Systematic Study of Computer Virus Using Virus Definition and Scanning Techniques

ABSTRACT

Background: Computer virus pose increased risk to computer data integrity, they cause loss of important information and data and cost a huge enormous amount in wasted effort in restoration or duplication of lost and damaged data. Everyday new viruses are discovered, and as the problem increases, there is the need to employ modern tools to identify them and to eliminate them from the devices.
Objective: This study will examine computer viruses and their impact to computer system.
Methodology: The study adopted the mixed research method of quantitative and qualitative techniques. In this study data was collected using both interview and observation methods, as well as comprehensive search of the literature review of various scholars from relevant databases such as Research Gate, Google Scholar, and dimension.
Result: This study identified how devices are vulnerable to computer virus attack and how modern technologies can actually increase the propagation of computer viruses. It also describes the technologies that are available to identify and eliminate the computer viruses, hence, propound probable good uses of virus technology.

Keywords: *Attack, Computer virus, Malicious, Risk, Scanning.*

1. INTRODUCTION

When possibilities of viruseswere first revealed in scholarly articles that were published in 1980's, many people did not take it seriously. However, within a short span of time the first wide-scale personal computer boot sector virus called "brain", which was created in the year 1986 by two siblings named Basit and AmjatFaroot Alvi in the city of Lahore, Pakistan.This virus infection caused a media sensation, but not an outrage. People were generally fascinated by the novel concept of a computer virus but few saw its full dangerous potential [1].

Right from that time up to present day modern computing, millions of individuals, organizations, and businesses have continued to sufferfrom one computer virus to another. Unlike older viruses which usually display a funny message on the screen or just a ball bouncing; the present day viruses such as Michelangelo Virus are extremelydamaging, which were programmed purposely to corrupt data, destroy data, erase files, format hard drives and encrypt partition tables [2].

According to the literature computer viruses are programs that are mostly created intentionally to wreak havoc to computer systems [3]. The computer virus can duplicate itself and infect systems without express authorization or knowledge of the user. The virus can also modify the copies or the copies may modify themselves. Everyday computers and other communication devices are being affected by viruses, which readily affects their performance. According to [4] the major problems caused by computer viruses consist of:

- Slow system performance
- Inconsistent system behavior,
- Unexplained data loss
- Frequent computer crashes amongst others.

A computer virus usually spread from one system to another when its host is taken to the un-infected system. For example, when one user sends the virus over a network or through some external storage devices such as CD (Compact Disk) or DVD (Discrete Versatile Disk), or USB(Universal Serial Bus) flash memoryor external hard disks.The computer virus danger is here to stay! In most part of the world, it has reached an epidemic proportion and the number of the viruses seem to be increasing day by day.As the field of computer viruses is evolving at an alarming pace with increased sophistication of new viruses which are network - aware and the rapid increase in the use of network by the users. There is also a need for a corresponding increase in the security measures to be taken by the users [5].

Signature scanning may not be able to detect all the possible viruses where signature was not predefined in the signature database [6]. A systematic study of computer virus using virus definition becomes a challenge in the information technology sector in order to design antivirus strategies that can be adopted by every organization, banks, individual computer users and various companies in order to reduce high level of insecurity of computer systems, data/documents loss, etc. Computer virus poses increased threat to computer data integrity. It causes the loss of important data and costs a huge amount of wasted effort to restore the corrupted data. Every time new viruses are discovered, and as the number increases, there is the need to put measures and techniques to identify and eliminate them from the computer systems. Hence, this study is projectedto assumes its dimension.

Since designing and implementing an effective antivirus software requires proper understanding of the structure of the virus,how they work, and mechanism of the action,the types or part of area they affect in the system. This study will serve as motivation for understanding the internal setting of the computer system, particularly the operating registry.

