# Virus and Content Filtering 15

## 15.1 Definitions

As the size of global computer networks expands and the use of the Internet skyrockets, the security issues do manifest themselves not only in the security of computer networks but also in individual user security on individual PCs connected to the Internet either via an organization's gateway or an Internet service provider (ISP). The security of every user, therefore, is paramount whether the user is a member of an organization network or a user of a home PC via an independent ISP. In either case, the effort is focused on protecting not only the data but also the user.

The most effective way to protect such a user and the data is through content filtering. Content filtering is a process of removing unwanted, objectionable, and harmful content before it enters the user network or the user PC. The filtering process can be located in several locations including on a user's PC, on a server within an organization, as a service provided by an ISP, or by means of a third-party site that provides the basis of a closed community.

In their report to the Australian Government on Content Filtering, Paul Greenfield et al. [1] divide the process of content filtering into two approaches: inclusion filtering and exclusion filtering.

## 15.2 Scanning, Filtering, and Blocking

Scanning is a systematic process of sweeping through a collection of data looking for a specific pattern. In a network environment, the scanning process may involve a program that sweeps through thousands of IP addresses looking for a particular IP address string or a string that represents a vulnerability or a string that represents a vulnerable port number. Filtering, on the other hand, is a process of using a computer program to stop an Internet browser on a computer from being able to load certain Web pages based upon predetermined criteria such as IP addresses. Blocking, like filtering, is also a process of preventing certain types of information



J.M. Kizza, *Guide to Computer Network Security*, Computer Communications and Networks, DOI 10.1007/978-3-319-55606-2_15

from being viewed on a computer's screen or stored on a computer's disk. In this section, we are going to look at these three processes and see how they are used in computer networks and personal computers as a way to enhance security.

### 15.2.1 Content Scanning

All Internet content inbound into and outbound from either an organization's network, an ISP gateway, or a user PC is always scanned before it is filtered. So scanning is very important in content filtering. Let us look at the ways scanning is done on the content of the Internet, either inbound or outbound. There are two forms of scanning: pattern-based and heuristic scanning.

#### 15.2.1.1 Pattern-Based Scanning

In pattern-based scanning, all content coming into or leaving the network, an ISP gateway, or user PC is scanned and checked against a list of patterns, or definitions, supplied and kept up to date by the vendor. The technique involves simply comparing the contents, which can be done in several ways as we saw in Sect. 12.2.1. Nearly all antivirus software packages work this way. This approach can, however, be slow and resource intensive.

#### 15.2.1.2 Heuristic Scanning

Heuristic scanning is done by looking at a section of code and determining what it is doing and then deciding whether the behavior exhibited by the code is unwanted, harmful like a virus, or otherwise malicious. This approach to scanning is complex because it involves modeling the behavior of code and comparing that abstract model to a rule set. The rule set is kept in a rule database on the machine and the database is updated by the vendor. Because of the checking and cross-checking, this approach takes more time, and it is also resource intensive, if not more than the previous one. Theoretically heuristics has several advantages over pattern-based scanning including better efficiency and accuracy. It can, potentially, detect viruses that haven't been written yet.

### 15.2.2 Inclusion Filtering

Inclusion filtering is based on the existence of an inclusion list. The inclusion list is a permitted access list—a "white list" probably vetted and compiled by a third party. Anything on this list is allowable. The list could be a list of URL for allowable Web sites, for example, it could be a list of allowable words, or it could be a list of allowable packet signatures for allowable packets. The nature of the list is determined by the security policy of the organization or a committee of a community. As Greenfield noted, this type of filtering can be 100% effective—assuming the person or organization that has compiled the white list shares the same set of values as the Internet user.

But the inclusion list approach, despite its effectiveness, has several drawbacks including the following:

- The difficulty to come up with a globally accepted set of criteria. This is a direct result of the nature of the Internet as a mosaic of a multitude of differing cultures, religions, and political affiliations. In this case, it is almost impossible to come up with a truly accepted global set of moral guidelines.
- The size of the inclusion list. As more and more acceptable items become available and qualify to be added on the list, there is a potential for the list to grow out of control.
- Difficulty of finding a central authority to manage the list. In fact, this is one of the most difficult aspects of the inclusion list approach to content filtering. For example, even though we have been suffering from virus attacks for years, there is no one authoritative list managed by a central authority that contains all the virus signatures that have ever been produced. There are currently highly inclusive lists managed by either private antivirus companies or publicly supported reporting agencies such as the Computer Emergency Reporting Team (CERT) Center.

### 15.2.3 Exclusion Filtering

Another approach to content filtering is the use of an exclusion list. This is the opposite of the inclusion list process we have discussed previously. An exclusion list is actually a "black list" of all unwanted, objectionable, and harmful content. The list may contain URLs of sites, words, signatures of packets, and patterns of words and phrases. This is a more common form of filtering than inclusion filtering because it deals with manageable lists. Also it does not pre-assume that everything is bad until proven otherwise.

However, it suffers from a list that may lack constant updates and a list that is not comprehensive enough. In fact, we see these weaknesses in the virus area. No one will ever have a fully exhaustive list of all known virus signatures, and antivirus companies are constantly ever updating their master lists of virus signatures.

