SESSION XIII TUTORIAL: COMPUTER VIRUSES Walter Schneider, Presider University of Pittsburgh Computer viruses: What they are, how they work, how they might get you, and how to control them in academic institutions WALTER SCHNEIDER University of Pittsburgh, Pittsburgh, Pennsylvania A computer virus is a program that replicates itself and spreads to computers with the goal of disrupting or destroying normal computer use. In academic computing, viruses represent a serious problem that costs millions of dollars in losses annually and hinders the free exchange of information so critical to education. Viruses operate in incubation, infection, and destroy phases. The nature, mechanisms, and preventive measures for personal-computer viruses are reviewed. Different procedures are recommended to protect research laboratories, instructional laboratories, and software lending libraries. Tradeoff'sbetween providing adequate protection and not having the security become too burdensome are considered. Computer viruses are programs that replicate themselves to spread to other computers; they have the potential of altering the behavior oftheir computer hosts. They can destroy research and instructional data and computer equipment, and they can easily be spread by honest, unknowing individuals, who are themselves using the host computers appropriately. Researchers need to take basic steps in order to prevent any catastrophic loss of data due to computer viruses, because universities, which typically encourage free exchange of information among many individuals, unfortunately thus make it very easy for computer viruses to do extensive damage. This paper provides a tutorial on what computer viruses are and how one can deal with them in academic settings. A complete description of computer viruses can be found in R. Roberts's (1988) book on the topic. A computer virus can affect any laboratory in which honest individuals are using programs imported from other sources. Viruses can be spread through the normal use This work was supported in part by Office of Naval Research Contracts N ware backto the University of Delaware, wherethe virus spread through the medicalcenter and on to the University of Delaware in general. It is also believedthat someonefrom the University of Pittsburgh whouseda computerat the University of Delaware imported the same virus back to the University of Pittsburgh, where it then replicated itself in the university's public laboratories. One of the students who was operating a computer in the psychology department's laboratorytook a floppydisk from the undergraduate lab and ran it on one of the public sites (perhapshe was doing word processing both at the public sites and on the laboratory computer). Thiswas, of course, a totally legitimate use of computers on campus. Unfortunately, while the student was word processing at the public site, the virus attached itself to the student's copy of the wordprocessing program. Whenthe student returned hisfloppy disk to the psychology lab, the virus attacheditself to the operating system on a lab computer. The virus spread withinour laboratorywhendata from all of the computers were merged on one master file in themain computer. Afterseveral daysof replicating itself, the virus beganto erase the disks of the computers in the undergraduate laboratory. With the exception of the very first activitythat occurred in Pakistan,probablyall of the other activities that enabled the virus to spread resulted from honest individuals' appropriate use of computers. The net effect of the virus attack was the destruction of several months' worth of data collected in the undergraduate labs. When the virus destroyed the data from 120 students from my laboratory class, I was more infuriated than I have ever been in my academic career. It was as if someone had broken into my office and gone through my filing cabinets destroying all my data. Fortunately, because the data had been backed up, after several days of work the lab wasfunctioning again. This type of spreading of a computer virus can and probably will occur in any laboratory that allows disks to corne in from the outside. It is important to take precautions to reduce the virus threat. One should think of controlling viruses as one thinks aboutthe securityof one's horne. Almostany horne can be broken into even whenextreme securitymeasures have been taken. Most people use basic security measures, suchas lockingtheir doors, to makeit at leastsomewhat difficult for a would-bethief. Such basic measures inhibit robberiesenough so that they are infrequent, and we can proceed with our lives relatively unincumbered by either robberies or extreme security measures. But if robberies become more of a problem, one may have to considermore extensive measuresagainstintrusion(such as installing a security system that requires one to enter passwords whenever entering or leaving the premises). One must trade off ease of access against security. Fortunately, however, a few simpleprocedures can provide protection from most viruses. It is important not to become paranoid about the virus problem, but rather to choose an appropriate level of security that will allow COMPUTER VIRUSES 335 computers to accomplish one's tasks while the virus problem is kept in check. What Is a Computer Virus? A computer virus is a program that installs itself upon a systemto infectand/or destroy (or alter) othersystems. It is very important to understand the characteristics of a virusso thatone may reducethe likelihood of itsspreading. A virus is an executableprogram that attaches itself to other programs in order to spread. A simple example would be a virus that alters a computer's operating system so that whenever the system is started up (booted), the virus code will be executed. The virus then examines other programs that can carry it (e.g., executable programson any floppy disksinsertedintothe machine), and it will reinstall itself on floppy disks, which