The major objective of this study is a systematic study of computer viruses using virus definition. Other specific objectives include:

1. To study the phenomenon of virus and other malicious codes, that is how virus works,their method of attachment, as well as how they cause damage to data programs and files.
2. To describe the technologies that are available to identify and eliminate computer viruses and other malicious codes, worms and Trojan horses.
3. To provide well defined steps on how to protect and prevent illegal stealing of confidential data such as credit card number, account password, etc.
4. To provide procedural advice on how to fight the problem.
5. To educate and create user awareness on how to prevent computer virus and other malicious codes from infecting their systems.

2. RELATED LITERATURE REVIEW

According to [7] in his book titled Computer Virus defined computer virus as an unsolicitedsoftware programs that can attack the computer hard disk drive and cause various types of destruction. A computer virus is usually created when someone writes software programs and inserts harmful codes within the program.
According to the literature, the computer virus is a destructive and harmful computer software programs that duplicate itself autonomously through file systems in the computer. In essence, they are sectors of a program that are stored, and when they are executed, are able to create a duplicate copy of themselves in another stored program [1].
The canonical formal definition of computer viruses was stated by Cohen, in which computer virus is defined as a class of Turing machine strings capable of replicating itself. In this study, Cohen's definition of computer virus was adopted [8].

A common misunderstanding is that computer virus is simply an information security issue, such as Denial-of-Service (DoS) attacks or stack buffer overflows, that is a contemporary issue that are best left for the industry, or thecommercial software vendors to solve. In fact, computer viruses are integrated in stored program computers, since over time they spread from one computer program to another. Most modern computers are centered based on the stored program (von Neumann) model, and therefore are vulnerable to attacks. As the ability to process information increases, and the financial costof computers decreases, we expect to reach a stage of ubiquity, or pervasiveness, at which point computer systems will be fully integrated into everyday life [9].

Computer viruses are of different types, which are classified according to their origin, techniques, types of files they infect, where they hide the kind of damages they cause,and the types of generating system or platform they attack. Every computer virus is made up of at least two basic parts orsubroutine. Firstly, they contain a search routine that locates new files, which are worthwhile targets for attack. This routine determines how well the virus can reproduce. For example, whether it will do so quickly or slowly or whether it can attack every section of the disk or just some specific areas. The more sophisticated the search routine is, the more space it will take up. So, there is always a size versus functionality trade-off of the virus. Secondly, virus always contains a routine to duplicate itself into the area which the search routine locates. The copy is always sophisticated enough to do its job of getting detected [10].

Computer virus is a software program that can perform the same thing as any other program running in a system. The definiteconsequence of any particular virus depends on how it was programmed by the person who writes it. Some computer viruses are purposely created to damage programs or system files or otherwise interfere with the computer operations, while others usually spread themselves around, but even these ones that just spread themselves are also destructive in nature, since they destroy programs and files and can also cause other problems in the process of spreading [4].

A computer system can be attacked by a virus when infected file is copied into the hard drive, then this will likely activate the code inside by executing the infected application or by opening the infected document. An email attachment, a download, or via a shared external storage device like flash drive, CDs. Another easier channel of getting computer virus is via e-mail attachments, which are downloaded into the computer system via an internet connectivity. As soon as the corrupted or infected file is open through an application, the malicious code duplicates itself into a file on the computers' hard drive. Where it wants to deliver its payload (whatever the programmer designed it to do to your system), by simply deleting the email after opening the attachment would not eliminate the virus, since it has already entered the machine. A computer virus written can enable the payload to trigger execution of a specific command, such as when a file is opened or saved. For example, the Michelangelo virus was designed to release its payload on March 6th of any year of the artist's birthday.Computer viruses cannot do any physical damage to hardware, they would not melt down our system unit or cause the computer monitor to explode. However, some potential warnings about viruses destroying computers are mostly not real, they are just hoaxes put forward in order to cause panic to computer users [11].

3. METHODOLOGY

This section describes the methodology used in the study, which include the research design, the methods and the instrument used in the data collection, the validity and reliability of the instrument, as well as the justification of the chosen method.

