

WHAT IS WILD?

Sarah Gordon
IBM TJ Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
sgordon@watson.ibm.com

Abstract

“In the Wild” virus detection is part of the criteria of *National Computer Security Association (NCSA) Anti-virus Product Certification*, *SECURE COMPUTING Checkmark Certification*, the proposed *UK IT Security Evaluation and Certification (ITSEC)* anti-virus product certification and other product review and evaluation schemes. However, companies which use “certified” products, based on “In the Wild” (ITW) detection continue to suffer the effects of viruses. This paper considers the various definitions of “In the Wild”, as well as how well the “In the Wild” criteria as defined by the individual testing organizations measure the ability of products to deliver adequate protection. Inherent problems with such approaches are discussed from both a development and user perspective. Some alternative testing, development and protection strategies are offered.

Introduction

There are currently over 10,000 computer viruses in existence. Most of these have little likelihood of spreading and exist only in collections; they are known as “Zoo” viruses. Even an anti-virus virus researcher would be hard pressed to list a significant percentage of these viruses, let alone provide detailed information on how they operate. Most users have only even *heard* of a handful of them. Yet when a virus is encountered within a company, it is usually the case that a call to an anti-virus vendor, or a search through a virus encyclopedia will provide further information on that particular virus. This is because vendors, researchers and testers have begun to focus their attention on those viruses “In the Wild”.

The concept of “In the Wild” is an important one. Tests of anti-virus software have, until recently, focused on Zoo detection figures. These tests did not necessarily measure the ability of a product to meet the real world threat [1]. Consider two products tested against a corpus of infected files: by simply measuring which product detects more infected samples, we would be given no information concerning how well the product detects and repairs those viruses which are known to pose an active threat to a real world PC. A meaningful test of the efficacy of a product would be to measure the product’s ability to detect and remove *those viruses that the user is likely to encounter*: that is, *those viruses that are “In the Wild”*.

In order to understand the issues surrounding “In the Wild”, we will examine a history of the term. As far as we can determine, the actual phrase “in the wild” was first used informally to describe real-world virus incidents by Dave Chess, of *IBM’s TJ Watson Research facility*, in 1990/91 [2]. Around this time, Alan Solomon remembers using the term in telephone conversations in the UK [3]. The phrase subsequently cropped up in *Virus Bulletin* in 1992, in a message from Roger Riordan [4], where he referred to real-world incidents: “As Dave Chess pointed out on Virus-L (May 8th, 1991), few of your specimens have ever been seen in the wild...” [5]. It formally appeared in 1992, in “Measuring Computer Virus Prevalence”[6] where it was shown that a small number of viruses accounted for most actual virus incidents, i.e. were “in the wild”. Early *Virus Bulletin* tests featured an “In the Wild test-set”, a collection of viruses designed to measure real-world performance. The contents of this list were garnered from virus reports sent to *Virus Bulletin*, along with those viruses which researchers believed to be spreading, as opposed to those which were known to exist but which were not observed to be spreading (Zoo samples). While this was not entirely scientific, it is the first test of which we are aware that made a reasoned and logical attempt to move away from Zoo testing. During this time period, most new viruses were not initially discovered spreading in the wild and the vast majority of Zoo viruses were not considered to be an active threat. At the same time, Zoo testing remained prevalent [7] and users tended to judge products based on how many viruses the product could detect.

There were some obvious problems with this approach. As the number of viruses “In the Wild” continued to rise, the need for a better definition of “In the Wild” led to the creation of *The WildList* by Joe Wells [8]. Wells’ idea

was to take the concept of “In the Wild” used by *Virus Bulletin* and expand it internationally. To this end, he culled virus reports from virus researchers worldwide. Any virus that was reported by two or more researchers was considered to be spreading in the wild. In some cases, viruses were reported by only one contributor. Those viruses were placed into their own section of *The WildList*. (This supplemental list provides some idea of which viruses might be moving onto or off of the main section of *The WildList* and tends to be a more regional reporting mechanism.) A supplemental frequency table has been recently added. It does not show how common each virus is. Rather, it is *The WildList* sorted by the number of participants that report each virus. It gives the names, types, and aliases of the most frequently reported viruses. These viruses have been reported by at least one third of *The WildList* participants. They are sorted with the most frequently reported first. *The WildList* clearly states it should not be considered a list of the most common viruses, as no commonness factor is provided; however, it can serve to help determine which viruses are spreading.

The WildList represents the most organized effort to catalogue those viruses which are spreading, yet it would be wrong to define “In the Wild” as those viruses which are listed in *The WildList*. Rather, *The WildList* should be considered to be a subset of this set - a core set of viruses that every product *must* be able to detect. Making this subset more complete is fraught with problems, and falls prey to different definitions of what “In the Wild” actually means. Before considering the impact of tests based on *The WildList*, we shall examine some of these problems.

