FEATURE SERIES

The Usual Suspects – Part 1

Andreas Marx
University of Magdeburg, Germany

Most of today's anti-virus software detects nearly every known virus, even very complex polymorphic ones. However, to be good in the 'virus scanning' category there is much more out there to detect than simply all in-the-wild or zoo viruses.

This feature concentrates on virus-related problems in AV scanners, and how developers can avoid them. It is based both on results from our various tests (see http://www.av-test.org) and on comments received from IT representatives in large, international corporations. Most of the points raised look very simple, but they are all too often overlooked. This first part starts with trivial issues while the second instalment will reflect on more complex problems.

File Extensions

Most scanners do not scan all files by default – they use an extension list. Since new viruses have started to target 'new infectable' extensions, a program has to update this list with every scanner update. A better idea would be to scan all extensions by default to avoid this problem. However, there is usually an associated performance dip and, sometimes, additional heuristic false positives will be triggered.

Scanning everything on-access will cause huge performance problems if the scanner is dumb enough to scan everything every time, even if the file is unchanged or cannot be infected at all. Other problems will be caused with temporary files and very large files – it is not a solution, but it helps if the maximum size of the file to scan can be configured. However, at email gateway level 'scan all files' should be the default setting, since files can be renamed too easily to avoid detection.

Some scanners do not actually scan all files even when set to 'scan all files' or when the mask '*.*' is used. Most of the time at least some infected .BAT, .VBS and .COM files will be missed if they have non-standard extensions. This happens when the scanner checks the file extension, not the content, in order to scan solely for this kind of virus. It would be a good idea for vendors to make a 'smart' scan to find out the (hopefully) correct file format. If there is more than one possibility (like ASCII text or a .COM file), all possible supported formats should be scanned.

In most programs, the inclusion or exclusion extension list allows only 3-byte long strings. This is fine for .COM or .EXE files, but what about larger extensions like .CLASS – these have been found in the *Windows* world since at least *Windows 95*. Some scanners do not allow them to be scanned (unless in 'all files' mode), others look for extensions like '(*.)CLA(*)'. The latter is probably the best as there are often old volumes on file servers which cannot handle long file names. A user should be aware how the scanner handles such 3-byte extensions. Currently there are no known ItW viruses which infect files with more than a 3-byte extension but there are some zoo viruses which do. An interesting idea would be to export the 3-byte extension limit into the Unix world: some scanners under *Solaris*, *FreeBSD* or *Linux* show the same behaviour in this regard.

On-demand scanners usually use an extension list different from that of the on-access scanner (e.g. without archive file extensions). The on-access extension list cannot be configured in many programs, and in some scanners there is not even the option to scan all files on-access.

Another problem is caused by files with no extension at all. For example, many of the *Excel* macro viruses drop a file into the XLSTART directory. For this, many scanners have a special option on their default extension list – 'Scan files without an extension'. Unfortunately, not all of them handle extensionless files correctly – some do not scan for them, taking the real name as the extension – and often the file is left unscanned. If the option to scan all extensionless files does not exist, there is usually no way to add an empty value and all the files have to be scanned. A good point to make while discussing extensions is that no scanners seem to have a problem with double extensions like '.TXT.VBS'.

Scanning Options

Some scanners have really interesting default settings – usually they are optimized for speed, but not for security. Such settings start with a list of ten file extensions for the on-demand scanner to look for. No archives or packed programs will be scanned at all. Therefore, infected files could be missed, even if the virus scanner is capable of finding them. It would be better to scan all files by default, if not all archives too, in the first (automatic) scan of the whole system. If no virus is found, it can be switched back to an extension list until an infection is flagged.

Often, only one possible option exists for dealing with many types of infected files. Even on a desktop and especially on servers and mail servers, it is important to have different settings at least for macro and non-macro viruses. It would be better to divide them into boot, file, script viruses and other malware. For example, a user would be able to specify that script viruses and Trojans be deleted and macro viruses be cleaned.

