IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. XII (Mar-Apr. 2014), PP 116-121
www.iosrjournals.org

Detection and Removal of Infection Sources in a Network

Ms. Rintamol Mathew[1], Ms. Preethymol B[2]
[1]*Department of Computer Science and Engineering, Nehru College of Engineering and Research Center, India*
[2]*Department of Computer Science and Engineering, Nehru College of Engineering and Research Center, India*

Abstract: *Identifying an infection source in a network is a challenging problem. An infection source can be either a disease, virus or a rumor spreading through a network. Detection of infection source in a network, allows timely quarantine of the infection like a virus spreading in a network, thus limiting the damage caused. Here we take the computer virus as an infection source and tries to identify the infection source based on the knowledge of which nodes are infected and their connections. For the detection of an infection source, we use signature scanning method. Signature scanning is the most popular method of virus scanning and is capable of detecting more than 80% of viruses.*

Keywords**:** *Infection Source, Computer Virus, Signature Scanning, Source Estimation*

I. Introduction

An infection can be a disease, virus or a rumor spreading through a network. ie, in a population network, the infection is the disease that is transmitted between individuals. In the example of a computer virus spreading in a network, the infection is the computer virus, while for the case of a rumor spreading in a social network, the infection is the rumor [1]. Our goal is to find an infection source in a network based on the knowledge of which nodes are infected. The modern computer virus was conceived and verified by Fred Cohen in 1983. A computer virus reproduces by attaching itself to a normal program or document and take control of the execution of that program to infect other programs. Early viruses could spread slowly regularly by floppies, but the internet has made it much easier for viruses to move among computers and spread quickly.

A computer virus is a piece of software containing malicious code that can install, propagate and cause damage to the files or data stored in the computer without the knowledge of the user. There are different types of viruses and their characteristics, patterns and behaviors are all different. However, each virus has at least two fundamental characteristics, first, it is executable, second, it replicates itself. Without these two basic characters no virus can survive. There is a wide variety of virus symptoms that differ based on what type of virus you have and how far the virus has progressed. Some viruses alter data on the disk, some of them scramble the characters on the video display screen, some others display confusing messages to users, and some may consume computer resources thereby making the computer slow.

A virus infects others programs by attaching itself to those programs. Virus can transmit from one computer to another by using an infected file or infected disk. Some viruses are capable of transmitting themselves across networks and bypassing security systems. As it remains in the hard disk, it never goes even if the power is turned off. A virus can infect program files used for word processing, spread sheet, or operating system programs and document files such as windows.doc files that contain macros, information stored on disks by attaching to special programs in areas called boot records and master boot records.

Modern viruses may take advantages of network services such as World Wide Web, email and file sharing systems to spread. The internet consists of hundreds of millions of computers distributed around the world. Millions people use internet daily, taking full advantage of the available services at both personal and professional levels. The internet connectivity among computers on which the World Wide Web relies, however renders its nodes on easy target for malicious users who attempt to exhaust their resources or damage the data or create havoc in the network. Computer viruses, particularly in recent years, have increased noticeably in number. One of the most high profile threats to information integrity is the computer virus.

Generally a computer virus causes damage to the host machine. The damage can be done to a number of different components of the computer's operating and file system. These include system sectors, files, macros, companion files and source code. The always connected world of internet is a soft target for viruses. Viruses use internet connectivity to spread across the world faster and create havoc. A virus is basically an executable file which is designed such that first of all it should be able to infect documents, then it has to have the ability to survive by replicating itself and then it should also be to avoid detection. A computer virus reproduces by making copies of itself in the computer memory, storage, or over a network.

Signature scanning is a method of detecting a virus by scanning a target program to detect the presence of any virus signature. If the signature is found in the target program, then it is considered as infected otherwise the target program is considered as uninfected.

II. Related Works

The method used for spreading infection in a network depends on the parameters of diffusion process like outbreak thresholds and the effect of network structures [1]. Diffusion processes in networks are used to model active phenomena such as spread of information and social influence. Diffusion occurs in different contexts and usually involves a network of entities that work together in some way. Through these interactions, some property, e.g. information, idea, innovation, disease, etc., is diffused through the network. The diffusion process occurs on a network whose nodes represent individuals or organizations and edges represent interactions between nodes. In linear threshold model an individual is infected based on how many of their neighbors are already infected. There is a weight on the edge between two nodes, which defines a measure of influence. Each node has a threshold value, which is drawn randomly from some specified probability distribution. This threshold determines how many neighboring nodes have to be activated before the node itself becomes active. If the sum of the weights of all active neighbors exceeds the threshold, then the node will become active [2].

