Computer viruses

S.R.Subramanya
Natraj Lakshminarasimhan

© PHOTO DISC/J. NISHINAKI/COMP. MKC

Computer viruses generally refer to programs that unintentionally get into computers, disrupt the normal operation, and cause damage to data and programs. However, not all programs that cause damage are real viruses. For instance, *Worms and Trojan horses* are two common types that are not viruses: A worm finds its way into system registries and spreads as an independent program, while the Trojan horse is run unwittingly by the user. Thus, a virus is best identified not by what it does, but by how it spreads and infects other programs.

Computer viruses are similar to biological viruses. A biological virus enters a body, damages the body, spreads to other bodies, and eventually is eradicated by the internal immune system or by external means. Similarly, the computer virus enters the computer system and gets attached with a program (or set of programs or applications). As the application(s) is invoked, the virus becomes activated and spreads to the other parts of the system. By attaching itself to another program, a virus gains a host. When the host program is executed, the virus is executed. Each infected program in turn infects other programs, and the virus spreads.

Viruses can be either benign or destructive. If they are benign, they cause no real damage, but may produce harmless changes like disrupting the display on a monitor, or making some sounds or false alarms. However, if they are destructive, the viruses can cause real damage to the system such as hogging disk space and/or main memory, using up CPU processing time and introducing the risk of incompatibilities and conflicts. Over 40,000 different viruses have been cataloged so far. In recent years, the number of viruses unleashed has increased dramatically and the extent of the damage (in lost data, time and productivity) is estimated to be several billions of US dollars per year.

Types of viruses

This section presents a broad classification of viruses. Most viruses are in fact "hybrid" combinations of various properties from multiple classes.

File infectors. These most often attach to program files, but can infect any file with executable code, including script files or program configuration files. When the program, script or configuration is executed, the virus is executed as well.

System or boot-record infectors. System or boot-record infectors do not necessarily infect a file. They target, instead, certain areas of a hard disk used exclusively for system processes. These areas include the boot-record, which is a section of the disk dedicated to booting the operating system. On a diskette, these infectors attach themselves to the DOS boot sector; on a hard disk, they attach themselves to the Master Boot Record. Having infected a Master Boot Record, the virus spreads to the boot sectors of the inserted media.

Multi-partite viruses. Multi-partite viruses infect boot records as well as files. With its hybrid nature, a multipartite virus inherits the worst qualities of each of its parents, and consequently is far more contagious and destructive than either.

Macro viruses. Macro viruses infect macro-enabled documents, especially in the Microsoft(r) Office suite of applications-more specifically, Microsoft(r) Word and Excel. When opened, an infected document executes a macro automatically, or the user does so accidentally. The macro inflicts damage and then infects other documents on the disk. A macro is a set of executable commands designed to run in place of a repetitive task. Although macros are not unique to Microsoft products, it is through Microsoft products that the vast majority of macro viruses spread. Macro viruses are the most common type of viruses, however they do fairly little damage.

Stealth viruses. Stealth viruses use many techniques to thwart detection. One technique is to redirect the addresses within a program that point to other programs or system information, and have them point to the virus file instead. When the program calls for that supplementary program or system information, it actually runs the virus code. This infects the file without actually injecting additional code, which could show up as a symptom to virus scanning software. Another common stealth technique changes a file, but displays its size as it was before infection. Thus, it nullifies the ability to use the file length as an indicator of infection.

Encrypted viruses. Encrypted viruses enjoy the advantages of other encrypted material. Initially, encrypted viruses appear not as viruses, but as nondescript gibberish. But when an infected program is executed, a small piece of plain, unencrypted code decrypts the rest of the virus, which

0278-6648/01/$10.00 © 2001 IEEE

then proceeds to do its damage. When, and if, an encrypted virus is detected, it is very difficult to analyze since it is not subject to reverse engineering like the unencrypted viruses. This makes it hard to determine the structure of the virus and the precise scope of its payload. Encryption is most useful when coupled with a polymorphic strategy.

Polymorphic viruses. Polymorphic viruses try to evade detection by altering their structure or the encryption techniques. Each time an infection occurs, a polymorphic virus changes its form, confusing virus (detection) scanning software. Because virus scanners use certain unique "signature" characteristics to identify viruses, any virus that changes its form presents a formidable new challenge.

Stages in the life of a virus

During its lifetime, a virus typically goes through the following four stages:

Dormant phase. In this phase, the virus is idle and will eventually be activated by some event such as the occurrence of a certain date or the presence of another program or file. For instance, according to the system date, the virus gets activated.