may travel to other computers.It can then installitself on other systems, whenever the infected programs on the disks are run at new installations. The virus threat is very real. The National Security Agencyof the UnitedStateshas estimatedthat over 40% of the nation's college campuses have been hit by computer viruses. It does not take an exceptional ability at programming to writea new virus; onlyaboutthreecomputer courses and some detailed reading will suffice. A single individual almost anywhere in the world can thus inflict damage in hundreds of countries. In the future, there will be more viruses, and they will be more dangerous. We may even find academic terrorists targeting academicdepartments (e.g., animal rightsgroupstargeting programsthat collectanimaldata). Disillusioned students may injectviruses to disrupt classesso that they do nothaveto tum inassignments (similar to the way "bomb scare" reports became a problem in the 1960s in the United States). There are three phases to the operation of a computer virus; they reflectmetaphorical similarities between computer scienceand biology. The first phaseisincubationstaying dormant for a period of time. A computer virus can remain dormant, doing nothing, for an extended period. For example, it might only replicate itself after a certain number of starts of the operating system (e.g., every 50threboot). An incubating virusis thuslikea mole in a spy network. It sits there and operates normally for a long time, so that nobodysuspects that it is there. Users are frequently suspicious of newprograms thatcausetrouble on their computers, so that a virusthat wouldimmediately alter the operation of a computer might quickly be detected. A virus that would allow normal operation for severalmonths, however, and onlythen beginto alter the operationsof the system, wouldbe more likely to go undetected. Note that there is virtually no way to detect a virus while it is in its incubation phase. Unless one has a copy of the program before a virus has hit it, or particular signature information for a specific virus, there is no way to detect a virus during this period. The secondphase of a virus is infection, during which the virus tries to replicate itselfand spread to more com-puters. During the infection phase, the virus program tries to identify new host programs and install itself on them. A sophisticated virus will install itself on other programs without doing damage, so that it can spread before it is detected. Typically a virus will install itself on the boot block or operating system, or on executable programs (e.g., .COM, .EXE, .SYS, .BAT, or overlay files on IBM-compatible computers). Note that a virus can only spread by installing itselfon executable programs. A virus cannot be spread by modifying text or data files. A virus may install itself on a boot block or the operating system (e.g., COMMAND.COM file). During the infection phase, a virus can be detected. For simple viruses, this is done by means of identifying the change in the date or the length of existing files. Complex viruses may alter a file without changing its date or length; these viruses can typically be detected if there are changes in the check sum of individual files. However, the check sum requires a signature file thar records the check sums of the files to be recorded. In the third phase of a virus, the destroy phase, the virus destroys or modifies the operations of the host computer system. The destroy phase typically occurs after a period of time during which the virus has spread. For example, one virus might allow the system to be rebooted 20 times before it would shift from the infection to the destroy phase. The typical user might run a system for nearly a month of infecting other disks before the virus would go into the destroy phase. Many different malicious activities have been carried out during the destroy phase. A benign virus might put up a scare message or lock the system. For example, a program might put on the screen "beware of the virus," and not allow any other actions by the user. Some viruses will destroy all data and programs (e.g., the BRAIN virus will erase the hard disk). Note that a virus can destroy data passing through a machine, as well as destroy data on the system disk. For example, a virus may destroy just a subset of data files and then only destroy those files on the backup floppy disks. Such a virus could destroy a person's complete dissertation work. The student, for example, might run an analysis and find that the data file can no longer be read by the analysis program. The student might then take the backup data, stick it into the disk, and copy that data file back onto the disk. However, at this point the virus could attack the original data file, destroying it while the copying takes place. In this way, both the original file and all backups of the data file could be lost. One should remember that a virus can attack files even if they are not being read or written by the user. Some viruses consume computer resources. For example, a virus might execute a simple loop repetitively on each interrupt from the time-of-day clock, thereby consuming 50% of the computer's capacity. Another virus might create hidden files that use up disk space. At least one virus has been reported as destroying computer hardware. This virus could continuously move the disk back and forth from its minimum to its maximum travel. The hardware was not designed to withstand such continuous movement; it resulted in the heating up ofthe disk coil or motor, and it was claimed to have started fires within the computer. Viruses have also been known to alter the writable control store. Computers often contain a small amount of nonvolatile memory, which typically encodes information such as the kinds of disks or special devices that are attached to the computer. Ifa virus writes this writable control store, it can alter the machine so that it doesn't even recognize