An Analysis of How Antivirus Methodologies Are Utilized in Protecting Computers from Malicious Code

Daniel J. Sanok Jr
Kennesaw State University
1000 Chastain Road
Kennesaw, GA 30144, USA
404-514-9052
danny.sanok@gmail.com

ABSTRACT
Antivirus software utilizes several methodologies in scanning, detecting, and protecting computers and systems from viruses. As understanding increases about the vectors malicious code uses to attack and how antivirus software protects computer systems from the viruses, people will be able to more effectively help in creating an environment that is secure and virus free. This paper examines the techniques of signature detection, heuristics, and general decryption that antivirus software uses to detect and clean viruses.

Categories and Subject Descriptors
K.6.5 [**Security and Protection**]: *Authentication, Insurance, Invasive software (e.g., viruses, worms, Trojan horses), security, unauthorized access.*

General Terms
Security, Management.

Keywords
Antivirus, Scanning, Security, Encryption.

1. INTRODUCTION
Computer viruses are a two-edged sword in teaching society something about computers and information systems. One edge of the virus sword teaches how interconnected and sophisticated human beings have become as a result of computers. When a worldwide virus quickly strikes, it can have paralyzing implications on business and life until normalcy is returned. On the other edge of the virus sword, it shows how vulnerable society can be to attacks due to the lack of security from top to bottom in computers, operating systems, and other networked information system components. Viruses are much like their biological counterparts in many ways. Viruses are parasitic programs that attach themselves to a host, which is usually a computing device operating system like Windows for computers or Symbian for cell phones, and spread by transferring from host to host. Biological viruses operate in much the same way to biological hosts as electronic viruses do to their computer hosts.

Information Security Curriculum Development (InfoSecCD) Conference '05, September 23-24, 2005, Kennesaw, GA, USA.

2. VIRUS TYPES
Personal computer based viruses first appeared in the 1980s with the first one targeting the MS-DOS operating system in 1986; the first Internet based virus appeared in 1988, known as the Morris Worm. As of 2002, there were over 60,000 types of known viruses [3]. Viruses have different forms and are categorized by what their primary function is by their propagation method. Several types of viruses are listed below [2]:

•Boot Virus: a virus that infects the boot sector of disk storage.
•Parasitic Virus: a virus that attaches itself to executable files as part of their code. It runs whenever the host program runs.
•Polymorphic Virus: a virus that mutates with every new host to prevent signature detection.
•Macro Virus: a virus that exploits the macro language of programs like Microsoft Word or Microsoft Excel for malicious purposes.
•Trojan horse: a virus that enters a system disguised as something else. Their payloads include viruses, remote access methods, and data destruction.
•Worm, a virus that propagates on its own by a variety of means including hijacking email accounts, user IDs, and file transfer programs.
•Bomb: a virus that doesn't propagate itself at all; it is placed by a human or another program and activated by a trigger such as time or event.

In order to protect against these attacks, antivirus software companies employ several methodologies in defending against these attacks. They make use of file signature detection through scanning methods, which include email attachment scanning, download scanning, and static file scanning. They also utilize heuristic scanning and General Decryption (GD) scanning in fighting some of today's new complex and sophisticated viruses.

3. ANTIVIRUS METHODOLOGIES
With the threats and vulnerabilities to computers from viruses, there needs to be a way to fight the viruses. Just like their biological counterparts, computer programs known as antivirus software operate somewhat like antibiotics and even go a step further by attempting to be proactive in protecting against infection from computer viruses and malware programs that have

been created with the intent on doing something adverse to computers, data, and information systems.

The early viruses were simple machine language programs that attached themselves to programs and spread by creating identical copies from computer to computer. This made early virus detection methods simple. Antivirus applications were able to scan entire files looking for what is called a virus signature, which is a sequence of language instructions that is unique to the virus [1]. However, as virus authors wrote increasingly complex viruses to take advantage of new programming technologies, exploit new vulnerabilities, and spread in more devious ways to avoid detection by antivirus software. Antivirus companies had to evolve with them and develop new methods to scan, detect, and protect computers, data, and systems from viruses and the nuisance, havoc, and destruction they create.

Antivirus scanning methodologies main categories include signature detection through file scanning, heuristics, and general decryption. From those three classifications the various virus threats fall into one of the scanning methodologies based on which vector the virus uses.

