

A Day in the Life of an Anti-Virus Lab

Robert Vibert 2000-06-17

If you're an average computer user, the first time you may hear of a new computer virus is when it appears in your morning paper or makes the evening TV news. However, by the time the general press gets wind of a new virus, virus researchers are already considering it old news and have moved on to the next bit of malicious code that needs tending to. Every month, hundreds of viruses are discovered and fewer than one percent of them ever make the headlines.

A look at the analysis processes used by researchers to handle the hundreds of new viruses each month reveals that they range from a mix of manual investigation and semi-automated software tools to some nearly fully-automatic systems with little, if any, human intervention.

Most new viruses arrive at a virus lab by way of customers who discover something amiss on their systems and send any suspicious files to their Anti-Virus vendor. These suspicious files are not always viruses - they may be damaged, or simply some innocent file that someone has sent along for analysis in the hope that it contains the answer to a mysterious problem from which their PC is suffering. For many users, a virus is the first suspect when something goes wrong, and Anti-Virus labs are constantly receiving innocuous submissions. Needless to say, if a file arrives from a customer, it receives top-priority handling. Other files received from researchers and collections which are exchanged between virus labs are dealt with after the customer files have been processed.

The wheat from the chaff

The first stage of the process at any lab is to sift out the definitely-innocent non-virus files from the potentially infected ones. Many labs use one or more processes to accomplish this, usually by identifying the known "normal" files by matching them against an ever-growing database of files which have been thoroughly analysed and verified as virus-free. Using checksums and other pattern matching technology, each new file is compared to the known clean files to determine if it should be subject to further investigation or simply culled from the analysis process. If a file matches a known-clean file, it is removed from the process and the customer informed that it is not a virus. On any given day, the ratio of safe to suspect files received in a lab can vary tremendously.

The next stage is to inspect the now reduced group of suspect files to see if any are known viruses. For this test, staff in an AV vendor's virus lab will use their own virus scanner set to its maximum detection levels. Some will also use a number of virus scanners from competing vendors to increase their chances of pruning out the known viruses. Those files which are identified as known viruses by the lab's internal scanner are once again put aside from those which are still an unknown quantity and require further analysis. If they are viruses and were picked up by a competitor's scanner and not by the in-house scanner, they must be considered to be in the wild and detection needs to be added to the product represented by the lab.

The resulting files are then analysed for viral content, using an array of tools. Given that there are two main families of viruses seen these days, program file viruses and macro viruses, they are often sent for specialized analysis. Each file is dissected, probed, and prodded to determine just what it contains. Mike Pavluschick, a virus researcher at Kaspersky Lab in Russia, says "Within 5 seconds of looking at the source code of a macro, I can tell if it is a virus or not." Of course, one has to get a look at that source code first, before making such lightning-fast assessments.

To peer inside files and decipher their contents, virus labs use a range of software tools. One popular one is the IDA Pro disassembler from DataRescue in Belgium, which decomposes program files into their basic instructions, permitting virus analysts to examine the flow of code. Although to the layman the reports produced by a program such as IDA might look like so much gibberish, to the highly trained anti-virus researcher, it is easy to find the characteristics of a virus amidst the programming instructions.

Most anti-virus labs have also developed their own tools, some of which remain for their internal use and others which they share with other virus researchers. For example, Dmitry Gryaznov of Network Associates has developed a number of tools, including one called SCRID, which aids in the analysis of text-based viruses and worms, such as those using VBS (LoveLetter et al.). Fridrik Skulason of FRISK Software in Iceland, another long time virus researcher, has made available his program F-VBACRC, which facilitates the creation of checksums for new macro viruses.

Internal software tools are abundant at virus labs, performing such tasks as extracting macro source code from the wide range of macro viruses, creating detection patterns to be used in the scanning products, verifying if internal OLE structures in files are correct, comparing the macro code in one sample with a database of known macro viruses, etc. In many cases, these tools

are used by the professionals to simplify their work and have grown out of the repetitive nature of virus analysis. The vast majority of viruses seen each month in the labs are derivative works of existing viruses. Only occasionally does a truly new virus come along, and these require a special treatment.

How many thousand copies did you need?

Once a suspect file has been determined to contain what looks like an actual virus, it must pass a simple test - it must be able to self-replicate. There are a number of definitions of a virus, but they all come back to the same basic feature - the program must have self-replication capability. In the case of non-polymorphic viruses, it is sufficient to replicate a few copies of the virus, as each should look identical to the other. Prepared so-called "goat" files, which are "sacrificed" on the altar of research, are used to prod the virus into infecting. Each goat file has a well understood structure, which makes it easy to detect exactly how the virus infects a file and what changes it makes.

Goat environments are also used, with complete systems comprising of common configurations of the different versions of Windows operating systems and Microsoft Office applications, for example. Modern viruses often interact with the applications on computers, exploiting security holes in programs such as Outlook, Access, Excel, and Word. To understand the actions of these viruses, a simulated environment is needed to let the virus do its dirty work, while under constant observation. Once again, once the goat environment has been infected, careful inspection is made to determine all the actions undertaken by the virus are understood. This can be especially important in deciding what level of alert needs to be issued to customers. If the virus is fast-spreading or carries a potentially destructive payload, users need to know about it as soon as possible.