### 15.2.4 Other Types of Content Filtering

In the previous two sections, we have discussed the two approaches to content filtering. In each one of these approaches, a list is produced. The list could be made up of URLs, words (keyword), phrases, packet signatures, profile, image analysis, and several other things. Let us now look at the details of content filtering based on these items [1].

#### 15.2.4.1 URL Filtering

With this approach, content into or out of a network is filtered based on the URL. It is the most popular form of content filtering, especially in terms of denial of access to the targeted site. One of the advantages of URL filtering is its ability to discriminate and carefully choose a site but leave the IP address of the machine that hosts functioning and therefore providing other services to the network or PC.

Because of the low level of and fine-tuning involved in URL filtering, many details of the setup and format of the target are needed in order to be able to provide the required degree of effectiveness. In addition, because of the low-level details needed, when there are changes in the files in the URL, these changes must be correspondingly affected in the filter.

#### 15.2.4.2 Keyword Filtering

Keyword filtering requires that all the inbound or outbound contents be scanned, and every syntactically correct word scanned is compared with words either on the inclusive white list or exclusive black list depending on the filtering regime used. Although it is the oldest and probably still popular, it suffers from several drawbacks including the following:

It is text based which means that it fails to check all other forms of data like images:

- It is syntactically based, meaning that it will block words with prefixes or suffixes that syntactically look like the forbidden words, ignoring the semantics of the surrounding text.

#### 15.2.4.3 Packet Filtering

As we discussed in Chap. 1, network traffic moves between network nodes based on a packet, as an addressable unit, with two IP addresses: the source address and the destination address. Throughout this book we have discussed the different ways these addresses are used in transporting data. As we saw in Chap. 11, content is blocked based on these IP addresses. Because of this approach, if content is blocked or denied access based on IP addresses, this means that no content can come from or go to the machine whose address is in the block rules. This kind of blocking is indiscriminate because it blocks a machine based on its addresses, not content, which means that a machine may have other good services but they are all blocked. As we discussed in Sect. 11.2, packet filtering can also be done based on other contents of a packet such as port numbers and sequence numbers.

#### 15.2.4.4 Profile Filtering

The use of artificial intelligence in content filtering is resulting into a new brand of content filters based on the characteristics of the text "seen" so far and the learning cycles "repeats" done to discriminate all further text from this source. However, because of the complexity of the process and the time involved and needed for the filters to "learn," this method, so far, has not gained popularity. In the preprocessing

phase, it needs to fetch some parts of the document and scan it—either text based or content based—in order to "learn." This may take time.

#### 15.2.4.5 Image Analysis Filtering

Ever since the debut of the World Wide Web with its multimedia content, Internet traffic in other formats different from text has been increasing. Audio and video contents are increasing daily. To accommodate these other formats and be able to filter based on them, new approaches had to be found. Among these approaches is the one based on analyzed images. Although new, this approach is already facing problems of preloading images for analysis, high bandwidth making it extremely slow, and syntactic filtering making it indiscriminate semantically.

### 15.2.5 Location of Content Filters

At the beginning of the chapter, we stated that there are four best locations to install content filters. These four locations include, first and foremost, the user's PC, at the ISP as the main gateway to and from the Internet to the user PC, at the organization server, and finally by the third-party machine. Let us briefly look at each one of these locations.

#### 15.2.5.1 Filtering on the End User's Computer

At this location, the user is the master of his or her destiny. Using software installed on the user machine, the user can set blocking rules and blocking lists that are expressive of his or her likes and dislikes. Because this location makes the user the focus of the filtering, the user is also responsible for updating the blocking rules and lists. In addition, the user is responsible for providing the security needed to safeguard the blocking rules and lists from unauthorized modifications.

#### 15.2.5.2 Filtering at the ISP's Computer

Unlike filtering at the user PC, filtering at the ISP removes the responsibility of managing the filtering rules from the user and lists and places it with the ISP. It also enhances the security of these items from unauthorized local changes. However, it removes a great deal of local control and the ability to affect minute details that express the user's needs.

Because this is a centralized filtering, it has several advantages over the others. First, it offers more security because the ISP can make more resources available than the user would. Second, the ISP can dedicate complete machines—called proxy servers—to do the filtering, thus freeing other machines and making the process faster. Finally, the ISP can have more detailed lists and databases of these lists than a user.

In Sect. 11.2.2, we discussed the use of proxy servers and filters as firewalls. So we have a basic understanding of the working of proxy servers. The proxy servers are installed in such a way that all traffic to and from the ISP must go through this

proxy server to be able to access the Internet. A proxy filter can be configured to block a selected service.

#### 15.2.5.3 Filtering by an Organization Server

To serve the interest of an organization, content filtering can also be done at a dedicated server at an organization. Just like at the ISP, the organization's system administrator can dedicate a server to filtering content into and out of the organization. All inbound and outbound traffic must go through the filters. Like ISP filtering, this is centralized filtering, and it offers a high degree of security because the filtering rules and lists are centrally controlled.

#### 15.2.5.4 Filtering by a Third Party

For organizations and individuals that are unable to do their own filtering, the third-party approach offers a secure good alternative. Both inbound and outbound traffic on the user and organization gateways are channeled through the third-party filters. The third party may use proxy servers like the ISPs or just dedicated servers like organization servers. Third-party filters offer a high degree of security and a variety of filtering options.