- ***The WildList* lags behind viruses spreading in the wild.**

Time delay is by far the most commonly reported complaint against using *The WildList* as a complete set of viruses “In the Wild”. Put simply, the logistics of compiling reports from more than 45 contributors worldwide, who in turn have to compile their own list of Wild viruses based upon their technical support calls, can quickly date the contents of *The WildList*. In practical terms, a virus may make it onto *The WildList* two or even three months after it is first discovered at a user’s site. At present, this is difficult to improve upon, though streamlining analysis of submissions may help. Procedures to facilitate this have now been implemented, by way of automatically distributed, standardized reporting forms for participants, which will decrease the time required to process incoming submissions. These new forms should make reporting much simpler for the volunteer reporters, and it is hoped this may help aid in getting submissions in faster.

- **The viruses on *The WildList* are those viruses reported, not necessarily those viruses which are in the wild.**

As *The WildList* contributors are mostly made of those working within the anti-virus industry, it is not unreasonable to assume that *The WildList* represents those virus infections which are reported directly to developers/resellers. However, this group of viruses is not necessarily a complete list of viruses “In the Wild”. Consider the case of a hypothetical virus named Foo, spreading in the wild. This virus is detected and removed perfectly by all anti-virus products. Also, consider another virus spreading, called Bar. When a certain product detects Bar and attempts to remove it, it corrupts the file. It is entirely believable that anti-virus companies will receive far more reports of the Bar virus than the Foo virus, even if they have equal prevalence. Thus, researchers may preferentially receive reports of those viruses which are not adequately dealt with by anti-virus products.

One apparently obvious solution to this bias this would be to include businesses (as opposed to solely developers, resellers and researchers) in *WildList* reporting. However, although many companies require that all virus incidents are reported to a central body, some studies of computer virus epidemiology strongly suggest there are other problems with organizational reporting. Indeed, it is possible that problems with corporate statistics can reflect the inclusion of wild guesses, reporters being unaware of virus incidents and reporting bias toward problematic viruses [9].

- **The samples named in *The WildList* may not be the same viruses actually spreading in the wild**

This question highlights one of the areas which is actively being improved within *The WildList*; correlating reports of viruses *by name* with actually binary samples of infected files. It is believable that discrepancies over virus naming could both lead to viruses being inadvertently added to *The WildList*, as well as viruses being omitted.

Consider the case of a single virus, Foo, which is identified by product A as Foo.A, and identified by Product B as Foo.B. In such a case, if researchers simply correlated reports from the field with their own virus collections, both Foo.A and .B might be placed on *The WildList*. Next, consider two viruses, Foo and Bar, which are both identified by products A and B as FooBar. In such a circumstance, only one virus name would be added to the list of those viruses known to be “In the Wild”, where in actuality, there should be two different entries.

To help solve this problem, Wells has added additional criteria to those contributing to *The WildList*: if a sample is being reported for the first time, the reporter must supply a sample of that virus along with his report. This allows submissions of the same virus reported by different name to be caught, and similarly allows one to discriminate between two different viruses inadvertently identified by the same name. *The WildList Organization* has begun distributing the responsibilities associated with *WildList* sample replication and identification of the viruses amongst *The WildList* board members, and is including a peer review process for samples. It is hoped this will significantly decrease the workload and increase the naming accuracy.

***Note that even though we have outlined a number of shortfalls in *The WildList*, the author still believes that it is currently by far the best resource for tracking those viruses which are believed to be in the wild.**

Tests Based Upon *The WildList*

Using confirmed samples of every virus on the list as a test suite for testing anti-virus software can tell you whether or not a product detects the viruses on the list. While this is clearly only a minimal test, by monitoring the tests over time we should theoretically be able to determine whether or not a vendor continually meets the test conditions. The problems with practical aspects of doing this will be addressed later in this paper. Also, it is important to remember that while *The WildList* clearly defines what is meant by “In the Wild” for the purposes of the list and for tests which use the list, it is not a definitive measure of viruses that are causing incidents. For example, certain viruses have been found spreading in isolated areas of the world. While they are definitely in the wild causing incidents, the fact that they are reported by only one person keeps them from being included in the main section of *The WildList*. This creates a problem for users who rely on ITW-based certifications as the only measure of a product’s effectiveness, because current certification schemes, as will be shown later, do not test against even the upper portion (viruses in the wild, reported by 2 or more contributors) of the most current release of *The WildList*.

In this section, we will briefly examine three testing/certification schemes, before discussing how much assurance each gives to the user that his/her product will truly detect those viruses which are known to be spreading. Interestingly, each scheme uses a slightly different definition of “In the Wild” for the purposes of its tests. With this in mind, we will then re-examine *The WildList* as the baseline measure used by several certification bodies.

NCSA Criteria

Founded in 1989, the NCSA is a for-profit organization based in Carlisle, Pennsylvania. Its anti-virus product certification began in 1992. The main thrust of the criteria currently is to provide a way to measure the effectiveness of detection capabilities of virus scanners. The scheme requires that the scanner components of certified products detect 100 per cent of viruses found on the upper portion of *The WildList*, using a *WildList* that is two months old at the time of testing. This is said to allow for development time. We will now briefly examine relevant aspects of the scheme.