3.1 Research Design

The studywas designed to focus on studying the phenomenon of viruses and other malicious codes, how they work, their methods of attachment, as well as how they cause damage to data programs and files. It also describes the technologies that are available to identify and eliminate computer virus and other malicious codes, worms and Trojan horses. It provides well defined steps on how to protect and prevent illegal stealing of confidential data such as credit card number, account passwords, etc. It provides procedural advice on how to overcome the challenges posed by the malicious computer viruses. Finally, the research design was structured to educate and create user awareness on how to prevent computer virus and other malicious codes from infecting their systems (such as workstations and personal computers).

3.2 Methods and Instruments of Data Collection

There are several methods can be used for designing, collecting,and conducting the study in the form of data. This part of the study focuses on the procedures and processes for data collection methods and the instruments used. In this methodology, there are two major concepts that are used in the study, these are the primary research method, which is also known as the current analysis technique and also secondary research methods. Also known as the conceptual analysis technique.

3.2.1 Primary Research Method

The primary research method consists of the original sources of data and information, which are normally extracted from the source data and is then analyzed after being included in the assessment sheet. Assessment criteria plays a major role in the gathering of vital information from people in form of fieldwork. The vital information is normally collected via structured, semi-structured, or unstructured interviews or conversations with the relevant stakeholders, in this case, the computer users. Other methods can also be used to extract the information, such as questionnaire distribution, telephone conversations, radio communication, and also email messages [12].

For this present study in focus, the methods used for collecting the primarydata are the structured interview as well as observation technique.

a. **Structured interview:** This is primary data collected by conducting oral interviews with the computer users in order to understand how they feel about computer viruses and the possible solutions they are expecting from the researchers. Five computer users were interviewed using a structured interview with their consent, among them are novice, intermittent and expert users.

   Some of the interview questions includes:

   i. What do you understand by the term “computer virus”?
   ii. How often do you use your computer system?
   iii. For what purpose(s) do you use your computer system?
   iv. Has your computer been attacked by a computer virus program?
   v. How do you think computer viruses are spread?
   vi. Which indication have you experienced after your computer got infected with a virus?
   vii. Which virus have you come across in your computer?
   viii. How can you protect your computer system from virus attack?

The responses of the interviewees were given accordingly which were presented and analyzed in subsequent sections.

b. Observation: This is another primary data that was collected by a close look and the existing antivirus software's such as clam antivirus for researchers, which are also open source will be undertaking to understand how they are being designed, implemented and maintained. I have observed that most viruses are purposely programmed to harm the computer which may include slow system performance, irregular computer behavior, unexplained loss of data, destroying important programs, deleting files as well as reformatting the hard drive.

3.2.2 Secondary Research Method

Secondary data can be defined as information that is collected through studying the articles and journals of other scholars in the field of literature. The data that was collected from the secondary method was utilized in a layer of thorough examination before being incorporated into the needs assessment. This is expected to achieve better and accurate precision in the report [13].
The secondary data can be categorized into different classes that may contain data with different specifications. Examples of secondary data include but not limited to web materials, communication media reports, published journals, conference papers, magazines, newspaper clips, e-books, acknowledgements and information that has been cleansed, broken down and gathered for a reason other than the needs assessment [14].

a. Reviewed articles: This is a secondary data that was collected by studying the relevant scholarly works of other researchers in the area of computer virus. A comprehensive search across multiple databases such as ScienceDirect, ResearchGate, and Google Scholar was used to identify and retrieve relevant articles, which include journals, conference papers, e-books, magazines that were searched using the search terms "computer virus", "anti-virus software programs", and "malicious codes".

3.3Validity and Reliability of the Instruments

Validity and reliability are the two major techniques that are used when developing and testing a research instrument (e.g., face-to-face interview, observations, content assessment test, questionnaire, etc.) for use in a study. These techniques improve the quality of the measurement and the collected data used in the study [15].

Validity can be defined as the extent to which an instrument measures accurately what it is designed to measure. There are three common types of validity that are used by researchers and evaluators, which include construct validity, content validity, and criterion validity.Reliability on the other hand refers to the stability, dependability and procedure of the method explored [16]. The methods used in this study are valid and reliable, since the data collected do not in any way contradict each other.