## 15.3 Virus Filtering

Our discussion of viruses started in Chap. 3, where we introduced viruses as a threat to system security. We discussed the big virus incidents that have hit the Internet causing huge losses. In Sect. 5.3.5, we looked at viruses as hackers' tools. Although we did not specifically define the virus, we discussed several types of viruses and worms that hackers use to attack systems. Now we are ready to define a computer virus on our way to filtering it.

### 15.3.1 Viruses

A computer virus is a self-propagating computer program designed to alter or destroy a computer system resource. The term *virus* is derived from a Latin word *virus* which means poison. For generations, even before the birth of modern medicine, the term had remained mostly in medical circles, meaning a foreign agent injecting itself in a living body, feeding on it to grow and multiply. As it reproduces itself in the new environment, it spreads throughout the victim's body slowly, disabling the body's natural resistance to foreign objects, weakening the body's ability to perform needed life functions, and eventually causing serious, sometimes fatal, effects to the body.

Computer viruses also parallel the natural viruses. However, instead of using the living body, they use software (executable code) to attach themselves, grow, reproduce, and spread in the new environment. Executing the surrogate program starts them off, and they spread in the new environment, attacking major system

resources that sometimes include the surrogate software itself, data, and sometimes hardware, weakening the capacity of these resources to perform the needed functions, and eventually bringing the system down.

The word virus was first assigned a nonbiological meaning in the 1972 science fiction stories about the G.O.D. machine that were compiled into a book *When Harlie Was One* by David Gerrold (Ballantine Books, First Edition, New York, NY, 1972). Later association of the term with a real-world computer program was by Fred Cohen and then a graduate student at the University of Southern California. Cohen wrote five programs, actually viruses, to run on a VAX 11/750 running Unix, not to alter or destroy any computer resources but for class demonstration. During the demonstration, each virus obtained full control of the system within an hour [2]. From that simple and harmless beginning, computer viruses have been on the rise. Computer viruses are so far the most prevalent, most devastating, and the most widely used form of computer system attack. And of all types of systems attacks, it is the fastest growing. As we reported in Chap. 2, Symantec reports that on the average there are between 400 and 500 new viruses per month [3]. The virus is, so far, the most popular form of computer system attack because of the following factors:

- Ease of generation. Considering all other types of system attacks, viruses are the easiest to generate because the majority of them are generated from computer code. The writing of computer code has been becoming easier every passing day because, first, programming languages are becoming easier to learn and develop programs; second, there are more readily available virus code floating around on the Internet; and finally, there is plenty of help for would-be virus developers in terms of material and physical support. Material support in form of how-to manuals and turn-key virus programs is readily available free on the Internet.
- Scope of reach. Because of the high degree of interconnection of global computers, the speed at which viruses are spread is getting faster and faster. The speed at which the "Code Red" virus spread from the Philippines through Asia to Europe to North America attests to this. Within a few days of release, Code Red had the global networks under its grip.
- Self-propagating nature of viruses. The new viruses now are far more dangerous than their counterparts several years ago. New viruses self-propagate, which gives them the ability to move fast and create more havoc faster. One of the reasons that the Code Red virus was able to move so fast was that it was self-propagating.
- Mutating viruses. The new viruses are not only self-propagating, which gives them speed, but they are also mutating which gives them a double punch of delaying quick eradication and consuming great resources and therefore destroying more in their wake, fulfilling the intended goals of the developers.
- Difficult to apprehend the developer. As the Code Red virus demonstrated, owing to legal and other limitations, it is getting more and more difficult to apprehend the culprits. This in itself is giving encouragement to would-be virus developers that they can really get way with impunity.

#### 15.3.1.1 Virus Infection/Penetration

There are three ways viruses infect computer systems and are transmitted: boot sector, macro penetration, and parasites [4].

**Boot Sector Penetration** Although not very common nowadays, boot sectors are still being used somehow to incubate viruses. A boot sector is usually the first sector on every disk. In a boot disk, the sector contains a chunk of code that powers up a computer. In a nonbootable disk, the sector contains a file allocation table (FAT), which is automatically loaded first into the computer memory to create a roadmap of the type and contents of the disk for the computer to access the disk. Viruses imbedded in this sector are assured of automatic loading into the computer memory.

**Macros Penetration** Since macros are small language programs that can execute only after imbedding themselves into surrogate programs, their penetration is quite effective. The rising popularity in the use of script in Web programming is resulting in macro virus penetration as one of the fastest forms of virus transmission.

**Parasites** These are viruses that do not necessarily hide in the boot sector, nor use an incubator like the macros, but attach themselves to a healthy executable program and wait for any event where such a program is executed. These days, due to the spread of the Internet, this method of penetration is the most widely used and the most effective. Examples of parasite virus include Friday the 13th, Michelangelo, SoBig, and the Blaster viruses.