According to documentation published on the NCSA World Wide Web site, “NCSA tests and certifies that anti-virus scanners pass a number of stringent tests.” As our own tests have shown that some NCSA certified products should not pass the documented certification criteria, we asked NCSA for information regarding their virus test suite, to see if we could determine the cause of the discrepancy. At NCSA’s invitation we visited its’ virus lab where several virus test suite related problems were noted. One problem we noted was related to the replication of polymorphic viruses. Some viruses attempt to hide from virus scanning programs by keeping most of their code garbled in some way, and changing the garbling each time they spread. When these viruses run, a small header “de-garbles” the body of the virus and then branches to it. A polymorphic virus’ de-garbling header changes each time the virus spreads. The polymorphic test-set had not been fully replicated; only 6 viruses had been replicated. As some products may have unreliable polymorphic detection, a more complete polymorphic test suite is desirable. Ideally, all polymorphic viruses “In the Wild” should be replicated onto appropriate hosts, but this is a difficult and time-consuming task. Additionally, some viruses are multi-partite, which means they are capable, for example, of infecting not only files, but Master Boot Records of hard disks, or floppy disk boot sectors. These types of viruses should be replicated onto all appropriate media; we observed that this is not being done in NCSA tests at this time. Again, this replication presents some unique problems and will take some time to sort out. The macro virus test-set consisted of two replications of each macro virus except for ExcelMacro.Laroux, of which there was only one sample. This number of replicants is insufficient to allow for measurement of the reliable detection of macro viruses. There were not any macro viruses replicated onto *Office97* documents. As some of the macro viruses will replicate upwardly into *Office97* documents, inclusion of such documents is required in order to measure the level

of protection afforded the user. Nine of the boot sector viruses had not yet been replicated; work is currently in progress to rectify this. It can be difficult to replicate some of the viruses, requiring special expertise and in some cases, additional equipment. Plans are underway to initiate testing of master boot records of hard disks; however NCSA is some time away from this type of test. NCSA has demonstrated its commitment to fully expanding its test suite. NCSA virus lab technicians are currently in the process of solving these virus related problems by replicating the macro and polymorphic viruses into a larger suite, and by replicating the multipartite viruses onto appropriate media. According to NCSA spokesperson Jon Wheat, these issues are a "top priority".

Some administrative problems should be noted. Here, we are concerned with the timeliness of the tests and consistency of the scheme. On March 17th, 1997, the NCSA Web Site listed *F-PROT Professional 223a* as a certified product. Version 2.23a was released in August 1996. Similarly, *Eliashim's ViruSafe* version 7.1 was shown as certified; however, the most current version of *VirusSafe* on March 17th was 7.3. Two *Intel* products (*LanDesk for NT* and *LanDesk for NetWare*) were shown as last being tested in December 1996. *Dr. Solomon's* software versions 7.65 were the most current tested versions; their current release on March 17th was 7.69. When asked about the discrepancies, NCSA stated these products would be retested. As documented in [10], the scheme has not been without problems; however, at the time of the completion of this paper, these problems appear to be in the process of being resolved, and products appear to be tested on a regular basis.

NCSA is working hard to make the tests of the scanners complete, thorough and accurate and to keep the administrative aspects of the scheme functioning smoothly.

SECURE COMPUTING Checkmark

SECURE COMPUTING magazine, published by *Westcoast Publishing*, regularly reviews and evaluates anti-virus software, provides a venue for marketing of various security software products, and offers security-related articles for its subscribers. A recent addition to these services is the *SECURE COMPUTING Checkmark* scheme, which is designed to establish a standard for computer security products, test them against that standard and produce a certificate which shows that they meet the standard. They claim this is similar to governmental standards that seek to indicate to a buying public whether they can have faith in certain products. According to the *SECURE COMPUTING* Web page:

"Whatever your needs, you should know that the products you are buying are worth the money and give you a sensible level of security...It shows that the product has been tested and approved to an industry-recognized standard by an independent organization....There are also one or two other schemes run by private companies or by students in universities but in our opinion these are not worth bothering with..." "To obtain a Checkmark an anti-virus product has to detect all the viruses that are in-the-wild; that is those which are actually out in the real world causing infections (not held in the private collections of anti-virus researchers). The Checkmark in-the-wild list is updated on a monthly basis. Products are tested on the basis of the in-the-wild list current three months previously and there are a number of practical reasons for doing this. It is a fairly typical approach and in the real world gives a high standard for the anti-virus developers to achieve."

As with the NCSA library, not all multi-partite samples have been replicated onto both files and boot sectors; this process is underway. Tests of boot sector infections are done on real infected floppies; there are as yet inadequate resources for testing of Master Boot Records of virus infected hard disks. There are a thousand replicants for each polymorphic virus which has been replicated so far. *Secure Computing* will be including *Office 97* capable *Word* viruses now that some have appeared on *The WildList*; they are not testing any of the upwardly mobile *Office 95* viruses on *Office 97 Word* goat documents until these viruses are explicitly reported on *The WildList*.