3.1 Signature Detection

The first viruses that infected personal computers were parasitic in nature. They often attached themselves to other files and were executed or propagated whenever the host file was in use. In the early days of virus writing, many PCs utilized DOS as the base operating system and the viruses attacked file types called COM files, which are command executable files. The viruses usually attacked the executable's entry point, which is the beginning of the file [5]. Antivirus applications discovered this pattern of behavior and for speed and performance, simply scanned the PC for COM files and looked for a signature of the file. The signature would be a specific string of code that could be recognized as a virus. It compared the signatures it saw in the files against its database of virus signatures to verify the virus and take the appropriate actions for detection and cleaning.

When Windows 95 was introduced, most executable files migrated over to the EXE file type for executable files. As viruses migrated to the new file structures, antivirus software evolved as well. As machine resources such as memory, processor speed, system bus, and disk I/O performance increased, coupled with advances with antivirus technologies, file scanning was able to be implemented in two methods in the antivirus applications. It was performed on-demand when the user or system administrator wished to inspect the file system and files for viruses or on-the-fly, which is a background process in which as files are accessed, they are scanned for the presence of signatures that may positively match those in the antivirus application's database [5]. On-the-fly scanning is a feature that antivirus applications use to scan files as they are downloaded from the Internet, as email messages enter into an inbox of an email client, or when an application is launched. The method of performing this on the fly scanning is through hooks that antivirus applications can take advantage of in operating systems. In Microsoft Windows, these are known as Application Programming Interfaces (API). APIs are a set of software functions used by an application program as a means for providing access to a system's capabilities such as operations or communications. Since email is a predominant application that brings viruses into an organization, Microsoft has created an antivirus API that allows antivirus applications to utilize these API hooks in their Exchange Server software [8]. These AVAPIs were an answer to assist in signature detection by on-the-fly file scanning of email attachments in the email server database or in the message transfer agent (MTA) and help make email more protected. Most of the popular commercial antivirus applications allow users to configure their antivirus software to run scheduled on-demand scans of their fixed disks or run them on command by executing the on-demand scan. The scans can be configured to be intrusive and scan all files on a disk, compressed files, and other specified configurable options. These scans are resource intensive and allow the antivirus application to compare all the files on a system's signatures to those in the antivirus application database to ensure a system is clean of known static virus files.

Macro scanning of popular Microsoft Office applications like Word and Excel are easily performed also. Macros are powerful tools within the Office applications that expose APIs that viruses like to exploit [3]. Since the macros are stored in specific areas that antivirus file scanners can directly evaluate, the scanners can thoroughly interrogate the file for signature patterns and perform the appropriate actions for detection and cleaning if discovered.

3.2 Heuristics

The word heuristics comes from the Greek word "heuriskein", which means to discover (Munro, 2002). Heuristics describe a methodology of virus scanning by evaluating patterns of behavior and discovering when applications are performing out of the norms for operating. Heuristics antivirus methodologies perform in a manner that utilizes a search and discover operation to find viruses that do not have a known signature. Since heuristics look at file behavior, they are useful when combined with other antivirus technologies in determining when a program or file is a virus, such as with metamorphic viruses, which have a mutation engine that causes the properties of a file to change every time it is propagated [7]. This allows the antivirus application to look at the behavior of the program or file and determine from a probability the likeliness that it is a virus. Heuristics tends not to be as exact as signature matching, where an absolute match can be determined. In heuristics, the probability of the likelihood of a program or file being a virus by examining characteristics and behaviors matched to an antivirus heuristics database containing hundreds of indicators. These decision support systems aides a heuristics scanner in performing its job in monitoring suspicious code.

The decision support systems work by giving the heuristics scanner a weighting system. Points are determined for things like file date, file size, attempts to access system resources, or change other file properties. When a heuristics scanner is monitoring a file or program and scoring it, if the suspect being watched earns a certain number of points that crosses a threshold, the determination has to be made if the suspect is a virus. Heuristics scanners can have their thresholds configured different level of strictness. The strictness will help determine how many viruses are caught, but since heuristics works on a probability basis, chances are there will be a lot of false positives as well. One of the risks an organization would have to assume then is that if there are too many false positives, people would start to tend to ignore the false alarms thinking they are false in nature.

Heuristics methodologies typically are invoked when scanning methods do not detect abnormal behavior. One of the benefits of heuristics scanning is that it can work together with other scanning methodologies to aide in providing an overall protection package against virus and malicious code attacks. Heuristics is a useful when a signature can not be determined or when the file is suspicious but can not be determined and the general decryption engine is teamed up with the heuristics scanner to determine its intentions [4].