A network may contain multiple infection sources. For example, an infectious disease may be brought into a country through multiple individuals. Multiple individuals may collude in spreading a rumor or malicious piece of information in a social network [1]. To identify the infection source, we need to develop an infection spreading model. Here we use susceptible infected model for infection spreading. According to this model once a node has an infection, it spreads the infection to each of its neighbors after some random time. These spreading times are arbitrary, but we assume that they are independently and identically distributed. This model is widely used for developing or modeling viral epidemic. Inference of viral epidemic process in populations has been studied in [2], [4], [5], where various features related to the propagation of a viral epidemic, such as the rates of infection and the length of latency periods are investigated [1].

Different types of methods are used for detecting the presence of virus in a file or data stored in a computer [14].

A. Integrity Checking: It is the process of comparing the current state of stored data and/or programs to a previously recorded state in order to detect any change. Here an integrity checker records integrity information about important files on a disk, usually by check summing. If the content of a file is changed due to the activity of a virus, then the file will no longer match the recorded integrity information. The user is provoked, and can usually be given an option to restore the file to its pre-corrupted state. Integrity checking is the only way to determine whether a virus has damaged a file. The method detects the existence of viruses by comparing the hash values of a file with the hash value of its uninfected version. If no difference is found between the two hash values then the file is deemed to be uninfected.

B. Heuristic Scanning: This is a generic method of virus detection. It is used to detect new, unknown viruses in a computer that have not yet been identified. Heuristic methods are based on the piece by piece inspection of a virus looking for a sequence or sequence of instructions that differentiate the virus from normal programs. The advantage of this method is the capacity to detect known and unknown viruses, based on general characteristics shared by different viruses. One of the disadvantages of this method is that the virus writing community learns the rules used by heuristic software very quickly and starts writing viruses, which circumvent them. Virus writers can easily write viruses that don't follow the rules, making the current set of virus recognition rules obsolete. Heuristic scanning allows the computer to detect some viruses, which are not specifically known. It does not need any updates. Heuristic scanning is discovering only. In order to clean, it is necessary to know what changes the specific virus has made to the affected files.

C. Anomaly Based Virus Detection: This method is used for detecting noval attack. It focuses on normal system behaviors. In this method the system uses the collected heuristics to categorize an activity as normal or malicious. Even though probability of false alarm is comparatively higher in this method, it is more consistent because it is also capable of detecting new viruses. Anomaly detection includes two phase, they are training phase and detection phase. In training phase the behavior of the system is observed in the absence of attacks and machine learning techniques are used to create a profile of such normal behavior. In detection phase this profile is compared against the current behavior of the system and any deviations are flagged as potential attacks. The system is based on customized profiles, it is very difficult for an attacker to identify with assurance what action he can do without setting off an alarm.

D. Emulation Based Detection: The emulation based detection is an effective method where a virus is executed in virtual environment by emulating the instructions in the virus code. This type of detection is used to detect polymorphic as well as metamorphic viruses. The virus instance can be executed in the virtual environment in order to recognize instruction sequence or manners of the virus. The emulation method may not work well for

memory resident programs. This technique is able to detect the presence of encrypted virus. But this method is costly to implement.

E. Interception: Interception software detects virus like behavior and warns the user about it. Heuristics are used to detect virus like behavior. Interception is a good generic method to stop logic bombs and Trojan horses. Most of the interceptors are very easy to disable. When a Trojan is activated on your computer, it causes serious damage by deleting the files and destroying information on your system. Trojans are also known to create a backdoor on your computer that gives malicious users access to your system, possibly allowing confidential or personal information to be compromised. Unlike virus and worms, Trojans do not reproduce by infecting other files nor do they self replicate. A logic bomb is a programmed malfunction of a legitimate application.

III. Identifying Infection Sources

In this fast moving world, the importance of computer is increasing as they are used for almost every purpose. Most personal computers are now connected to the Internet and to local area networks which facilitate the spread of malicious code. With each day, computer viruses are becoming more and more complex, not easy to detect and avoid them. Hence we need a mechanism that protects files and other documents in the computer from very harmful viruses which can cause huge damage. In order to protect the computer systems from being attacked by viruses, suitable actions should be taken to guarantee internet security and firewalls must be installed on every computers.

We can define an infection sources either as a disease, virus or a rumor. A computer virus is nothing but a computer program that can infect a computer without the knowledge of the person using the computer. Virus can spread from one computer to another when an infected file is copied or transfer from host to target system. Static and dynamic detection methods are used for identifying the presence of virus. In static method a virus is detected by examining the files or records for the occurrences of virus patterns without actually running any code. In dynamic detection method, the virus code is detected by running the program and observing its behavior. The program monitors identified methods of virus activity including attempts to infect and avoid detection. This may include attempts to write to boot sectors, modify interrupt vectors, write to system files, etc. Often a virus requires a host and their goal is to infect other files so that the virus can live longer. Many viruses attempt to hide from being discovered.