This scheme could attempt to address the problem of *The WildList* lagging behind the current threat as it retains the option to add viruses at the discretion of the test administrator [11]. It remains to be seen if this results in tests which provide a good measure the protection provided; however, all indications are that the tests should be thoroughly and competently performed. *Westcoast Publishing* is well positioned to promote the scheme in both the United States and Europe, using *SECURE COMPUTING* magazine as well as its' recently purchased *InfoSecurity News*.

ITSEC Certification

The proposed *UK IT Security Evaluation and Certification (ITSEC)* model of anti-virus software certification consists of several criteria, all of which are designed to measure how the product meets the dynamic real world threat. Several anti-virus product vendors have been involved in helping draft guidelines, and the specialty

magazines *Virus Bulletin* and *SECURE COMPUTING* have also sent representatives to the *Anti-Virus Working Group* meetings. The evaluation process is in the developmental phase although significant progress has been made in the past year, particularly in the area of formalization of criteria. The main areas with which the process is concerned are Standard, Threat Assessment, Virus Attack Techniques, *Anti-Virus Working Group Virus Collection*, Comprehensive Virus Collection, "Advice Documentation", and Certificate Maintenance Scheme.

With the *ITSEC* scheme, an increasing level of stringency would be applied and associated with the commonality of the virus or observed technique, i.e. weighted testing. The current plan is to perform tests with common and wild viruses (note that *ITSEC's* definition of "In the Wild" is not the same as that used by *The WildList*) listed concurrently and cumulatively and to require a 100% score to pass. Common viruses are defined by the *ITSEC* scheme as those which are frequently reported as causing attacks; it defines "In the Wild" viruses as those having been recorded as responsible for attacks. Determination of which are common and which are "In the Wild" is to be made by a national authority that monitors the changing situation reported from worldwide centers of virus expertise. The current strategy for Zoo testing is detection of at least 90% of the different named viruses in an approved collection for a passing score. The *Anti-Virus Working Group* recently announced that the University of Hamburg has agreed to act cooperatively with the *ITSEC* evaluation scheme and do Zoo testing on-site at the University, as part of the evaluation process. Other collections may be used, providing they meet certain requirements, yet to be determined. In addition to this detection criterion, a number of other proposed criteria are put forth in the formal documentation provided by the *ITSEC Anti-Virus Working Group* drafts for Standard Functionality Class for Anti-Virus Products. These criteria include, but are not limited to, areas including recovery means, false positive levels, common compression, self-checking, logging and naming. Discussion of these is beyond the scope of this paper.

A Virus Attack Techniques Encyclopedia has been developed (under contract) by the *Anti-Virus Working Group*. This document is intended to detail all known techniques used by viruses, and currently includes the following: boot record infectors, parasitic viruses, multipartite viruses, companion viruses, stored code modification, environmental format considerations, stealth, execution infectors, system infectors, interception of system services, defense mechanisms, payloads, permanent configuration changes, hardware/software specific viruses and macro viruses. It is a dynamic document. The encyclopedia will be used to help in formulating ways to more fully analyze and test products; for security reasons, it is a limited distribution document.

By attempting to measure a product's performance against the threat by scanning a comprehensive large collection of all viruses, testing extensively against those viruses which are known to be "In the Wild" according to designated reporting authorities, and measuring product abilities against a range of different attack strategies, the *ITSEC* scheme is focusing on the current and *future* "In the Wild" threat. By evaluating the product's ability to defend against the different techniques used by viruses, they hope to provide a measure of a developer's ability to track a rapidly changing threat. The *CLEF* would maintain close contact with the developer of the product currently under evaluation, with developers being required to demonstrate that not only are they up to date with the current threat, but that they have in place sufficient procedures to monitor the threat as a function of time and update the software to meet this threat. This would be documented through the use of the Certificate Maintenance Scheme, which includes extensive paperwork on the part of the developer to document their resources and plans in various areas including intelligence activities related to monitoring the threat, threat analysis and countermeasures. This "vendor evaluation" is something that almost no other evaluations of anti-virus software includes, and is one of the biggest benefits of the proposed *ITSEC* approach. It is also one of the areas which appears to meet with the most resistance within the USA. One concern which has been cited is the sharing of information between *CLEFs*: "Even though the UK requires that all techniques and lessons learnt from evaluations be documented at the end of an evaluation and made available to the UK evaluation community, it is felt that *CLEFs* prepare this information from a position of non-disclosure of information which is of a proprietary interest to them. There is some concern UK based evaluations, by virtue of their commercial nature, do not encourage the sharing of evaluation techniques amongst the evaluation community" [12].

There is another potential problem with this type of approach. As documented in [1], this solution can lead to possible problems as new threat types may be as yet unanalyzed, and the virus itself is not in the wild. There is no guarantee as to the time sequence that a virus may be found to exist, be found in the wild, be obtained and analyzed by an evaluation or certification service, and its threat type documented. This is illustrated by the recent spate of macro viruses, where initially there was a noticeable lag between the knowledge of the threat type by anti-virus researchers, the discovery of the first in the wild *Word* macro virus [13] and the first in the wild *Excel* macro virus

[14], and the implementation of detection and prevention for these virus threat types on the part of some developers.