Once a computer attack is launched, most often a virus attack, the attacking agent scans the victim system looking for a healthy body for a surrogate. If it is found, the attacking agent tests to see if it has already been infected. Viruses do not like to infect themselves, hence wasting their energy. If an uninfected body is found, then the virus attaches itself to it to grow, multiply, and wait for a trigger event to start its mission. The mission itself has three components:

- To look further for more healthy environments for faster growth, thus spreading more
- To attach itself to any newly found body
- Once embedded, either to stay in the active mode ready to go at any trigger event or to lie dormant until a specific event occurs

#### 15.3.1.2 Sources of Virus Infections

Computer viruses, just like biological viruses, have many infection sources. Again like biological viruses, these sources are infected first from first contact with either a newly released virus or a repeat virus. One interesting fact about computer virus attacks, again following their cousins the biological viruses, is that a majority of them are repeat attacks. So like in human medicine, a certain type of proven medications is routinely used to fight them off. Similarly with computer viruses,

the same antivirus software is routinely used to fight many of the repeat viruses. Of late, however, even known viruses have been mutating, making antivirus companies work harder to find the code necessary to eliminate the mutating virus.

Of the known viruses, there are mainly four infection sources: movable computer disks such as floppies, zips, and tapes; Internet downloadable software such as beta software, shareware, and freeware; e-mail and e-mail attachments; and platform-free executable applets and scripts. It is important to note that just like biological viruses, infections are caused by coming in close contact with an infected body. Likewise in computer viruses, viruses are caught from close contact with infected bodies—system resources. So the most frequently infected bodies that can be sources of viruses are as follows [4]:

- Movable computer disks: Although movable computer disks like floppies, zips, and tapes used to be the most common way of sourcing and transmitting viruses, new Internet technologies have caused this to decline. Viruses sourced from movable computer disks are either boot viruses or disk viruses.
- Boot viruses: These viruses attack boot sectors on both hard and floppy disks. Disk sectors are small areas on a disk that the hardware reads in single chunks. For DOS formatted disks, sectors are commonly 512 bytes in length. Disk sectors, although invisible to normal programs, are vital for the correct operation of computer systems because they form chunks of data the computer uses. A boot sector is the first disk sector or first sector on disk or diskette that an operating system is aware of. It is called a boot sector because it contains an executable program the computer executes every time the computer is powered up. Because of its central role in the operations of computer systems, the boot sector is very vulnerable to virus attack, and viruses use it as a launching pad to attack other parts of the computer system. Viruses like this sector because from it, they can spread very fast from computer to computer, booting from that same disk. Boot viruses can also infect other disks left in the disk drive of an infected computer.
- Disk viruses: Whenever viruses do not use the boot sector, they embed themselves, as macros, in disk data or software. A macro is a small program embedded in another program and executes when that program, the surrogate program, executes. Macro viruses mostly infect data and document files, templates, spreadsheets, and database files.
- Internet downloadable software: Historically, it used to be that computer viruses were actually hand carried. People carried viruses on their floppy disks whenever they transferred these infected disks from one computer to the other. Those were the good old days before the Internet and the concept of downloads. The advent of the Internet created a new communication and virus transmission channel. In fact, the Internet is now the leading and fastest virus transmission channel there is. Internet downloads, bulletin boards, and shareware are the actual vehicles that carry the deadly virus across the seas in a blink of an eye.
- E-mail attachments: As recent mega virus attacks such as the “Code Red,” “SoBig,” and the “Blaster” have demonstrated, no computer connected to the

Internet is safe any longer. E-mail attachments are the fastest-growing virus transmission method today. With more than one half of all today's Internet traffic made up of e-mails and millions of e-mails being exchanged a day going through millions of other computers, the e-mail communication is the most potent channel of infecting computers with viruses. Incidentally straight-texted e-mails, these are e-mails without attachments, are free from viruses. Since attachment-free e-mails are pure texts, not executables, they cannot transport viruses. Viruses, as we have already seen, are executable programs or document macros that can be embedded into other executables or application documents.

- Platform-free executable applets and scripts: Dynamism has made Web application very popular these days. Web dynamism has been brought about by the birth of scripting languages such as Java, Pearl, and C/C++. As we discussed in Chap. 6, the Common Gateway Interface (CGI) scripts let developers create interactive Web scripts that process and respond to user inputs on both the client side and the server side. Both CGI scripts, which most often execute on the server side and JavaScript and VBScript that execute within the user's browser on the client side, create loopholes in both the server and the client to let in viruses. One way of doing this is through a hacker gaining access to a site and then changing or replacing the script file. The hacker can also lay a "man-in-the-middle" attack by breaking in a current session between the client browser and the server. By doing so, the hacker can then change the message the client is sending to the server script.

#### 15.3.1.3 Types of Viruses

Just like living viruses, there are several types of digital (computer) viruses, and there are new brands almost every other day. We will give two classifications of computer viruses based on transmission and outcomes [4, 5].

**Virus Classification Based on Transmission**

- *Trojan horse viruses*: These viruses are labeled Trojan horse viruses because just like in the old myth in which the Greeks, as enemies of Troy, used a large wooden horse to hide in and enter the city of Troy, these viruses use the tricks these legendary Greeks used. During transmission, they hide into trusted common programs such as compilers, editors, and other commonly used programs. Once they are safely into the target program, they become alive whenever the program executes.
- *Polymorphic viruses*: These viruses are literally those that change form. Before a polymorphic virus replicates itself, it must change itself into some other form in order to avoid detection. This means that if the virus detector had known the signature for it, this signature then changes. Modern virus generators have learned to hide the virus signatures from antivirus software by encrypting the virus signatures and then transforming them. These mutations are giving virus hunters a really hard time. The most notorious mutating virus was the "Code

Red" virus which mutated into almost a different form every other day, throwing virus hunters off track.