Propagation phase. In this phase, the virus places an identical copy of itself into other programs or into certain system areas on the disk. Each infected program will now contain a clone of the virus, which will itself enter a propagation phase.

Triggering phase. At this point, the virus is activated to perform its intended function. As with the dormant phase, the triggering phase can be caused by a variety of system events, including a count of the number of times that this copy of the virus has made copies of itself

Execution phase. The actual function of the virus is performed in the execution phase. The function may be harmless, such as displaying some messages on screen, or damaging, such as destroying programs and data files.

Naming conventions

Vendors of anti-virus software name viruses based on their types. Even so, virus naming conventions are sometimes confusing, especially to someone simply looking for a quick fix. Since as much information as possible commonly goes into a name, virus terminology often seems lengthy and cumbersome. To further complicate matters, antivirus vendors tend to have their own

A small sample of viruses and their actions

New Zealand (Stealth boot-record infector)

Upon infection, this virus upon gives the messages "Your PC is now Stoned!" and "LEGALISE MARIJUANA!" It may also damage the associated floppy disks. This virus was originally created by a New Zealand student.

WM/Wazzu (Word macro virus)

Each time a document is infected, a few words in the document move to a different location, and the word "wazzu" is inserted in random places. The word "wazzu" can be removed with a simple search and replace command, but the relocated words must be carefully sought out.

WM/Concept (Word macro virus)

This virus spreads much like WM/Wazzu by automatically executing in an opened document that has been infected. The virus displays a pop up message box with the single digit "I." The virus is completely harmless. It was the first macro virus ever discovered, and is still common.

Michelangelo (Boot record infector)

This virus overwrites the first 17 sectors of a hard disk. It was named Michelangelo because it was triggered on Michelangelo's birthday: the 6th of March.

I Love You (Macro virus)

This virus could wipeout music and image files, and send copies of itself to contacts in a user's address book once the mail is opened. Replication is what made the virus spread so fast. The virus also launched what may be the biggest Distributed Denial of Service (DDOS) attack on a Website.

Bubble Boy (Macro-Microsoft mail client virus)

Unlike previous viruses, this does not need an accompanying attachment to be opened to unleash the virus. Bubble Boy does not come embedded in an attachment, and has the ability to infect a computer just by the act of a user reading the e-mail. It works on Microsoft(r) Outlook, and the virus requires that the recipient open the e-mail; it will not run if the e-mail is only viewed through the Preview pane.

Melissa (Macro-Microsoft mail client virus)

Melissa macro virus gets activated through Microsoft(r) Word 97 or Word 2000 and the Microsoft(r) Outlook e-mail program. When it infects a computer, it sends copies of itself along with whatever document is open at the time to the first 50 people in the computer's e-mail address book. It also infects all other Word documents that are subsequently opened.

Code Red

Code Red virus exploits a security hole in the Microsoft Internet Information Server (IIS) to penetrate and to spread. It also disables the system file checker (SFC) in Windows. It probes random IP addresses to spread to other hosts. During a certain period of time, it only spreads. It then initiates a Denial-of-Service (DoS) attack against a government website, and finally suspends all activities. This repeats every month. Code Red II is a variant of Code Red and targets Microsoft IIS Web servers, and does not pose a threat to end users.

Nimda

Nimda is a complex virus with a life cycle consisting of four parts: a) local file infection, b) mass mailing, c) Web server infection, and d) LAN propagation. Nimda locates EXE files on local machines and infects them. It locates email addresses (via MAPI from email client and local HTML files) and sends an email to each of the addresses, containing an attachment called README.EXE. It also locates Web servers and infects them by using several known security holes. It modifies web pages such that surfers of that site get infected. It will also search and infect files shared in the local network.

individual naming conventions. (And all vendors agree on a very limited number of points.) Thus, two anti-virus software vendors may call the same virus by two radically different names.

Thankfully, certain prefixes tend to show up frequently. W95 means Windows 95, while W97M means Word 97 Macro. WM just means Word Macro. If you know you are looking for a Word Macro virus, expect to find WM or W97M in the name. A few other common prefixes include XM (Excel Macro) and AM (Access Macro), as well as W32 (Windows 95 or later) and WNT (Windows NT only).

Virus generations

When the first viruses were written, they could be classified into five generations of viruses. Each new generation of viruses has incorporated new features that make the viruses more difficult to detect and remove.

First generation (simple). The first generation of viruses consisted of simple viruses. These viruses did nothing very significant other than replicate. In the case of boot sector viruses, this could cause a long chain of linked sectors. In the case of program-infecting viruses, the repeated infection might result in a continual extension of the host program each time it was re-infected. Many of the new viruses being discovered today still fall in this category.