Finally, there are problems with issues of legal liability. Whereas German law demands someone be liable for failure in *ITSEC* certified products, the United States makes specific disclaimers assuming no responsibility. Drawing again from Borrett [12], we find “the political implications of legal liability for Europe and North America merits further investigation. In the interim, it may suffice to place an appropriate caveat alongside any US evaluated products which appear in UK Certified Product List publications.”

Future Trends: New Paradigms and Epidemiological Shifts

While each of these schemes use *The WildList* as a basis for wild virus detection only one (*ITSEC*) represents more than a series of snapshots of particular product’s detection. We see several dangers associated with the current situation. In order to better illustrate these dangers, let us first build a perfect set of review criteria. Note that here we shall attempt to address only those aspects required for virus detection; properties such as virus removal, product usability and technical support are beyond the scope of this paper.

The shift from Zoo to “In the Wild” testing marked the beginning of a move towards measuring the protection provided by a product. However, this shift is only the beginning of a true measure of protection provided. A cursory examination of *The WildList* shows us that a particular computer is more “at risk” of infection by certain viruses on the list than certain others [15]. For example, Ping_Pong.B and Wazzu are both on *The WildList*, yet few would argue that for the average computer, the probability of infection with Wazzu is considerably higher. However, in most tests carried out against the set of viruses catalogued in *The WildList*, each sample is equally weighted. Clearly, this is not a complete approach; in a “perfect” world, we would weight each virus by the actual probability one had of encountering it and it effecting one’s work. Extrapolating onward, we would include all Zoo viruses in this weighting; for example, those viruses which are difficult to replicate would have a low rating (not in the wild, and not likely to spread even if released), whereas those viruses which have been actively circulated in newsgroups and which are highly viable in the wild would have a higher rating. However, even this approach is not complete. It is easy to argue that while the overall features of such a weighting scheme for viruses would vary relatively slowly as a function of time, its details may fluctuate rapidly. Consider, for example, a situation where a certain virus is distributed widely on a set of mass-produced CDs. The threat posed by this virus (that is, the probability that you will encounter it) has increased somewhat, even though it may have only actually infected one PC at this time. Another layer of complexity which we will not address here is that such a weighting scheme would vary depending on whom the review was being carried out for; *Word* macro viruses, for example, pose little threat to those who do not use *Microsoft Word*.

Initially, we believed that testing based on criteria that involved this type of weighting was impossible. We have since determined that the tests could be done using data gathered from *IBM* studies. However, problems with formalizing such a scheme remain. While using the data to formulate test criteria that could measure threats on a global scale is feasible, we believe certification using these methods is not practical at this time, due to the need for the certification body to independently gather the necessary comprehensive data. Other methods of measuring real world virus prevention provided by a scanner need to be compared to this model. Making such a comparison, we observe that for each of the certification bodies we have examined, all fall short in terms of the currency of the viruses used for testing. At one time, we believed that this was not a serious problem [16]; however, recent shifts in the way viruses appear in the wild are rapidly altering this perspective.

The most serious change in the ways viruses spread since perhaps the beginning of the virus problem is posed by macro viruses. These viruses attach themselves to data items that are frequently shared. Moreover, this sharing is often done via the LAN or Email, making such macro viruses highly virulent. Indeed, macro viruses have been so successful in the wild that the two most reported viruses to *Virus Bulletin* in January 1997 were both macro viruses: Concept and Npad. We have observed that once a virus begins to spread rapidly, it can reach epidemic proportions within an organization very quickly. It is the combination of large spread rate and lag in *WildList* testing times of *WildList*-based certification schemes which poses the biggest problem to those relying on *The WildList* for certifications. Since a virus must be reported by two or more *WildList* contributors, it is possible for a virus to be rampant within one organization and still be observed by only one *WildList* reporter. By the time a virus discovered in the wild is actually observed by two reporters and included in the certifying body’s test-set, the virus may have already been spreading within any given organization for several months. A good illustration of this is the Concept virus. Discovered in July 1995, the virus first appeared on *The WildList* on Sept 10th, 1995. Thus, by the rules of a certification body using the criteria of detection of a collection based upon a two month old *WildList* compliancy, a

product which was certified would only be required to detect this virus by Nov 10th, by which time it was already spreading rapidly in the wild.

Another problem has developed which may impact the ability a certification body has to measure that vendor's ability to meet the threat posed by macro viruses. A macro virus which replicates under *Office 95's* version of *Word* may be automatically converted by *Word* to the new *Office 97 Word* format. This is referred to as 'up-conversion'. The problem is related to a controversy within the anti-virus community regarding this up-conversion of *Office 95 Word* macro viruses and testing of anti-virus products. Some anti-virus researchers have indicated they feel this up-conversion of in the wild *Office 95* macro viruses is the creation of "new" viruses, and as such, represents an unethical act for any anti-virus product tester. Others researchers maintain the opinion that *Office 95 Word* macro viruses which are in the wild and able to replicate into *Office 97* documents, (via *Word*), should be part of the in the wild test set in both their *Office 95* and *Office 97* form, as to do otherwise could expose users of certified products to unnecessary risk.