- *Stealth virus*: Just like the polymorphic virus that uses mutation to distract its hunters from its track, a stealth virus makes modifications to the target files and the system's boot record, and then it hides these modifications. It hides these modifications by interjecting itself between the application programs the operating system must report to and the operating system itself. In this position, it receives the operating system reports and falsifies them as they are being sent to the programs. In this case, therefore, the programs and the antivirus detector would not be able to detect its presence. Once it is ready to strike then it does so. Jasma [5] gives two types of stealth viruses: the size stealth which injects itself into a program and then falsifies its size and the read stealth which intercepts requests to read infected boot records or files and provides falsified readings, thus making its presence unknown.
- *Retrovirus*: A retrovirus is an antivirus fighter. It works by attacking antivirus software on the target machine so that it can either disable it or bypass it. In fact, that is why it is sometimes called an *anti-antivirus* program. Other retroviruses focus on disabling the database of integrity information in the integrity checking software, another member of the antivirus family.
- *Multipartite virus*: This is a multifaceted virus that is able to attack the target computer from several fronts. It is able to attack the boot record and all boot sectors of disks including floppies, and it is also able to attack executable files. Because of this, it was nicknamed *multipartite*.
- *Armored virus*: Probably the name is fitting because this virus works in the target computer by first protecting itself so that it is more difficult to detect, trace, disassemble, or understand its signature. It gets the coat or armor by using an outer layer of protective coat that cannot easily be penetrated by antivirus software. Other forms of this virus work by not using a protective coat but by hiding from an antivirus software.
- *Companion virus*: This is a smarter virus that works by creating companions with executables. Then it piggybacks on the executable file and produces its own extension based on the executable file. By so doing, every time the executable software is launched, it always executes first.
- *Phage virus*: This virus parallels and is named after its biological counterpart that replaces an infected cell with itself. The computer counterpart also replaces the executable code with its own code. Because of its ability to do this, and just like its biological cousin, it is very destructive and dangerous. It destroys every executable program it comes into contact with.

**Virus Classifications Based on Outcomes**

- *Error-generating virus*: Error-generating viruses lunch themselves most often in executable software. Once embedded, they attack the software to cause the software to generate errors.

- *Data and program destroyers*: These are viruses that attach themselves to a software and then use it as a conduit or surrogate for growth, replication, and launch pad for later attacks and destruction to this and other programs and data.
- *System crusher*: These, as their name suggests, are the most deadly viruses. Once introduced in a computer system, they completely disable the system.
- *Computer time theft virus*: These viruses are not harmful in any way to system software and data. Users use them to steal system time.
- *Hardware destroyers*: While most viruses are known to alter or destroy data and programs, there are a few that literally attack and destroy system hardware. These viruses are commonly known as *killer viruses*. Many of these viruses work by attaching themselves to micro-instructions, or "mic," such as bios and device drivers.
- *Logic/time bombs*: Logic bombs are viruses that penetrate the system, embedding themselves in the system's software, using it as a conduit, and waiting to attack once a trigger goes off.

#### 15.3.1.4 How Viruses Work

In Sects. 15.3.1.2 and 15.3.1.3, we discussed how computers get infected with viruses and how these viruses are transmitted. We pointed out that the viruses are usually contracted from an infected computer resource and then passed on. We discussed those most likely resources to be infected and from which viruses are passed on. We have also pointed out in other parts of this chapter that over time, the methods of virus transmission have actually multiplied. In the beginning, viruses used to be transmitted manually by users moving disks and other infected materials from one victim to another. Since the birth of the Internet, this method has, however, been relegated to the last position among the popular methods of virus transmission.

Let us look at how the Internet has transformed virus transmission by focusing on two types of viruses that form the biggest part of virus infection within the network environment. These are the macro virus and the file virus. Of the two, the macro viruses have the fastest-growing rate of infection in networks. This is a result of several factors including the following:

- Big software warehouses innocently intend to provide their users with the flexibility of expanding their off-the-shelf products capabilities and functionalities by including macro facilities in these products. For example, popular Microsoft products include these macros [5]. Using these macro facilities, able users can create their own macros to automate common tasks, for example. But as we saw in Sect. 15.3.1.1, these macros are becoming a vehicle for virus infection and transmission.
- Microprogramming languages are now built into popular applications. These microprogramming languages are getting more and more powerful and are now packing more features. They can be used to build macros to perform a variety of functions. For example, Microsoft *Visual Basic for Applications* (VBA) is such a language that is found in a number of Microsoft popular applications including

PowerPoint, Excel, and Word. Again as we pointed out in Sect. 15.3.1.1, this creates ready vehicles to carry viruses.

The problem with these macros is that they introduce loopholes in these popular Internet applications. For example, VBA can be used by hackers to define viral code within the applications. Other macros that are not built using programming and scripting languages are included in applications that can be used by hackers as easily. The fact that macros behave as executable code within the applications is very attractive to hackers to use it and introduce viral code into the computer and hence into the network.

