common and macro viruses are not working over the network, the viruses that targets executable files were very common as compared to other type of viruses. In modern times such viruses are modified to infect files in other ways to replicate over the internet which is a common medium now-a-days.

One more target of a virus can be the macro coding language. A number of application programs are developed using macro coding languages. A macro can be defined as a tiny code written to perform some specific task in that application software. For example, we use macros in ms-access and ms-excel to perform some repeatable lengthy tasks in one click. These macros are executed within the application software in which they are written. The VBA code is very popular on Microsoft platform to write macros in ms-office applications to automate specific operations and can be attached to the menus as well. This method can be used to destruct files and automated emails over the internet connection. The viruses that target the macro codes can be considered as more hazardous than the viruses who targets boot sector or executable files. Since they reside inside the document files in the form of script, so it is very difficult to track such viruses and stop them from execution. The viruses that target executable files are relatively easy to trace and stop their execution. Just opening a ms-word or ms-excel file can execute the macro virus and we cannot find it just by checking the size of the document because it depends upon the contents written in the document and it is not fixed. These viruses that target macro codes are platform independent and can replicate from one operating system to another such as windows to Linux and vice versa. It makes such viruses more dangerous than the viruses that targets boot sector of floppy diskettes or hard disk drives and viruses that targets executable files.

Another target of viruses can be web pages and similar applications. These viruses are developed using script languages such as vbscript or javascript that are used in web pages and similar applications. These viruses are automatically executed when a web page is opened or the applications containing such scripts are accessed. In modern times, surfing the web is a very common activity on computer systems, so these viruses that targets the web pages are very common and very easily spread over the internet. These viruses are very dangerous and are able to infect similar files that supports such scripts. The web pages are designed to include advertisements and other codes that run in background. Such web pages are main targets of script viruses. The viruses hide themselves in some advertisement.

V. VIRUS DETECTION TECHNIQUES AND THEIR LIMITATIONS

From the birth of viruses, the techniques to detect viruses are also being developed so as to protect computer systems from these malicious threats. These techniques can be categorised into following different categories:

A. Static Signature Scanning Technique

This technique is a common technique to detect known viruses that gets attached to a computer file. The virus attaches some specific set of instructions to the file and these instructions need to be executed in the same sequence in order to function. These instructions are actually strings of bits called “virus signatures”. Mostly the computer antivirus software searches for these virus signatures in order to identify the virus infection in a file. If the virus signature is found in a particular file then that file is marked as “virus infected file”. The requirement to this technique is that the virus signature must be stored in some database from where the antivirus program reads that information to find it in the file. If in the database, virus signature of some virus is not stored, then that virus can’t be detected by the antivirus software. More the entries in the virus signature database, more is the capability of antivirus program to detect various types of virus.

The disadvantage of String Scanning technique is that it can detect only those viruses whose signatures are present in the database. It means that the technique is unable to detect unknown viruses. Thus the database needs to be continuously updated with new virus signatures so that the antivirus program can detect new viruses. Finding the signature of a new virus is not so easy task, it requires deep analysis of the virus and it is a task of very skilful researcher. It may take time to find the signature of a new virus, by the time the virus may cause considerable damage to the files without being detected. Also this technique can’t detect dynamic signature viruses because they have the capability to change their signatures after they are attached to a file.

B. Generic Signature Scanning Technique

This is a modified version of Static Signature Scanning technique with wildcards. Some viruses are able to modify their signature after they are attached to a file, so they become difficult to detect. This Generic Signature Scanning technique is used where wildcards are used to compare and search for virus signatures instead of exact matching of string of bits. In this way the antivirus software can detect all variants of same family of viruses. The technique is somewhat similar to regular expression matching where a pattern of string of bits is found from a long string of bits. Mostly new viruses are created from existing ones by changing some pattern so this technique is able to detect those new viruses whose signature’s pattern matches with existing virus signatures.

The disadvantage of this technique is that, it may not be able to detect those viruses whose virus signatures does not match with wildcard scope. If a new virus is created with totally new signatures and its signature does not match with the pattern defined by wildcards, then it can’t be detected by using this technique. Also again there is a main role of database of signatures be present in order to detect viruses.

C. Heuristic Analysis Technique

There are many types of viruses that the Signature Scanning Techniques are unable to detect due to their specific properties. For example in metamorphic viruses, the virus signature continuously keeps on changing so it can’t be detected using signature scanning. Similarly, encryptor/decryptor viruses can’t be detected by using Signature Scanning Techniques because they become unreadable because of encryption. Such type of viruses can be detected by Heuristic Analysis Technique using static or dynamic analysis of the behaviour of infected binary file. This technique probabilistically identifies unknown and new viruses. In static heuristic analysis, there is a database of harmful code fragments they may cause damage to files. The binary file to be inspected is reverse engineered and the obtained code is analysed by comparing it with code fragments stored in the database. Some suspicious behaviour of a virus may include deleting files, adding fake registry entries, replication etc. If some file analysis shows that type of virus like behaviour then it is marked as infected. In dynamic heuristic analysis, the binary file is executed in a virtual environment and it is allowed to perform its functions normally and its behaviour is analysed in that virtual environment. In the analysis if it is observed that the file performs some suspicious functions such as deleting files, adding fake registry entries, replication etc. Then the file is marked as infected. Although, this method is slow but it is more capable to identify an unknown virus.