In the case of polymorphic viruses, thousands of samples must be replicated to ensure a proper detection pattern is created, as each infection of the virus is designed to look different from the others. Most virus labs have automatic replication routines and facilities established to provide an environment for virus files to replicate. These test beds must be returned to a pristine state after each replication exercise, to ensure that there is no cross-contamination of one virus with another. The more polymorphic the virus, the more samples must be created to ensure that all permutations are understood.

Catching the critters

Having determined that a file is in fact a new virus, the next stage is to create a means of consistently detecting the virus, without triggering any false alarms on otherwise innocent files. Virus researchers such as Joe Hartmann of Trend Micro, are constantly working to improve the detection capabilities of their products. "At this point, we often use patterns rather than exact identification (such as CRC checksums). However, one of my goals is to make the Trend engine more accurate in its virus detection."

Once a researcher has created a virus detection pattern, it must be thoroughly tested. The replicated samples from the replication stage can be re-used for this purpose. The virus definition must be able to catch all infected samples - if one or two are missed, they will have the potential to spread undetected. The pattern is also tested against massive libraries of known clean files, in an effort to ensure that no innocent file is incorrectly identified as infected - what the virus researchers call a false positive.

With the virus properly analysed and understood, and a detection routine or pattern ready for customers, a description is prepared of the virus and its effects. Some labs have tools which simplify this task, providing templates for the researcher to complete, in order to ensure that the details of the virus and its effects are properly documented. At this point, the customer who submitted the virus sample is contacted with the information and the update to the detection database which can be used to remove the virus from their systems.

If the virus has been identified as "hot", a news release is prepared for sending to customers and to the general press. "Hot" is the term used to describe a virus which has the potential to spread very quickly, as was the case with Melissa and LoveLetter, or to cause serious damage, as was the case with CIH. The news release will usually indicate the characteristics of the virus, how widespread it has been up to that point in time, what damage it may cause, and where to get updated protection from it. Most Anti-Virus companies have mailing lists for their customers and send the information to them at the same time as it is sent to the press. A hot virus will also warrant that the update be posted to the vendor's website immediately and that their online virus encyclopaedias will also be updated. Viruses which are not considered as likely to be of immediate concern to most users are handled in the regularly scheduled updates to both the detection and information sources.

Autopilot engaged

Some vendors, such as Trend and Symantec, are hard at work on automating their processes to

ensure that suspect files are collected from customers in an orderly and controlled fashion, analysed systematically and any responses (explanations and virus detection updates) returned to them within a suitable time frame. This time frame is typically described as between two and forty eight hours, depending upon such variables as the workload in the lab, the complexity of the virus, and the need to incorporate any new technology into the virus detection engine.

Trend has a system called TICK (Trend Internet Customer Key), which provides an automated Internet-based facility for submitting suspect files and receiving updates and answers. The TICK system relieves both customers and virus researchers from the time wasted talking on the phone. Customers can upload a suspect file to TICK, have one or more of the fifty virus researchers in Trend's Philippines labs analyse it and report back to them with a fix and description. Customers can also track the progress of their file using TICK, much as UPS customers can track their parcels being shipped. The target response times range from a few hours for a simple request for information to twenty four hours for a complex new virus or malware.

Symantec has been working together with IBM for several years on what they call the Immune System. IBM first presented the theory behind the Immune system in an article published in 1993. At Anti-Virus industry conferences throughout the 90s, IBM researchers presented their developing concepts, and a working prototype was eventually demonstrated in 1997. In 1998, IBM stopped work on their own Anti-Virus product and signed an agreement with Symantec, under which development of the Immune System would continue. Currently, the system is in trials with users and is expected to go live before the end of 2000.

The Immune System is an ambitious project, involving the most automated virus analysis system to date. Samples received from customers undergo all the processes explained above (from collecting a suspect sample on a client's machine to providing the cure), but without human intervention - the entire effort is conducted in an automated system. Only if a sample cannot be automatically handled will a human virus researcher intervene, with two goals in mind - to discover what has caused the problem and provide a manual fix, and then to improve the design of the Immune System so that it can handle the anomaly itself next time round. The stated objective of the Immune System is to respond to viruses within a few hours of their appearance, thereby reducing the chances of any major outbreaks.

As the number of new viruses increases, the Anti-Virus vendors will be continually investing in personnel, tools and automation to improve the speed and accuracy of their virus labs. And,

chances are, by the time users are hearing about a new threat, protection is just a few mouse-clicks away.

Relevant Links

[Anatomy of an Immune System](#)

IBM AntiVirus Online

[Automatic Extraction of Virus Signatures](#)

Jeffrey O. Kephart and William C. Arnold

[More Anti-Virus Advice](#)

Segura Solutions

Anti-Virus Labs' Submission Links:

[Command Software](#)

[Computer Associates](#)

[F-Secure Corp](#)

[Frisk Software](#)

[InoculateIT](#)

[Kaspersky Labs](#)

[Network Associates \(US\)](#)

[Sophos Plc.](#)

[Symantec](#)

[Trend Micro](#)

[Privacy Statement](#)

Copyright 2006, SecurityFocus