Computer virus can attach themselves to just about any type of files and are spread as files are copied and sent from individual to individual. Some computer viruses have damage routine. This routine can deliver the virus payload. The payloads may display messages, but they can also destroy files, reformat your hard drive, or cause other kinds of damage. If the virus does not include a damage routine, it can create problems by taking up storage space and memory, and reduce the overall performance of your computer. Virus attacks are becoming more common, more frequent, and more severe. The increase in number of virus incidents is attributed to recent rapid growth in the number of internal networks and network connections, particularly relating to the internet and intranet. The more files that are shared by users, the greater the risk of users being infected by a virus.

A virus might rapidly infect every file on the individual computer or slowly infect the documents on the computer. A virus is simply a computer program. It can perform anything that any other program you run on your computer can do. Some viruses are designed to intentionally damage files, and others may just spread to other computers. The three main components of a computer security include secrecy, accuracy and availability. Computer viruses usually have an impact on two of these, they are accuracy and availability. A virus may modify programs and data stored in the system, and sometimes, your machine may not work when you require it. Viruses use various techniques to hide from users and increase the time they have to spread freely. A computer virus is a malicious program that can spread across networks by making copies of itself, usually without the user's knowledge. Viruses can have dangerous side effects. These can vary from displaying irritating messages to deleting all the files on your computer. A virus program has to be run before it can infect your computer. Viruses have ways of making sure that this happens. Viruses can attach themselves to other programs or hide in code that is executed automatically when you open certain types of files. The virus can copy itself to other files or disks and make changes on your computer.

A computer virus is somewhat similar to a biological virus. Initially a biological virus infects a cell in the body and then starts to replicate itself to the entire body. Similarly a computer virus is a segment of machine code that will copy itself into one or more host programs, when it is activated. When these infected programs are run, the viral code is executed and spreads further. The viral code is executed before the code of its infected host. The structure of a virus includes four parts. They are search, copy, anti detection and payload. A computer must run a virus for it to become active. Viruses contain a search routine, which looks for programs that are regularly executed. So that the virus becomes active as soon as possible.

The search routine identifies how quickly a virus can reproduce and which types of programs it can infect. Once a virus has identified a program to infect, it must copy itself into that program. Virus contains a

copy routine to avoid detection by antivirus software. The copy routine appends the virus code to the original host program code or overwrites some or all of the original code, to infect the chosen program. The size of the copy routine depends on the structure of the program that the virus is intended to infect, executable files with the .exe extension, for example, have a complex file structure and require larger copy routines than some other file types. Virus creators usually design search and copy routines to prevent viruses from being identified by antivirus software, but the creators also include specific anti detection routines.

An anti detection routine may keep the last modified date of an infected file as same as to the original uninfected file. Search, copy and anti detection routines allow a virus to spread or reproduce, effectively, but a virus may contain payload routines, which actually perform the function or functions for which the virus was designed. All viruses do not contain a payload routine. If a payload routine is present, it may create annoyance, destroy data, etc.. A virus may execute a payload routine immediately or once again, wait for a set of predetermined conditions to occur before triggering the routine.

The figure 1 given below shows the functional diagram of a virus.

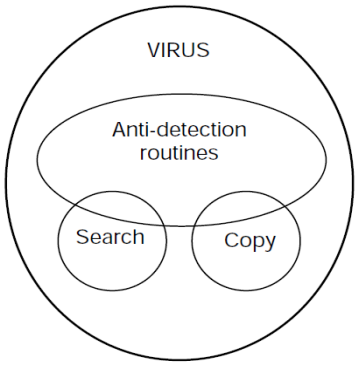

Fig 1: Functional block diagram of a virus

The necessary components of a computer virus include search, copy, and anti detection routines. Other routines can be added to the top of the basic three routines to prevent normal computer operation, to cause damage, or to play practical jokes. Viruses are written in assembly language.

In order to determine the presence of a virus we use signature scanning method. For performing signature scanning we need a virus scanner. A scanner works by reading the data from disk and applying signature matching operation against a list of known virus signatures. If a match is found for a virus signature, then we can conclude that the data is infected. Scanners can recognize viruses that have not been executed yet. A virus signature is a string of characters or numbers. One signature may contain several virus signatures. Signatures are algorithms that uniquely identify a specific virus. Different types of virus may share a single signature, allowing a virus scanner to identify viruses it has never seen before. The virus signature is like a fingerprint in that it can be used to detect and identify specific viruses.