The documents method used for collecting data in this study work provides the following benefits, which include:

a. Saving cost, which ensures judicious usage of money spent and time.
b. Accurate and concise information was collected for the study. The structured interview provides some advantages, which include:
    i. A better understanding of the collected information was achieved, through proper and careful explanations of the interview questions to the respondents.

ii. The collected datawere specially focused to the needs of the study and are therefore not restricted to culture specific.

3.4 Justification of the Chosen Methods

The rationale for selecting the methods of data collection, which included primary data collection techniques of observation and face-to-face interview with computer users, as well as secondary data collection technique of reviewed articles in the internet through Google Scholar search engine include the following:

i. Defining thestatement of the problem: A primary research methodology will help in identifying the real problems faced by investigators during malware detection and analysis. Hence, conducting an interview with the various computer users assisted the author in defining the problem statement of the study in a much more realistic manner.
ii. Understanding needs and requirements: Studies that use a primary research method assist a researcher in understanding the needs and requirements of the end user. Quantitative and qualitative analysis can be used to determine the precise and accurate needs of all end users facing the same dilemmas. Observation and face-to-face interview analysis gives statistical data that is more realistic and provides accurate information on the current needs and requirements of users of virus detection and analysis. In addition, bottlenecks in any virus detection process were acknowledged.
iii. Discovering evasion techniques: Conducting primary research for the study can provide information on various evasion techniques. An enhanced obfuscation applied to the malware code makes detection and analysis flexible and simplistic for an investigator. Thus, the gathered information helped in determining various recommended evasion techniques.
iv. Analyzing an attacker's perception: An interview method can also assist in determining the attackers' perception; this can be achieved by choosing questions on behaviour analysis carried out by investigators based on their experiences. This valuable information helped in determining the vulnerabilities of existing systems and assisted the author with determining underlying updated security features to be included in the proposed solution.

The truth behind choosing those methods mentioned above came after the need to make progress during the preparation for more of the phase of data analysis. There is a need for interactive interviews with computer users in order to set good social/feasibility.

After interviewing the computer users, then comes the need for observation to have practical understanding. The last method used is internet browsing which helps to generate the latest information to analyze systematic study of computer viruses using virus definition.

4. RESULTS AND DISCUSSION

The study conducted an interview as a means of primary data collection. The interview session consisted of asking some willing participants some questions regarding their knowledge and experience of computer virus attack.

The questions were earlier stated in Chapter Three under the methodology. However, in order for the answers obtained from the interviewees to be fully understood, the interview questions are again stated here with their corresponding answers according to the opinion of the participants.

i. What do you understand by the term “computer virus”?
One of the respondent defined computer virus as a malicious program that is coded and designed to spread from device to device. Most viruses are harmful and destructive to the computer that it infects.
Another respondent also defined virus as “a harmful program written to spread from one host to another and it also has the ability to replicate itself and spread rapidly.
Similarly, another respondent defined virus as “any unwanted set of codes written for the purpose of malicious intent.

ii. How often do you use your computer system?
Majority of the interviewees responded that they use their computers regularly.

iii. For what purpose(s) do you use your computer system?
Majority of the interviewees responded that they use their computer systems for academic purpose only while few replied that they use it for academic purpose and entertainment.
Most of the interviewees that uses their computers for entertainment such as watching YouTube videos, downloading films from the internet and downloading and installing applications online, etc. complained of high rate of virus spread onto their computers.

iv. Has your computer been attacked by a computer virus program?
Majority of those interviewed admitted that they were victims of computer virus attack in one way or the other.

v. How do you think computer viruses are spread?
Some of the interviewees responded that it spreads into their computers through email attachment, especially spam or junk mails. While others replied that it spreads through storage devices such as flash memory.