Next to macros in applications software in network transmission capabilities are file viruses. File viruses may be any of the types we have already discussed that attack system or user files. File viruses present as much danger to a network as the macro viruses as long as the infected computer is attached to a network. Notice that we would have nothing to say if a computer is not attached to any network. In fact, the safest computers are disconnected computers in bankers.

#### 15.3.1.5 Antivirus Technologies

There are four types of viruses that antivirus technologies are targeting. These are "in-the-wild" viruses that are active viruses detected daily on users' computers all over the world, macro viruses, polymorphic viruses, and standard viruses.

The "in-the-wild" viruses are collected and published annually in the *WildList* (a list of those viruses currently spreading throughout a diverse user population). Although it should not be taken as the list of "most common viruses," in recent times, the list has been used as the basis for in-the-wild virus testing and certification of antivirus products by a number of antivirus software-producing companies. Additionally, a virus collection based upon the *WildList* is being used by many antivirus product testers as the definitive guide to the viruses found in the real world and thus to standardize the naming of common viruses. For the archives and current list of the *WildList*, see *The WildList – (c)1993–2003 by Joe Wells* at http://www.wildlist.org.

The other three types of viruses—the macro viruses, polymorphic viruses, and standard viruses—have already been discussed in various parts of this chapter. Antivirus technologies are tested for their ability to detect all types of viruses in all these modes.

## 15.4 Content Filtering

As we noted in Sect. 11.2.1, content filtering takes place at two levels: at the application level where the filtering is based on URL which may, for example, result in blocking a selected Web page or an FTP site and filtering at the network level based on packet filtering which may require routers to examine the IP address of every incoming or outgoing traffic packet. The packets are first captured, and

then their IP address both source and destination, port numbers, or sequence numbers are then compared with those on either the *black* or *white* list.

### 15.4.1 Application-Level Filtering

Recall in Sects. 11.2.1 and 15.2.4 that application-level filtering is based on several things that make up the blocking criteria, including URL, keyword, and pattern. Application filtering can also be located at a variety of areas including at the user's PC, at the network gateway, at a third party's server, and at an ISP. In each one of these locations, quite an effective filtering regime can be implemented successfully. We discussed that when applying application-level filtering either at the network or at the ISP, a dedicated proxy server may be used. The proxy then prevents inbound or outbound flow of content based on the filtering rules in the proxy. With each request from the user or client, the proxy server compares the clients' requests with a supplied "black list" of Web sites, FTP sites, or newsgroups. If the URL is on the black list, then effective or selective blocking is done by the proxy server. Besides blocking data flowing into or out of the network or user computer, the proxy also may store *(cache)* frequently accessed materials. However, the effectiveness of application-level blocking using proxy servers is limited as a result of the following technical and nontechnical factors [6].

#### 15.4.1.1 Technical Issues

- *Use of translation services in requests can result in requested content from unwanted servers and sites:* If a user requests for content from a specified server or site, and if the requested content cannot be found at this site, the translation service operated by the request can generate requests to secondary sites for the content. In such cases then, the content returned may not be from the specified server unless secondary requests are specifically blocked.
- *The domain name server can be bypassed:* Since a user's request for a site access can be processed based on either a domain name or the IP address of the server, a black list that contains the domain names only without their corresponding IP addresses can, therefore, be bypassed. This usually results in several difficulties, including not processing requests whose IP addresses cannot be found on the black lists and doubling of the size of the black list if both domain names and equivalent IP addresses are used for every server on the list.
- *The reliability of the proxy server may be a problem:* The use of a single proxy server for all incoming and outgoing filtering may cause "bottleneck" problems that include reduced speed, some applications failing to work with specific servers, and loss of service should the server were to collapse.

#### 15.4.1.2 Nontechnical Issues

- *ISPs problems:* ISPs involved into the filtering process may face several problems, including the added burden of financially setting up, maintaining, and administering the additional proxy servers, supporting and maintaining reluctant clients that are forced to use these servers, and meeting and playing a role of a moral arbiter for their clients, the role they may find difficult to please all their clients in. In addition to these problems, ISPs are also faced with the problems that include the creation or updating and hosting black lists that will satisfy all their clients or creating, updating, and distributing black lists in a secure manner to all their clients.
- *The costs of creating and maintaining a black list:* There is an associated high cost of creating and maintaining a black list. The associated costs are high because the black list creation, maintenance, and updates involve highly charged local politics and a high degree of understanding in order to meet the complex nature of the list that will meet the basic requirements that cover a mosaic of cultures, religions, and political views of the users. In addition to these costs, there are also the costs of security of the list. Black lists are high target objects and prime targets for hackers and intruders

### 15.4.2 Packet-Level Filtering and Blocking

In Chap. 2, we saw that every network packet has both source and destination IP addresses to enable the TCP protocol to transport the packet through the network successfully and to also report failures. In packet-level filtering and blocking, the filtering entity has a black list consisting of "forbidden" or "bad" IP addresses. The blocking and filtering processes then work by comparing all incoming and outgoing packet IP addresses against the IP addresses on the supplied black list. However, the effectiveness of packet-level blocking is limited by both technical and nontechnical problems [6].