Is such testing required to make sure users are adequately protected [24], as part of an *ITW* based certification, or would this be an unethical act of irresponsible virus creation? No one can argue that the *Office 97* viruses are in many ways different from their *Office 95* origins. However, we question whether this difference in physical structure, form, and language supports the contention that these are in fact totally different viruses and that replicating the *Office 95* virus onto an *Office 97* document is unethical virus creation. It is the opinion of this author that such arguments are counterproductive and that certification bodies which perform *ITW* tests and certifications should simply replicate *Office 95* macro viruses onto *Office 97* documents, using due diligence in the care of such samples, as part of these *ITW* tests and certifications. As an industry, the anti-virus industry has long held the position that virus creation for any reason is unethical. This belief has been somewhat altered by the necessity to perform tests of viruses generated by virus creation 'kits', and the need to generate multiple polymorphic samples to allow for reliable detection and disinfection. The evolution of the virus threat may force us to re-examine our beliefs yet again.

Another serious problem for certification bodies brought about by macro viruses is the vast numbers of variants we are observing coupled with the concept of "In the Wild". Virus exchange sites appear to be less prominent for macro viruses, than is the case for file and boot sector infecting viruses. The majority of these macro virus variants are being discovered already spreading in the wild. We believe that there are a number of reasons for this. First, as current macro viruses are written in *WordBasic*, they essentially carry around with them a complete copy of their source code [17]. As the language is both simple to use and powerful, viruses are easily modified and released. Second, we have observed seemingly random corruption of macros within the *Word* environment. While we are as yet unable to reliably recreate such corruption in a laboratory environment, we can see that macro viruses seem to be more resilient to such corruption than binary viruses. Thus, whereas a corrupted binary virus frequently renders a virus non-functional, many *Word* macro viruses are quite capable of replication even when corrupted, leading to creation of a new variant. Thus, we have observed certain *Word* viruses spawn many variants in just a few months - something which rivals even the most prodigious of "ordinary" viruses. This rapid rise of new strains discovered in the wild has further clouded the concept of "In the Wild," as well as reduced the value of certifications carried out against *The WildList*. A more forward-looking approach would appear to be that described earlier as taken by *ITSEC*, which attempts to certify a company's ability to meet the current and future threat. It would appear that in terms of protecting the user, the most critical question is no longer whether a company can detect a specific virus, but how quickly that company can meet a new threat.

Some people have argued that all viruses are effectively "In the Wild", as many virus collections are available via virus exchange bulletin boards and web sites. However, a virus which is found on a Bulletin Board System or web site may not be viable in the real world. In 1992-1993, we examined the relationship between viruses found on virus exchange BBS compared with those known to be causing incidents [18]. It was determined there was little if any reason to believe viruses on underground BBS contributed significantly to the population of viruses spreading in the real world. The majority of these viruses simply were not found to be spreading. At the same time, individuals were reporting (and continue to report) infections caused by some of these barely viable viruses; this may be a result of the users obtaining the viruses and using them for testing (or reporting) purposes.

In 1994, we began to observe a change in the nature of virus exchange and distribution. It was concluded that with the growth of the Internet, viruses could reasonably be expected to spread using several different models. Specifically, Web virus distribution was predicted to make viruses widely available to the general computing population should they desire to obtain them; Usenet news was shown to be a potential distribution media for

viruses for both the willing and unwilling, and the Internet itself was examined as a potential hotbed for viral spreading which could occur almost instantly and worldwide. "The system is the perfect medium to host and transfer the very programs designed to destroy the functionality of the system itself"[19]. Whether or not the increased availability and relative anonymity afforded by the Internet will contribute to in the wild virus population remains to be seen. Viruses that have been released via Usenet have not become rampant in the wild. However, certification bodies who rely on detection of those viruses "In the Wild" should keep a careful eye on the role of global inter-networking, lest they be taken unaware by a paradigm shift in the way viruses spread. We are already beginning to observe real problems in this area, which we will discuss later in this paper.

Threat and counter-threat

The need for a new method of reviewing and certifying anti-virus software becomes more apparent when we examine some of the new threats resulting from the increased use of networks and desktop Internet connectivity. Although we have yet to see a virus spread in minutes/hours on a global scale via Email, we believe that the potential for such a virus exists. There have been several precursors to such a virus; here we shall discuss two of them: CHRISTMA EXEC [20] and ShareFun [21].

CHRISTMA EXEC is a well-known "chain letter", which was released on December 9th, 1987. It is a good example of how an e-mail worm can impact a network: CHRISTMA EXEC spread across BITNET, EARN and IBM's internal network, dramatically slowing the IBM worldwide network on December 11th, 1987. The program, written in REXX, spread on VM/CMS installations, and displayed a Christmas tree along with a message, before sending a copy of itself to all of the users' correspondents in the user files NAMES and NETLOG.