vi. Which indication have you experienced after your computer got infected with a virus?
Here, the responses differed from individual to individual. One of the respondent said he encountered slow computer performance, another person said he encountered unexplained data loss, another one said he encountered duplication of files, and another person said he encountered erratic computer behaviour, another person also admitted to experiencing hiding of files.

vii. Which virus have you come across in your computer system?
Majority of the respondents replied that they got infected with spyware, while others replied that they got infected with worms, while others replied that they got infected with Trojan horse virus.

viii. How can you protect your computer system from virus attack?
The interviewees gave many recommendations on how to protect a system from virus attack, among the recommendations made were as follows:

- Limit your downloads
- Keep your anti-virus software updated
- Don’t open emails with attachments unless you know the source
- Use a very strong password that will be difficult to crack
- Back up your computer system regularly in case of virus attack.

4.1 Structure of a Signature Scanning Virus Detection Software

The most common computer virus menace is virus scanning. Virus scanners consist of two parts; a scanning engine and a component that feeds the data to the scanning engine, the scanning engine searchers for virus signatures or a small pattern that uniquely identifies a virus [10].

Using “signature” of a virus or virus “definition” to detect its presence in an executable is the simplest, economical and most common approach for detecting the majority of current viruses.

Signature scanning program which can only detect known viruses which are included in the data base of a virus patterns is simple and cheap enough to be easily available and valuable to the public in general, and it has minimum impact on the existing code and hardware moreover, it is simple to add new virus patterns to the scanner database whenever a new virus is discovered.

Majority of the commercial scanners identify computer virus using a signature database. These scanners boast larger virus databases ranging anywhere from 65,000 to 120,000 patterns of fingerprints that have been built over long periods. They also use “heuristic” engine for scanning. The heuristic engine eliminates files that cannot contain viruses, and scans only the suspicious ones. Such heuristic typically includes identifying executable files, type appropriate file size, and scan only certain regions of files for viruses [10].

4.3 Virus Signature

Virus signature is a unique string of bits, or the binary patterns of a virus. The virus signature is like a finger print in such a way that it can only be used to detect and identify specific viruses. Simple signatures are usually specified as a string of characters in ASCII.A923BF7, for example, signatures of time that contain fixed pattern of hexadecimal digits, for some viruses, these fixed patterns are not sufficiently powerful to define the signature in a compact way for viruses, one would have to specify a large number of fixed patterns to identify a simple virus [18].

4.4 Drawback of Scanning Using Virus Signature

The main drawback of scanning as a method of detecting virus in a computer system is its failure to detect unknown or new viruses. Scanning also fails with self-excerpted viruses. It also fails with executables compressed with technique, making the same virus appear different.

Finally, with the discovery of more computer viruses, the scanning algorithms may tend to perform more slowly, because they have to match a large set of possibilities. Also, the more signatures there are, the more likely that any arbitrary signature will also match some legitimate code in some applications [17].

5. CONCLUSION

Computer virus issues have become an interesting aspect of research and development in the aspect of computing. This study identified various types and categories of computer viruses, as well as the various techniques used in identifying and detecting them. The study also discovered that as computer usage increases, there will be increase in the number of comber of computer viruses, and they may even become more sophisticated in terms of attacking method and difficulty in detecting. Hence, it becomes necessary for individuals and organizations to use third party tools that canmonitor and defend computer systems from virus attack.

The most important thing to remember when dealing with computer viruses is to avoid panicking; viruses do not possess mystical power. They are also software programs that have to adapt to the limitations of all other programs, they can only do their dirty work if they are executed. Having a very good antivirus program in the computer system can

systematically prevent any form of virus attack before it could even happen. As long as you are virus-conscious, not virus paranoid, you can prevent or recover from any attack.