#### 15.4.2.1 Technical Issues

- *Packet-level blocking is indiscriminate:* Blocking based on an IP address of a victim server means that no one from within the protected network will be able to reach the server. This means that any service offered by that server will never be used by the users in the protected network or on the protected user computer. If the intent was to block one Web site, this approach ends up placing the whole server out of reach of all users in the protected server or the user PC. One approach to lessen the blow of packet-level filtering to the protected network or user PC is the use of port numbers that can selectively block or unblock the services on the victim server. However, this process can affect the performance of the proxy server.

- *Routers can easily be circumvented:* Schemes such as *tunneling,* where an IP packet is contained inside another IP packet, are commonly used, particularly in the implementation of virtual private networks for distributed organizations and the expansion of IPv4 to IPv6: one can very easily circumvent the inside victim IP address by enveloping it into a new IP address which is then used in the transfer of the encased packet. Upon arrival at the destination, the encased packet is then extracted by the receiver to recreate the original message. We will discuss tunneling in Sect. 16.4.2, 17.4.2
- *Blacklisted IP addresses are constantly changing:* It is very easy to determine that a server has been blacklisted just by looking at and comparing server accesses. Once it is determined that a server has been blacklisted, a determined owner can very easily change the IP address of the server. This has been done many times over. Because of this and other IP address changes due to new servers coming online and older ones being decommissioned, there is a serious need for black list updates. The costs associated with these constant changes can be high.
- *Use of nonstandard port numbers:* Although it is not very common, there are many applications that do not use standard port numbers. Use of such nonstandard port numbers may fool the server filter, and the blocked port number may go through the filter. This, in addition to other filtering issues, when implementing a firewall may complicate the firewall as well.

#### 15.4.2.2 Nontechnical Issues

- *Increased operational costs and ISP administrative problems:* As we saw in the application-level blocking, there are significant cost increments associated with the creation, maintenance, and distribution of black lists. In addition, the ISPs are made to be moral arbiters and supervisors and must carefully navigate the cultural, religious, and political conflicts of their clients in order to maintain an acceptable blacklist.

### 15.4.3 Filtered Material

The list of filtered items varies from user to user, community to community, and organization to organization. It is almost impossible, due to conflicting religious, cultural, and political beliefs, to come up with a common morality upon which a list like a "black list" can be based. Lack of such a common basis has created a mosaic of spheres of interests based on religion, culture, and politics. This has caused groups in communities to come together and craft a list of objectionable materials that can be universally accepted. The list we give below is a collection of many objectionable materials that we have collected from a variety of sources. This list includes the following items [7, 6]:

- *Nudity* is defined differently in different cultures. However, in many cultures, it means the complete absence of clothing or exposure of certain living human body parts.
- *Mature content* is differently defined and lacks universal acceptance. However, in many cultures, it refers to material that has been publicly classified as bad and corrupting to minors. The material may be crude or vulgar language or gestures or actions.
- *Sex*: Verbal and graphic descriptions and depictions of all sexual acts and any erotic material as classified by a community based on their culture, religion, and politics.
- *Gambling*: There are many forms of gambling, again based on community standards. These forms include physical and online gambling and game batting.
- *Violence/profanity*: Physical display and depictions of all acts that cause or inflict physical and psychological human pain including murder, rape, and torture.
- *Gross depiction*: Any graphic images, descriptive or otherwise, that are crude, vulgar, and grossly deficient in civility and behavior.
- *Drug/drug culture and use*: Graphic images, descriptive or not, that advocate any form of illegal use of and encouraging usage of any recreational drugs, including tobacco and alcohol advertising.
- *Intolerance/discrimination*: Advocating prejudice and denigration of others' race, religion, gender, disability or handicap, and nationality.
- *Satanic or cult*: Satanic materials that include, among others, all graphic images descriptive or otherwise that contain sublime messages that may lead to devil worship, an affinity for evil, or wickedness.
- *Crime*: Encouragement of, use of tools for, or advice on carrying out universally criminal acts that include bomb making and hacking.
- *Tastelessness*: Excretory functions, tasteless humor, graphic images taken out of acceptable norms, and extreme forms of body modification, including cutting, branding, and genital piercing.
- *Terrorism/militant/extremists*: Graphic images in any form that advocate extremely aggressive and combatant behaviors or advocacy of lawlessness.

## 15.5 Spam

It may be difficult to define spam. Some people want to define it as unsolicited commercial e-mail. This may not fully define spam because there are times when we get wanted and indeed desired unsolicited e-mails and we feel happy to get them. Others define spam as automated commercial e-mail. But many e-mails that are unsolicited and sometimes automated are not commercial in nature. Take, for example, the many e-mails you get from actually worthy causes but unsolicited and sometimes annoying. So to cover all these bases and hit a balance, we define spam as *unsolicited automated e-mail*.