3.3 General Decryption

As virus writers worked to create increasingly complex viruses that avoided detection and were able to carry out havoc and data destruction, they created the encrypted virus. The viruses that generally fall into this category are polymorphic and metamorphic in nature. Polymorphic viruses were the first to appear in the encrypted virus category. An encrypted virus contains two parts, a small amount of code that produces the decryption routine and the encrypted virus body with the payload [7]. The virus and payload are simple viruses in their unencrypted state. The virus decryption technique also contains a small amount of code that can perform encryption once the virus has propagated itself. In a polymorphic virus, it has a routine to generate a new encryption key every time it re-encrypts its payload called a mutation engine, making the virus appear different each time.

General decryption scanners work to detect these threats is that this technology relies on virtual technology. It is comprised of three components: a CPU emulator, a virus signature scanner, and an emulation control module. When a request comes into the antivirus application to scan a questionable file, the general decryption engine loads the questionable file into its virtual environment. The virtual environment replicates a real computing environment. This allows the questionable file to execute and be observed under a microscope of the general decryption engine [7]. Risk is minimized, since the virtual system is a self contained operating environment that sandboxes the questionable file. In this environment the virus can be allowed to run its decryption mechanism. The general decryption engine monitors the behavior. Once the decryption occurs and the payload is unencrypted and begins to execute, the virus scanner can check the virus against its database of virus signatures to verify the virus and take the appropriate actions for detection and cleaning. The simplicity of combating new and sophisticated polymorphic and metamorphic viruses is that in this sandboxed virtual environment is that the virus does all the work while the general decryption engine observes the behavior and when the virus doesn't expect it, the antivirus software provides the proper protection by eliminating the threat before it can perform real damage to a system or data.

Metamorphic viruses are different from polymorphic viruses in that they do not contain a decryptor but are able to create a new generation that looks different every time they propagate. Much like polymorphic viruses, metamorphic viruses are detected using emulator based methodologies. But instead of performing file scanning at point of decryption, the metamorphic scanners observe a questionable file using heuristics methods. When the emulator based heuristics encounters a metamorphic virus, the heuristics technology is employed by observing and scoring the suspected file or program's behavior. Based on configured threshold settings, once the score is determined the decision is made whether the suspect is a virus or not and the program or file is allowed to execute in the real environment or have the appropriate actions for detection and cleaning take place.

4. CONCLUSIONS

Computer viruses have grown exponentially in numbers in recent years, from first appearing in the early 1980s, to a few hundred in the 1990s, to the tens of thousands by the end of the 1990s and early 2000 [3][6]. The complexity and sophistication of these viruses has increased over the years from simple parasitic pieces of code to morphing programs with the intention of avoidance and destruction. Antivirus scanning methods fall into three main categories that are signature detection through file scanning, heuristics, and general decryption. Most antivirus applications use a hybrid approach in detecting and protecting from viruses. As a result of the constant research and development of these methodologies, antivirus software has developed and evolved along with the viruses in an attempt to protect computers, data, and other information system components. Antivirus is an integral part of the overall defense-in-depth approach that organizations should employ in designing a security architecture with security controls at various levels of the IT infrastructure. Understanding the threats and how to counter them will better help protect the confidentiality, integrity, and availability of their systems and data.

5. REFERENCES

[1] Brain, M. (2004). How Computer Viruses Work. http://computer.howstuffworks.com/virus.htm

[2] Microsoft Corporation. (2004). The Antivirus Defense-in-Depth Guide. http://www.microsoft.com/technet/security/topics/serversecurity/avdind_0.mspx

[3] Munro, J. (2002, July 1). Antivirus Research and Detection Techniques. http://www.extremetech.com/article2/0%2C1558%2C325439%2C00.asp

[4] Muttik, I. (September 2000). Stripping Down an AV Engine. http://www.nai.com/common/media/vil/pdf/imuttik_VB_%20conf_2000.pdf

[5] Nachenberg, C. (1997). Computer virus-antivirus coevolution. Communications of the ACM, 40(1), 46-51.

[6] Schneier, B. (1999). The Trojan Horse Race. Communications of the ACM, 42(9), 128.

[7] Szor, P., & Ferrie, P. (September 2001). Hunting for Metamorphic. http://enterprisesecurity.symantec.com/PDF/metamorphic.pdf

[8] Trendmicro Corporation. (May 14, 2004). ScanMail for Microsoft Exchange 3.53. http://www.trendmicro.com/ftp/products/npf/sp/en/smex/readme-smex353-patch1.txt