Signature scanning is the most popular method of virus scanning. It is capable of detecting more than 80% of viruses. It uses a simple logic and uses less memory and system resources. It is able to detect encrypted viruses too. A signature database displays a major role in signature scanning. This is a collection of bit patterns found in files infected by a range of known viruses. The scanners check each file for the existence of any of the virus signatures in the database. If a signature is found, then the scanner checks for existence of another signature, before taking a decision that the virus is present. Signature scanning is the one of the most effective techniques to detect a virus. The notification and remediation process is much more effective when the malware has been positively identified and the remediation can be customized for specific malware.

Simple computer viruses copy themselves to each executable files they infect. These types of viruses replicate an identical copy of themselves byte by byte each time it infects a new file. These types of viruses can be easily detected by searching a specific string of bytes, called virus signature, which has been extracted from the virus. A virus signature is a sequence of bytes that may be found in a virus program code but unlikely to be found elsewhere. A signature has to be very carefully extracted as a wrong signature may detect uninfected files as infected. Extracting a virus signature is a practiced job and usually done by an anti virus researcher after carefully analyzing an infected file or viral body. Signature scanning is not only capable of detecting simple viruses it can also detect encrypted viruses. An encrypted virus includes a decryption routine and an encrypted viral body. When the virus is executed first the decryption routine gains control. The decryption routine decrypts the encrypted viral body and transfers control to the decrypted viral body. Then the decrypted viral body does its mischief. These types of viruses can also be detected by searching for a signature from the unchanging decryption routine.

A scanner will search all files in memory, in the boot sector and on disk for code snippets that will uniquely identify a file as a virus. For this purpose, it requires a list of unique signatures that will be found in viruses and not in benign programs. To avoid false alarms, most scanners will check the code of a suspected file against either the virus code itself or a checksum of it. Checksum is a method used to verify that if the data has been changed or not, it involves summing all of the bits in a file. This is the most common method used for virus detection. Signature based detection is fast and accurate since the chances of false alarms are very low. The requirement of the system is to have an updated database of all signature files of malware. The accuracy of this method is dependent on the signature database of the system.

A common technique that virus writers use to avoid detection is to enable the virus to change itself to by having some kind of self modifying code. This type of virus is usually known as metamorphic virus and can be particularly hard to detect. In order to replicate itself, a virus must be permitted to execute code and write to memory. Because of this reason, many viruses attach themselves to executable files that may be part of lawful programs. If a user attempts to launch an infected program, the virus code may be executed concurrently. One of the most commonly used method for the detection of computer virus is the virus scanner, which uses signatures to identify a specific virus in executable files, boot records or memory. Viruses will not damage your system as long as they are coded correctly. Any system damage resulting from a purely replicating virus happens because of the bugs in the code that clash with system configuration.

The most observable advantage of virus scanner is its ability to detect potentially harmful virus, including trogen horse, worms and email virus. Once a virus is detected, the scanners allow the user to quarantine the virus to prevent their damage or if possible to eradicate it completely. Scanners are very good at detecting virus that they have the signature for. The scanner utilities are mainly classified into two. They are on demand scanners and memory resident scanners. They perform a thorough scan of all executable files on a disk, searching for known viruses. Memory resident scanners are loaded into memory when the computer is turned on, and they continue until the computer is turned off, scanning files as they are accessed. If a virus is detected in a program or files stored in the system, the scanner reports it and warns the user. A memory resident virus scanner constantly runs in the back ground of the computer and automatically cheeks for viruses. Memory resident scanners offer better protection than regular scanners. Since memory resident scanners can identify a virus before a virus can do severe damage.

Computer viruses are not naturally destructive. The essential feature of a virus is not its ability to damage the data, but its ability to gain control of the computer and make a fully functional copy of itself. A virus can reproduce itself. When a virus is executed, it makes one or more copies of itself. The computer programs that are destructive may not be considered as viruses because they do not all reproduce. Similarly all viruses are not destructive in nature because reproduction is not destructive. The computer virus avoids the detection of operator control by hiding itself in other programs. Thus the virus gains access to the CPU by run programs that it happens to have attached itself to without the knowledge of the user. Most viruses are pretty harmless. The user may not even notice the virus for years. Sometimes virus cause random damage to the programs and over a long period they might destroy files and disks.

The figure 2 given below shows block diagram of a signature scanning.