ShareFun.A is a macro virus which spreads by infecting *Word* documents, and as such, operates just like most other macro viruses. However, ShareFun.A attempts to spread via desktop e-mail, attempting to send mail messages to addresses listed in the users' address book. The message has the subject line "You have GOT to read this!", and it carries with it an attachment which contains the infected document. Fortunately, the virus e-mail routine is not very effective relying on certain applications being active upon the user's desktop, and so is not likely to be spread rapidly via this mechanism.

A virus, by definition, replicates, and attaches itself to a host program. Although CHRISTMA EXEC did not attach itself to a host and therefore was not strictly speaking a virus, and ShareFun.A appears to be flawed in its design, these examples of malware provide a definite warning of things to come. Collecting virus samples, extracting signatures and distributing cures have traditionally been time-consuming tasks for the anti-virus researcher. The upgrade and updating processes have required frequent action on the part of users. As we have observed more and more viruses, some anti-virus vendors have developed automated methods to deal with the analysis of common viruses. This has helped cut the workload, but is still insufficient to deal with the virus problems of the future. In a time when viruses can spread worldwide in hours or even minutes, a day or two of waiting could render a company impotent. Even automation of the distribution of signature updates via techniques such as push-technology will not fully solve the response-time problem; for viruses which spread chiefly by computer-computer interaction, rather than human-computer interaction, the interactive and time-consuming element of isolation, capture, replication, and analysis is quite simply too slow. We believe that current levels of protection are not sufficient to defend well against an e-mail-aware virus. By the time such a virus could be isolated, sent to researchers, replicated, analyzed, a fix provided and that fix disseminated worldwide, the virus may well have already reached epidemic proportions.

In an attempt to address this problem, *IBM Research* has developed a biologically inspired anti-virus technique: a computer immune system that can automatically identify, analyze and remove the virus from the system [22]. The immune system provides for automated collection and analysis of viruses, but does not stop there. It prepares and distributes the immunization for the virus automatically. No human intervention is required in most cases. Simply put, the immune system monitors activity and filters it for virus-like behaviour. If it is determined that a known virus is present, it deals with the virus appropriately. However, if a known virus is not found, the system then automatically transmits a copy of the suspected infection (via a transaction center) to the *IBM Research Division* labs. There, with no danger to the user's machine, the system releases decoy programs, which seduce the virus into attacking. The decoys are examined for modification, and when such modification is found, viral signatures are extracted, and a repair algorithm is generated. This algorithm is automatically distributed throughout the system, curing both the virus which has been found there and on any other machines which have enabled the immune system. At the same time, immunity to that virus is provided throughout the system. All of this can take place in a matter of minutes, making use of secured authenticated transactions between the users PC and the *IBM Research*

Division secure lab. Although human input may still be required in some rare situations, it is hoped that the ability of the immune system to respond to new threats will far exceed conventional techniques.