Because Internet use is more than 60% e-mail, spamming affects a large number of Internet users. There are several ways we can fight spam including the following:

- *Limit e-mail addresses posted in a public electronic place*. E-mail addresses usually posted at the bottom of personal Web pages are sure targets of spammers. Spammers have almost perfected a method of cruising the Internet hunting for and harvesting these addresses. If you must put personal e-mail on a personal Web page, find a way of disguising it. Also opt out of job, professional, and member directories that place member e-mail addresses online.
- *Refrain from filling out online forms that require e-mail addresses*. Always avoid, if you can, supplying e-mail addresses when filling any kind of forms, including online forms that ask for them. Supply e-mail addresses to forms only when replies are to be done online.
- *Use e-mail addresses that are* not *easy to guess*. Yes, passwords can be successfully guessed, and now spammers are also at it trying to guess e-mail addresses. The easiest way to do this is to start with sending mails to addresses with short stem personal fields on common ISPs such as AOL, Yahoo, and Hotmail, fields like tim@aol, tim26@aol, joe@hotmail, and so on.
- *Practice using multiple e-mail addresses*. Always use several e-mail addresses and use one address for strictly personal business. When filling forms for nonserious personal business and pleasure, use a different e-mail address. In fact, it is always easy to determine who sells your e-mail address this way. By noting which address was used on which form and to whom, one can also easily track what sites are causing spam. These days there are also one-time disposable e-mail addresses one can easily get and use with little effort.
- *Spam filtering*. Always using spam filters at either the network level or application level to block unwanted e-mails. In either case, the spam is prevented from reaching the user by the filter. We will discuss this more in Sect. 15.3. While this approach has its problems, as we will see, it can cut down tremendously the amount of spam a user receives. Many ISPs are now offering spam filters.
- *Spam laws*. The outcry caused by spamming has led many national and local governments to pass spam laws. In Europe, the European Union's digital privacy rules are passed and are in force; these rules require companies to get consent before sending e-mail, tracking personal data on the Web, or pinpointing callers' location via satellite-linked mobile phones. The same rules also limit companies' ability to use cookies and other approaches that gather user information [8]. In the United States, efforts are being made to enact spam laws both at federal and state levels.
  - Federal spam law: The Senate approved a do-not-spam list and ban on sending unsolicited commercial e-mail using a false return address or misleading subject line [8].
  - State spam laws. All states have some form of spam laws on the books.

The European Union, leading the pack of antispam legislators, has passed a digital privacy law that requires companies to seek users' consent before sending

e-mails, tracking personal data on the Web, and pointing callers' location using satellite-linked cell phones unless it is done by the police or emergency services [9]. Other European countries have enacted spam laws with varying success, and these laws can be viewed at http://www.spamlaws.com/eu.html.

In the United States, the recently passed *Controlling the Assault of Non-Solicited Pornography and Marketing Act of 2003*, or the *CAN-SPAM Act of 2003*, tries to regulate interstate commerce by imposing limitations and penalties on the transmission of unsolicited commercial electronic mail via the Internet. In addition to the federal law, many states have much stronger antispam legislations.

In general, however good and strong antispam legislations are, it is extremely difficult and expensive to enforce.

Beside the United States, the EU, and European countries, several other countries outside Europe, including Australia, Canada, Japan, Russia, Brazil, and India, have or are in the process of enacted spam laws. This is an indication that there is a global movement to fight spam.

**Exercises**

1. What are the major differences between a boot virus and a macro virus? Which is more dangerous to a computer system?
2. List and briefly discuss three most common sources of virus infections.
3. In this chapter, we did not discuss the likely sources of computer viruses. Discuss four most likely sources of computer viruses.
4. Why is antivirus software always developed after the virus has stricken?
5. Describe the similarities between biological viruses and computer viruses.
6. What are the difficulties faced by a community that wants to filter the Internet content?
7. Describe how a virus is moved on the Internet.
8. Why it is that viruses are more dangerous on peer-to-peer networks than in client-server networks?
9. Study and discuss the virus infection rate in peer-to-peer network, client-server networks, and the Internet.
10. Why do macros have the highest infection rate in network virus transmission?

**Advanced Exercises**

1. Research and develop a comprehensive list of the currently known viruses.
2. Research, find, and study a virus code. Write an antivirus for that code.
3. Look at a popular application such as PowerPoint or Excel. Find and disable the micros. How do you enable them again?
4. Discuss and develop a policy for dealing with viruses.
5. What is a virus "in the wild?" Research and draw an estimate of all viruses in the wild. How do you justify your number?

## References

1. Greenfield P, McCrea P, Ran S (1999) Access prevention techniques for internet content filtering. CSIRO, Commonwealth Scientific and Industrial Research Organization (Australia), Australia. National Office for the Information Economy
2. Forcht K (1994) Computer security management. Boyd & Fraser Publishing, Danvers
3. BBC News. Battling the Net Security Threat. http://news.bbc.co.uk/2/hi/technology/2386113.stm
4. Kizza JM (2002) Computer network security and cyber ethics. McFarland and Company, Jefferson
5. Jasma K (2002) Hacker proof: the ultimate guide to network security, 2nd edn. OnWord Press, Albany
6. CSIRO (1998) Blocking content on the internet: a technical perspective. CSIRO, Mathematical and Information Sciences
7. Kizza JM (1998) Civilizing the internet: global concerns and efforts towards regulation. Jefferson, McFarland & Company
8. The Associated Press. Anti-spam law goes into force in Europe. Chattanooga Times-Free Press. Saturday, November 1, 2003. C5
9. The Associated Press. Anti-spam law goes into force in Europe. Chattanooga Times Free Press. C5, Saturday, November 1, 2003