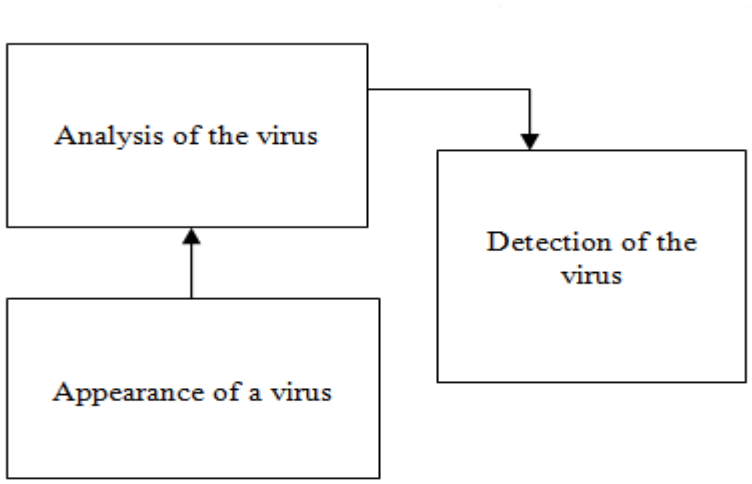

Fig 2: Block diagram of signature scanning

IV. Conclusion

We know that finding an infection source is a challenging problem. An infection source can be either a virus, disease or a rumor. An infection source can cause serious damage to a computer, because when a computer or network is infected by an infection source, it takes the control over the system and then perform the malicious activities like reformatting the disk, deleting important files etc. Here we take an infection source as the virus. For detecting the presence of a virus, we use signature scanning method. Signature scanning is the most popular method of virus scanning. It is capable of detecting more than 80% of viruses. A virus signature is a string of characters or numbers.

References

[1] Wuqiong Luo, Wee Peng Tay, and Mei Leng, "Identifying infection sources and regions in large networks," *IEEE Trans. Signal Process.*, vol. 61, no. 11, Jun. 2013.

[2] C. Moore and M. E. J. Newman, "Epidemics and percolation in Small-world networks," *Phys. Rev. E*, vol. 61, pp. 5678–5682,

[3] M. E. J. Newman, "Spread of epidemic disease on networks," *Phys. Rev. E*, vol. 66, no. 1, pp. 016 128+, Jul. 2002.

[4] P. D. ONeill, "A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods," *Mathematical Biosciences*, vol. 180, no. 1-2, pp. 103 – 114, 2002.

[5] A. Ganesh, L. Massouli, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proc. IEEE INFOCOM 2005*, 2005.

[6] N. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*. Griffin, 1975.

[7] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?" *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5163–5181, 2011.

[8] L. Han, S. Han, Q. Deng, J. Yu, and Y. He, "Source tracing and pursuing of network virus," in *Proc. 8th IEEE International Conference on Computer and Information Technology Workshops*, 2008, pp. 230–235.

[9] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, 2004.

[10] G. Brightwell and P. Winkler, "Counting linear extensions is #P-complete," in *Proc. 23rd Annual ACM Symposium on Theory of Computing*, 1991, pp. 175–181.

[11] United States. Federal Energy Regulatory Commission and North American Electric Reliability Corporation, *Arizona- Southern California Outages on September 8, 2011: Causes and Recommendations*, 2012.

[12] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

[13] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world' networks." *Nature*, vol. 393, no. 6684, pp. 440– 442, 1998.

[14] G. Hu, "Robust consensus tracking for an integrator-type multi-agent system with disturbances and unmodelled *International Journal of Control*, vol. 84, no. 1, pp. 1–8, 2011.

[15] L. J. Allen, "Some discrete-time SI, SIR, and SIS epidemic, models," *Mathematical Biosciences*, vol. 124, no. 1, pp. 83 – 105, 1994.

[16] Y. Wen, G. Shi, and G. Wang, "Designing an inter-cloud messaging protocol for content distribution as a service (CoDaas) over future internet," in *International Conference on Future Internet Technologies*, 2011.

[17] G. Hu, "Robust consensus tracking of a class of second-order multi-agent dynamic systems," *Systems and Control Letters*, vol. 61, no. 1, pp. 134–142, 2012.

[18] G. Hu, "Robust consensus tracking for an integrator-type multi-agent system with disturbances and unmodelled dynamics," *International Journal of Control*, vol. 84, no. 1, pp. 1–8, 2011.

[20] M. Kimura, K. Saito, R. Nakano, and H. Motoda, "Extracting influential nodes on a social network for information diffusion," *Data Min. Knowl. Discov.*, vol. 20, pp. 70–97, 2010.

[21] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a network," in *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 137–146.

[22] S. Aldalahmeh and M. Ghogho, "Robust distributed detection, localization and estimation of a diffusive target in clustered wireless sensor networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011.