The disadvantage of Heuristic Analysis Technique is that in case of Static heuristic analysis harmful code fragments from the database are to be compared with the code inside the file in order to identify harmful behaviour of a file. But a harmful behaviour can be implemented using a number of different ways. For example to delete a file from a hard disk a number of different ways are there and all those ways need different set of instructions. Although, dynamic heuristic analysis is capable to identify suspicious behaviour but this method is too slow. The delayed identification may cause considerable damage to the files. Also some viruses are activated with some user action such as pressing a button or pressing some combination of keys. In such cases dynamic heuristic analysis will be unable to detect such viruses. Similarly, date or time specific viruses will be activating until that date and time reached. Such viruses will remain undetected.

D. Integrity Checking Technique

Some viruses use vague mechanisms like mutation, obfuscation and activating periodically. Such viruses are very difficult to detect using Signature Scanning and Heuristic Analysis techniques. In integrity checking technique, the uninfected fingerprint of file is to be stored in a secure location on the disk and while integrity checking, the stored uninfected fingerprint is compared with present fingerprint. In the fresh stage of the system the fingerprints of all the files are calculated and stored at safe location on the disk. There are a number of algorithms like CRD32, MD4, MD5 etc. to calculate fingerprints. The same algorithm is used to calculate present fingerprint of a file while integrity checking. If the fingerprint match occurs, the file is considered as uninfected and is allowed to execute but if there is a mismatch in the stored and presently calculated fingerprint of a file then the file is marked as infected.

The disadvantage of Integrity Checking technique is the lack of accuracy. The method may report false positives because some useful programs may change themselves and are designed to store their configuration or user data in some binary files. The integrity checking technique may mark such programs as infected which is a false positive indication. Also when the fingerprint of files is being stored, it is assumed that the files are uninfected but it may not be true, they may already be infected. Such files are taken as uninfected by integrity checking technique. Above all, the technique is very slow for huge binary files where each time the file is executed it is analysed and takes lot of time for computing checksums.

VI. CONCLUSION AND SUGGESTIONS

A virus is not like a spyware developed for some mysterious behavior, it is developed to arise problems to use the computer and damage something. There are various types of viruses and at the same time various categories of virus detection techniques are available in antivirus software programs. These techniques include Static Signature Scanning Technique, Generic Signature Scanning Technique, Heuristic Analysis Technique and Integrity Checking Technique. Each technique is capable to identify some specific type of viruses but may fail to detect other types of viruses.

The limitations in each analysed technique shows that no technique is perfect in its function and more research need to be performed in the field of virus detection techniques.

VII. REFERENCES

[1] Wing Wong, Analysis and Detection of Metamorphic Computer Viruses, San Jose State University SJSU ScholarWorks, May, 2006

[2] Sulaiman Al Amro, Ali Alkhalifah, A Comparative Study of Virus Detection Techniques, International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:9, No:6, 2015

[3] Computer Associates Virus Information Center (www3.ca.com/virus/).

[4] Essam Al Daoud, Iqbal H. Jebril, Belal Zaqaibeh, Computer Virus Strategies and Detection Methods, Int. J. Open Problems Compt. Math., Vol. 1, No. 2, September 2008

[5] Computer Security Resource Center Virus Information (csrc.ncsl.nist.gov/virus/).

[6] Prabhat K. Singh, Arun Lakhotia, Analysis and Detection of Computer Viruses and Worms: An Annotated Bibliography, ACM SIGPLAN Notices 29 V. 37(2) February 2002

[7] F-Secure Security Information Center (www.datafellows.com/virus-info/).

[8] Umakant Mishra, Methods of virus detection And their limitations, http://www.trizsite.com

[9] IBM Antivirus Research Project (www.research.ibm.com/antivirus/).

[10] McAfee AVERT (www.mcafeeb2b.com/naicommon/avert/).

[11] Anita Thengade, Aishwarya Khaire, Devaj Mitra, Alok Goyal, Virus Detection Techniques and Their Limitations, International Journal of

Scientific & Engineering Research, Volume 5, Issue 10, October-2014 ISSN 2229-5518

[12] Sophos Virus Analyses (www.sophos.com/virusinfo/analyses/).

[13] Symantec Security Response (www.symantec.com).

[14] What You Can Do About Computer Viruses 17.

[15] Trend Micro Virus Information Center (www.antivirus.com/vinfo/).

[16] Virus Bulletin (www.virusbtn.com).

[17] Viruslist.com (www.viruslist.com).

[18] The WildList Organization International (www.wildlist.org).