6. RECOMMENDATIONS

The study made the following recommendations that can serve as a guide for dealing with the problem of computer viruses:

1. Reduce or avoid the usage of public-domain software: This type of software may easily contain virus, because they don't have stringent configuration controls. And being in the public domain, this software may become a likely target for computer virus authors. Therefore, if you are going to use the public domain software, it is recommended to subject it to test in an isolated environment.
2. Be conscious of any oddbehaviors: This does not mean to become fearful, but it simply means to be very conscious. When a computer becomes slow in booting at start-up or when loading a program, may not necessarily be a virus attack, but it can be a sign or symptoms of likely virus attack. Therefore, it is recommended to run a quick or full virus scan.
3. If a computer virus is identified in a disk, then it is recommended to isolate it with immediate effect. Thereafter, a virus scan can be done to eliminate the virus.
4. Adopt the usage of software protection mechanisms on disks and files where necessary.
5. Use licensed antivirus software programs for detecting and eliminating the computer virus. Avoid unlicensed antivirus software programs, as they can also become harmful programs and bring destruction to the computer system.
6. It is recommended to always back-up your important files and documents in an external storage device or in the cloud storage system. Having a back-up means even if your files and documents are destroyed by virus, you can always recover them.

REFERENCES

1. Alenezi, M. N., Alabdulrazzaq, H., Alshaher, A. A., &Alkharang, M. M. (2020). Evolution of malware threats and techniques: A review. *International journal of communication networks and information security*, *12*(3), 326-337.
2. Valli C, Brand M. (2008); Malware Analysis Body of Knowledge. Paper presented at the 6th Australian Digital Forensics Conference; Edith Cowan University; Mount Lawley Campus; Western Australia.
3. Gaikwad, R. (2016). Computer Viruses and Detection: A review.
4. Oyelere, S. S., &Oyelere, L. S. (2015). Users' perception of the effects of viruses on computer systems–An empirical research. *African journal of computing & ICT*, *8*(1), 121-130.
5. Gan, C., Yang, X., Zhu, Q., Jin, J., & He, L. (2013). The spread of computer virus under the effect of external computers. *Nonlinear Dynamics*, *73*, 1615-1620.
6. Naidu, V., Whalley, J. & Narayanan, A. (2018). Generating Rule-Based Signatures for Detecting Polymorphic Variants Using Data Mining and Sequence Alignment Approaches. *Journal of Information Security*, **9**, 265-298. doi: 10.4236/jis.2018.94019.
7. Hawkins, J. (2017). Computer Virus: The Damaging Facts about Computer Viruses! eBook.
8. McMullin, B. (2000). John von Neumann and the evolutionary growth of complexity: Looking backward, looking forward. *Artificial life*, *6*(4), 347-361.

9. Garfinkel, T., & Rosenblum, M. (2003). A virtual machine introspection based architecture for intrusion detection. In *Ndss* (Vol. 3, No. 2003, pp. 191-206).
10. Jaiswal, M. (2017). Computer Viruses: Principles of Exertion, Occurrence and Awareness. *International Journal of Creative Research Thoughts (IJCRT)*, 648-651.
11. Slade, R. (2012). *Guide to Computer Viruses: How to avoid them, how to get rid of them, and how to get help*. Springer.
12. Curtis, K. R. (2008). Conducting market research using primary data. *Assessment and Strategy Development for Agriculture*.
13. Schuurman, B. (2020). Research on terrorism, 2007–2016: A review of data, methods, and authorship. *Terrorism and Political Violence*, *32*(5), 1011-1026.
14. Church, R. M. (2002). The effective use of secondary data. *Learning and motivation*, *33*(1), 32-45.
15. Mohajan, H. K. (2017). Two criteria for good measurements in research: Validity and reliability. *Annals of Spiru Haret University. Economic Series*, *17*(4), 59-82.
16. Cohen, L., Manion, L., & Morrison, K. (2017). Validity and reliability. In *Research methods in education* (pp. 245-284). Routledge.
17. Wanjala, M. Y., & Jacob, N. M. (2017). Review of Viruses and Antivirus patterns. *Glob. J. Comput. Sci. Technol*, *17*, 1-3.
18. Driscoll, D. L. (2011). Introduction to primary research: Observations, surveys, and interviews. *Writing spaces: Readings on writing*, *2*, 153-174.