the hardware that exists in it, and the user must then reconfigure the system completely. Since the original configuration may have been made in the factory, most users can be at quite a loss when having to fix the writable control store. Perhaps the most dangerous mode of destruction occurs when viruses alter data in a manner that is undetectable. Viruses can be written to seek out programs of a particular file type and alter the data, yet maintain the format of the data so that the programs using the data operate normally. For example, a virus could be written to locate all the spreadsheet files (which have a common extension file name such as ".wks") and randomly alter a few of the data cells on the spreadsheet (e.g., it might alter some of the summation formulas to add 10% to the total). The spreadsheet program would then operate normally but give bad results. In a research application, this could cause someone to falsely report a result as significant because practically no one calculates data by hand anymore. To review: a virus infects a computer in three phases; only the third phase is detectable without special virus protection programs. In the first phase, incubation, the program typically waits until the user no longer suspects that a new applications program contains the virus. In the infection phase, the virus spreads to other programs that can be transported to other computers in a way that is not likely to become detected. In the destroy phase, the virus produces some sort of havoc within the host system. How to Limit the Spread of a Virus There are three approaches that can be taken in order to limit the spread of a virus. They involve: limiting access to the computer, installing virus protection programs, or installing disk-watch programs. Each method has its costs and benefits. No method is certain. Since viruses come from other computers, the first approach is to limit the number offoreign executable programsthat can be run on one's computer. Computer viruses cannot be spread through the air like some human viruses. A computer virus is more analogous to the human disease AIDS, which can bespread only through intimate physical contact or the exchange of blood. A computer will not pick up a virus unless one executes a program infected by another computer that has the virus. If executable programs are not imported from other sources, a computer cannot become infected. Viruses are particularly common on free computer bulletin boards, Christmas card programs, games, and some antivirus programs. One should bevery cautious about importing entertainment programs. Any popular game program is likely to have been run on many machines, and hence it is a good target for viruses. One should also be wary about using pirated software. Such software is more likely to have a virus because of its questionable history. The original manufacturers of software products will generally compile their programs from raw source code, which makes it very unlikely that a virus will be introduced by a manufacturer (although, there have been several instances in which commercial software houses have unintentionally spread viruses). The first line of defense I use in my lab is to have all staff members sign an agreement that they will not bring any executable software into the laboratory without written permission from the lab director. I also have my staff read sections of this paper for background. Students can still do word processing of their homework on the machines, but they must use the word-processing programs that are already in the laboratory. I find the risk of having a virus incubate for several months and then begin to randomly alter data to be too great to offset the minor benefit of allowing people to use external programs in my laboratory. The second method of limiting access is to use writeprotection tabs on disks whenever possible. Writeprotection tabs are usually small pieces of foil for 5.25-in. floppy disks, or a small tab on 3.5-in. diskettes, which can be flipped. The hardware in the disk drive checks to be sure that the disk is not write-protected before writing on it. By write-protecting a disk, one prevents a virus from altering the disk and also has the possibility of detecting a virus during its infection phase (e.g., if one gets a writeprotection error on the disk that one was not trying to write to, the probable cause is a virus). When one takes a floppy disk to a public site, it is particularly important to install write protection. Remember, even if one does not intend to write on a disk, a virus may copy itself onto the disk. Write-protection tabs can prevent this problem. The third method of limiting the access of viruses involves bringing copies of one's own operating system and any executable programs to external computers. In this way, one will be protected from being attacked by a virus at a public site. Note that it is critical that one boot off of one's own floppy disks rather than the public site's operating system, because the most likely source of the virus is the operating system itself. If it is necessary to execute programs at a public site, one should remove all executable files from any floppy disks to be operated there, and use write-protection tabs whenever possible. The second approach to virus protection, the installation of virus protection programs, has both benefits and costs. The Appendix provides a list of commercially available antivirus programs. Of course, any virus protection program is limited. Such programs can only detect viruses during their infection or destroy phase, and they can substantially disrupt normal computer functioning.' The typical antivirus program installs itself on the operating system and monitors the likely locations where a virus would attack. For example, an antivirus program might monitor to determine if any program tries to do a direct write to the disk, or install itself on the clock interrupt (common behaviors of