Damages from these simple viruses are usually caused by bugs or incompatibilities in software that were not anticipated by the author of the virus. First-generation viruses do nothing to hide their presence on a system, so they can usually be found simply by noting an increase in the size of files or the presence of a distinctive pattern in an infected file.

Second generation (self-recognition). First generation viruses repeatedly infect the host, which leads to depleted memory and then early detection. To prevent this unnecessary growth of infected files, second-generation viruses usually implant a unique signature that signals that the file or system is infected. The virus will check for this signature before attempting infection. If the signature is not present, it will place the signature soon after the infection has taken place; however, if the signature is present, the virus will not re-infect the host.

Third generation (stealth). Most viruses may be identified on a contaminated system by scanning the secondary storage and searching for a pattern of

A brief chronology of computer viruses

1986 First PC virus was created and termed the *Brain* virus. The virus was created in Pakistan and is a boot sector virus, i.e., it affects only the boot records. It falls under the stealth virus category.

1987 First memory resident file infector was discovered in Lehigh University and named *Lehigh*. Attacks executable files.

Jerusalem virus first appeared at the Hebrew University, Jerusalem. It's another memory resident file infector.

1988 First anti-virus was *Den Zuk* created in Indonesia. It was designed to detect and remove the Brain virus and immunize disks against a *Brain* infection.

Cascade Virus is found in Germany. It is an encrypted virus, meaning it was coded so that it could not be analyzed easily.

1989 *Data Crime* virus is on the loose and strikes on Friday the 13th.

Dark Avenger virus, attacks slowly, so that it goes unnoticed.

Fredo virus discovered in Israel. It is the first full-stealth file infector.

1990 Many anti-virus products are introduced including IBM's McAfee, Digital Dispatch, and Iris.

Viruses combining various characteristics spring up like the Polymorphism and Multipartite.

1991 Symantec releases *Norton* antivirus software.

Tequlia, a stealth, polymorphic and multi-partite virus is found.

1992 Media mayhem greeted the virus *Michelangelo* that March. Predictions of massive disruptions were made, and anti-virus software sales soared.

1994 A virus called *Kaos4* was posted on a pornography news group file. It was encoded as text and downloaded by a number of users.

Virus *Pathogen* appeared in England; the writer was tracked down by Scotland Yard's computer crime unit.

1994 The *SatanBug* virus appears; The anti-virus industry helps the FBI find the person who wrote it-a kid.

Cruncher was considered a good virus as it compressed infected programs.

1995 Anti-virus companies worry about staying profitable with the emergence of Windows 95 because the boot viruses cannot infect it, hence the possible catastrophic loss of business. But the macro viruses that do infect the Windows-95 applications soon appear keeping the anti-virus companies happy and in the green.

1996 *Concept*, a macro-virus, becomes the most common virus in the world.

Laroux is the first virus to infect Microsoft® Excel spreadsheets.

1997 The Anti-Virus Research Center in Cupertino develops an exclusive technology called Bloodhound-Macro to address the growing number of new and unknown macro viruses.

1998 Posing as an anti-virus software, an e-mail attachment virus infects computers worldwide.

1999 The *Melissa* virus, a macro, causes worldwide destruction. Primarily infecting Microsoft(r) Word and Outlook, it automatically sends mail to everyone in the user's address book.

2000 *I Love You* virus causes havoc; it is transmitted via e-mail and, when opened, it automatically sends mail to everyone in the user's address book.

2001 *Code Red* virus exploits a security hole in Microsoft Internet Information Server (IIS) to spread. It disables the system file checker (SFC) in Windows. It probes random IP addresses to spread to other hosts.

2002 *Nimda* is a complex virus with a mass mailing worm component, which spreads itself in attachments named README.EXE. Nimda affects EXE files on local machines, locates email addresses and spreads itself. It also locates Web servers and infects them.

data unique to each virus. To counteract such scans, some resident viruses employ stealth techniques. These viruses subvert selected system service call interrupts when they are active. Requests to perform these operations are intercepted by the virus code. If the operation exposes the presence of the virus, the operation is redirected to return false information.

Fourth generation (armored). As anti-virus researchers have developed tools to analyze new viruses, authors have turned to methods to obfuscate the code of their viruses. This "armoring" includes adding confusing and unnecessary code to make it more difficult to analyze the virus code. These viruses may also perform directed attacks against anti-virus software, if present on the affected system.