Most of the time, there are different options for what to do with infected files – clean, copy, move (isolate), delete, rename, allow or deny access, print a page, beep and shut

down the computer, and so on. Occasionally these options can be used together (like rename and move) but, more dramatically, if the option fails, nothing will be done. It should be possible to have a second option in case the first fails. For example, 'try to clean, and if that fails, delete the file' is often used by customers. Some scanners, especially ones on mail servers or gateways, allow only one setting – if the cure of an attachment fails, it will be delivered …oops! An extra option to make a backup copy of the original file in a special quarantine folder before taking any action should be a standard setting.

An option to switch on or off the protection against back-doors and similar malware should be implemented. This would avoid legal issues and provide the user who requests it with real protection. It would be useful to add a switch for detecting jokes, too, since most home users want to have these programs while corporations do not. In some cases it would be helpful to exclude only some 'virus names' from the detection. This could not only be useful in the case of false positives, but also if the user wants to use NetBus (and only NetBus), for example.

Report Files

A standard report file should at least include information about the scanner, the version and date of both the program and the signature file(s), and the options used for scanning. The current date and time, the user and computer names should be included, too – at least once with a desktop product or with every entry if it is a server or mail server scanner. Some anti-virus scanners still do not include this essential information in their log files.

Every virus found and the action subsequently taken should be included in the report file, together with the full path, file name and why the action has been performed. With a mail server product, information about the sender and the recipient should be added. In our tests, we frequently came across unusable log files – the exact path or file names were truncated and replaced by '..'. In the case of archive files, only the file names could be found, but neither the archive name nor the path to the archive were located.

The log files should be exportable to at least text or comma-separated value (CSV) files. HTML-only log files are better to read if a browser is available, but should not be the standard or only setting, since they are infectable and harder to import into other programs. All entries should be separated by correct line feeds (e.g. 0x0d/0x0a for *Windows* programs) and the length of the report file should be unlimited. However, some anti-virus scanners currently have problems exporting log files with more than 1,000 entries. Really huge log files of several MBs will often be truncated at a random position without an error message.

In good documentation it should be possible to include all the files which have been scanned, not just the infected ones. For desktop products, it is useful to truncate the log files automatically if they are too big (1–2 MB rather than 50 KB), but on server software this option should be turned off by default. A short statistic or overview function of how many files have been scanned, how many are infected, how many have been deleted etc. is also useful.

Error Messages

Many AV scanners try to avoid displaying error messages and others' messages are incomprehensible, like 'PK-F-Init failed. Return code = 0x25628'. If an operation has failed, for example the removal or cleaning of a file on a write-protected drive, an error message must be displayed and included in the log files. Some programs do not do this – they look as if they are cleaning viruses correctly even if they cannot do this for physical reasons. So, the virus is still there, even when the program says it has been 'successfully' deleted or cleaned.

The same happens if a file cannot be opened, changed into a directory or scanned, if it is locked or if the user does not have the right to access it. Most scanners will skip such files without any notice. This is not acceptable, especially on *NT* or Unix systems with a user rights system. In the case of a password-protected (archive) file, a scanner should write a comment into the log file indicating which encrypted files cannot be scanned. Most of the time, the scanner will not report anything, or it will give a wrong 'OK' message or report internal errors, not specifying the real reason. Of course, the scan statistics should show the number of files which could not be scanned.

Translation

In some programs the translation of documentation is really ugly. This applies not only to error messages (some, translated verbatim, are nonsense), but also to the program itself and the on-line help. An example would be the use of the word 'exchangeable' instead of 'removable' in the case of a virus being cleaned. Others describe scanner options wrongly or are shortened – the English version is usually shorter than most other language versions. In this case, there should be enough space left for the translated strings.

Command Line vs GUI Versions

In many cases, command line scanners are much more powerful than GUI versions. Even if virus researchers and some companies choose this kind of program exclusively, it should be made clear that most, if not all, the additional functions are implemented in the GUI version, which is used more often in general practice. These functions include some speed-up or exact detection of viruses, and also recursive scanning for more types of compressed and archived files in memory. The GUI version only scans for certain files and then not recursively, with temporary extraction onto hard disk. Some complex polymorphic zoo viruses can only be detected with command line options, which are obviously not available in the GUI version.

Next month we will look at more complex problems.