viruses). The problem is that many programs will appropriately do direct writes to the disk (it COMPUTER VIRUSES 337 is often much more efficient) or alter the operating system. Most terminate and stay resident (TSR) programs, such as Borland's Sidekick or various spell checkers function thus (see Duncan, 1988). Some virus programs will do a check sum on every sector ofthe disk when it is read into the system. This can quickly detect when a virus goes into its infection phase, but, the check sum may consume 20-30% of one's computer time during disk input/output and substantially decrease the disk storage space. A serious problem with antivirus programs is that they frequently "false alarm." I found this to be an acute problem with the public domain antivirus program called Flushot. The program detected and provided a warning every time a program did a direct write to the disk. However, the word-processing program that I use on my computer also does direct disk writing. Thus, every time I used the word processor I got many false alarms, warning me about potential virus writes. Just as the lamb that always cries wolf will be ignored, an antivirus program that reports many legitimate operations as potential viruses will come to be ignored too. I decided that the standard antivirus programs false alarm too frequently to be usable in my laboratory. The third approach to limiting access of viruses is to install disk-watch programs to detect the results ofthe infection phase of a virus. A disk-watch program will maintain a check sum (i.e., a code for the specific bit pattern for all the bytes on the file) for every appropriate file on the system disk. The disk-watch program is run originally to make a fingerprint of all the executable files (typically the .EXE, .COM, .SYS, .BAT, driver, and overlay files, the boot block, and the operating-system overlay files). Whenever the disk-watch program checks the system, it checks the date, length, and check sum of all the files to see whether or not they have been altered. In most applications environments, users modify data but rarely modify the executable forms programs (the notable exception occurs when programmers use compilers to make new programs). For example, the user might utilize a word processor to modify many word-processing files in the system. The disk-watch program will verify that the wordprocessing executable program (e.g., WP.EXE) is not altered, and it will report any such alteration. However, it does not report when any of the word-processing files are altered. Since a virus cannot be spread via the wordprocessing files, it is not necessary that these be checked for viral infection. For the typical user of a computer system, the .EXE and .COM files are not altered except when new software is installed. Thus, a disk-watch program will rarely false alarm, and usually one can recognize the legitimate occurrences of false alarms (e.g., when a new version of the software has recently been installed). The process of installing and using disk-watch programs is straightforward. On first installation, the disk-watch program scans all ofthe files creating a signature file with the date, length, and check sums of all the programs. Every time the computer is rebooted, all of the executable files are checked for their length and date of last modification. This is done by including the disk-watchprogram name in the "AUTOEXEC.BAT" file to automatically check the disks. It takes about 12 sec to do a date, time, and length check of all the critical files on a reboot. Every week I do an automatic check-sum test. This requires 5 min (on an AT computer with a 200 MB disk and 200 files to check). The program that I use allows one to set a check-sum verification interval, so it automatically does the check-sum test once a week and the fast length and date checks otherwise.! This eliminates the need to keep records on when to do the check-sum test, and it does not slow down the boot process substantially. If the system detects that a file has been altered, it reports the file name and the way in which the file has been altered (changed in length, date, check sum). The user must then respond by indicating that he or she has noticed this change and can tell the program to alter the signature file so that the new version of the file will be recognized in the future. The user may have to enter a password to change the signature file. Updating the signature file takes only a few seconds for each altered file. A disk-watch program can be used on a "quarantine" computer to detect new viruses. One computer in my laboratory is designated as the quarantine computer, on which any new software is first run. On this computer, the disk-watch program checks not only the executable files but all files on the disk. In this way, any attack of the virus will be detected. I do check sums on the quarantine computer at both the beginning and the end of each day's running of the programs. To be effective, antivirus procedures must be as automatic and as easy to execute as locking the door of a house. An antivirus program should be completely selfexplanatory. It should be possible to create and update any signature files by means ofthe user's simply indicating "yes" or "no" to questions that require judgment (e.g., when a changed version of a program appears). If the program false alarms, or if it requires one to consult a manual, the antivirus procedures will probably not be used reliably. Whether one limits access to a computer, installs virus protection programs, or installs disk-watch programs, users incur at least some type of inconvenience when they attempt to limit the spread of computer viruses. Any system can be broken into. In my own laboratory, I have found limited access in combination with a disk-watch program to be the best compromise. This combined approach has prevented further virus infection, and it required only minimal changes in the normal routine of the laboratory.