Fifth generation (polymorphic). These most recent class of viruses to appear on the scene are also called self-mutating viruses. These viruses infect their targets with a modified or encrypted version of themselves. By varying the code sequences written to the file, or by generating a different, random encryption key, the virus in the altered file will not be identifiable through the use of simple byte matching. To detect the presence of these viruses requires that a more complex algorithm be employed that, in effect, reverses the masking (decryption) to determine the presence of the virus.

Virus detection

Generally, anti-virus software can detect all types of known viruses, and it needs to be updated frequently for its effectiveness. Basically, there are four means of virus detection: signature-based scanning, emulation, heuristics, behavioral analysis and check summing.

Signature-based scanning. This scheme searches for unique strings of code, i.e., the virus's signature specific to particular viruses. When this string of code is found, the file is declared infected. Relying on signatures for detection poses two problems. First, users must update their virus scan software frequently in order to ensure the latest possible protection is installed. Secondly, signature-based scanners are only effective in identifying known viruses.

Emulation. This mimics the execution of the infected file to determine any malicious intent. Essentially, the file is contained in what is referred to as a sandbox or virtual environment. In plain language, the file is tricked into believing its interacting with the operating system, when in fact, it is not. Emulation can be time-consuming and result in a noticeable performance slowdown.

Heuristics. These attempt to detect unknown viruses and often employ generalized signature scanning geared to detect families of viruses. If the virus is related to a known family, heuristics will detect it and report it as suspicious or infected with an unknown virus. Heuristics also rely on emulation or a combination of signatures and emulation. Due to heuristics' penchant for false positives (identifying a clean file as infected) and performance concerns, many vendors have suppressed the level of heuristics employed. As a result, only a very small number of products have gained a track record for detecting previously unknown threats.

Behavioral analysis. This monitors the execution of the file and gives the user an opportunity to either prevent or undo any proposed or taken action. For example, if a file attempts to write to the system registry, the action can be blocked, either by the user or automatically, depending on configuration. In many respects, behavioral analysis and emulation are closely related, though behavioral analysis often lets the file execute in real-time, stopping it only when suspicious behavior is detected. This method overcomes the performance slowdown side effect of emulation.

Check summing. This is essentially a count of bits that is used to verify file integrity. An initial scan of system files is performed; the numerical quotient for each file is derived and stored in a database. When subsequent scans are performed, the program checks the database to ensure that the numerical quotient matches. If it has changed, the file is considered suspicious and/or infected.

Conclusions

Computer viruses have been around for over fifteen years. Over 40,000 different viruses have been cataloged so far. In recent years, the number viruses have increased dramatically. The damages they cause are estimated to be several billions of U.S. dollars per year. Most often, the origin of the virus is difficult to trace. Various kinds of anti-virus software have been developed which detect viruses and take corrective actions.

The anti-virus software needs to be continually updated to cope with newer types of viruses. The proliferation of the Internet and Web, have enabled viruses to spread quickly on a massive scale, by taking advantage of several security loopholes. The continual challenge is to have quick and effective responses to these virus attacks.

Read more about it

- F.B. Cohen, *A Short Course on Computer Viruses,* 2nd ed. New York: Wiley, 1994.
- D. Harley, U.E. Gattiker, and R.Slade, *Viruses Revealed.* New York: Osborne/McGraw-Hill, 2001.
- R. Skardhamar, *Virus: Detection and Elimination*. New York: Academic, 1996.

Websites

- This is the IBM research group Website on Anti-Virus, which provides information on virus attacks, and also on the virus time-line: http://www.research.ibm.com/antivirus/
- This Website provides information on various popular viruses: http://www.datafellows.com
- This is the Candian discovery channel's online Website, which provides information on virus attacks, and also on the virus time-line:http://exn.ca/Nerds/20000504-55.cfm
- This anti-virus vendor Website provides good information on viruses and how to prevent them: http://www.symantec.com/
- This is Mcafee's anti-virus library, which provides information on the latest and most popular worldwide viruses: http://vil.mcafee.com/default.asp?

About the authors

S.R. Subramanya obtained his Ph.D. degree from George Washington University, where he received the Richard Merwin memorial award from the EECS department in 1996. He received the Grant-in-Aid of Research award from Sigma-Xi for his research in audio data indexing in 1997. He is currently an Assistant Professor at the University of Missouri-Rolla. His research interests are in Multimedia Systems and Computer Security.

Natraj Lakshminarasimhan obtained his bachelors degree from the University of Madras. He is currently a graduate student in the Computer Engineering Department at the University of Missouri-Rolla. His interests are in Computer Networks and Security.