Immune System Overview

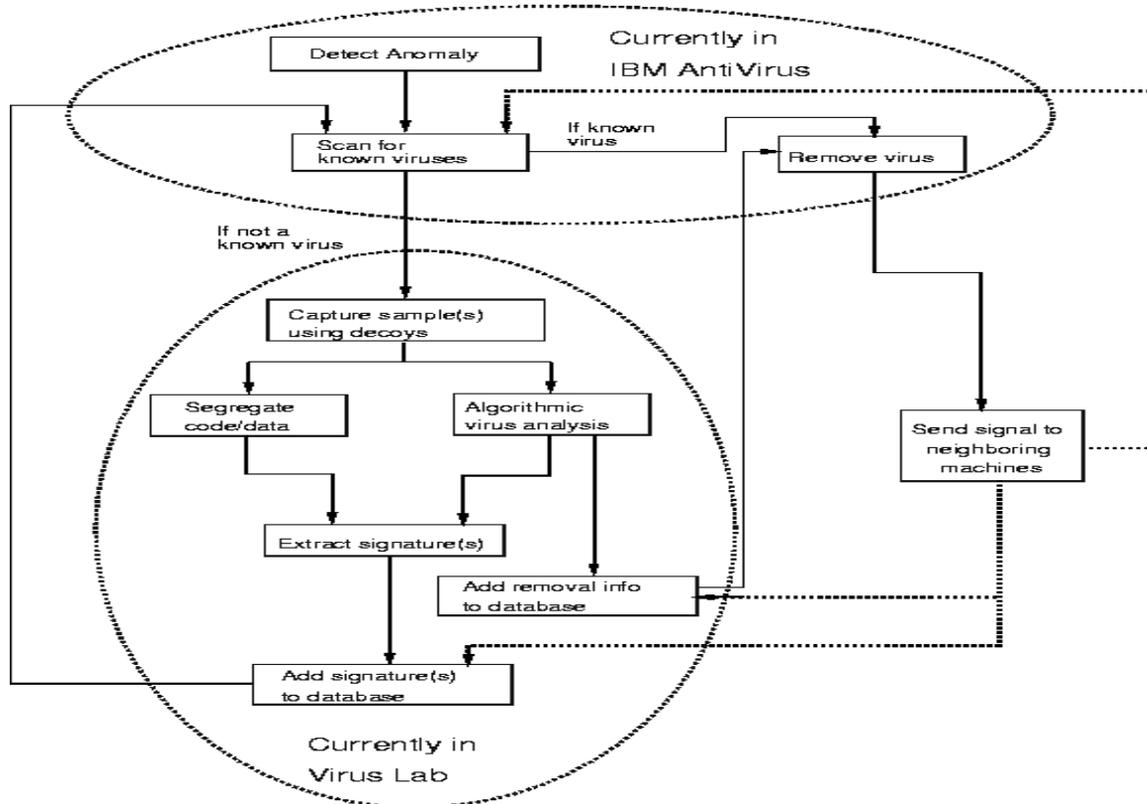


Figure 1. [26]

Certification Challenges

This immune system model offers obvious benefits to the user and administrator: detection, removal, product updating and product distribution to multiple sites are all done immediately and transparently. The benefits to the developer include freeing of time for the anti-virus researcher who no longer will spend time analyzing trivial viruses, and implementing their detection. However, this protection model offers both technical and administrative challenges to the certification body. The challenges are many; here, we will examine several of them.

Design Criteria and Test Administration

Currently, central to the test criteria for all anti-virus product certifications are various lists of viruses with obvious names. Indeed, as reflected in the *ITSEC* guidelines, naming compliance and consistency are sometimes important parts of the product [23]. The immune system model eliminates the need for developers to decide on names before detection and disinfection for new viruses can be implemented. *The WildList*, being name-based, would be unsuitable for use as a minimal detection criteria of these new, rapidly spreading, in the wild (but as yet unnamed) viruses.

The measurement of response times as currently designed into certification schemes such as those used by *NCSA*, and *SECURE COMPUTING Checkmark* is made to measure for protection models in which immediate response is unnecessary or unfeasible. An immune system model can meet the needs of the users in situations where Internet connectivity and viral increase will cause a two or three month time lag in documenting certification measurement

to be clearly unacceptable; the current certification model cannot effectively measure this response time. Automated updates, push technology enabled updates, and updates available via File Transfer Protocol sites and Bulletin Board Systems may be documented in the cooperative *CLEF/Developer* effort that is part of the *ITSEC Anti-Virus Working Group* model. However, while this current certification model has the potential to provide some measurement of vendor response time and reliability using these, in the future when new viruses spread worldwide in a matter of days, hours, or even minutes, response time problems will render even these approaches too slow. Indeed, we believe some of these approaches are already outdated.

The testing of an immune system model will require a high degree of competency and technical expertise with not only anti-virus software and virus sample replication, but with the Internet and networked systems in general.

Conclusions

We have looked briefly at the history of the term "In the Wild" and how this developed into *The WildList*, the *de facto* standard for building test-sets made up of those viruses in the wild. We have then examined three certification schemes based upon *The WildList*, and show that only one, *ITSEC*, appears to be constructed in such a way as to measure the ability of a vendor to track and match the current threat: the others are chiefly based upon *The WildList*, and suffer greatly due to the rapidly changing threat. In one case, we illustrated how a certified product might not even be able to detect viruses in the wild which were spreading 6 months prior to the current date.

We considered an alternate way of classifying viruses for certification purposes, and discovered that although the number of viruses is rising steadily, the actual threat posed by computer viruses to computers varies as a function of time. We have highlighted the importance of measuring a developer's ability to quickly respond to new viruses and supply updates in the field. In particular, we note that in the case of an e-mail-aware or Internet-aware virus, even automated signature distribution may be too slow to be of much practical help. The computer-computer interactions which are becoming more and more the models of the ways in which we conduct business on the Internet are rendering manual elements of viral isolation, sample capture, replication, and analysis too slow - only techniques such as *IBM's* immune system approach offer the type of response time needed to adequately protect from such a virus.

This has serious implications for those involved in the certification of anti-virus software. Tests based upon *The WildList* measure the ability of a product to protect the user far better than Zoo based tests. However, we question the long-term usefulness of *WildList*-based certification schemes, especially in light of the turnaround and maintenance time of certification. While we acknowledge *The WildList* to be much improved with definite scientific and practical value, we feel certifications based upon *The WildList* represent the bare minimum in terms of protection - their presence alone is insufficient to guarantee the protection of your company.

Acknowledgments

The author would like to thank Richard A. Ford and Steve R. White, IBM TJ Watson Research Center, for suggestions, corrections and general enlightenment.

Bibliography

1. Real World Anti-Virus Reviews and Evaluations-the Current State of Affairs. Sarah Gordon and Richard Ford. Proceedings 19th National Information Systems Security Conference. Baltimore Maryland. October 1996. pp. 526-535
2. David Chess. Private e-mail conversation. Used with permission.
3. Alan Solomon. Private e-mail conversation. Used with permission.
4. Roger Riordan. Private e-mail conversation. Used with permission.
5. Letters to the Editor. Virus Bulletin. July 1991
6. Measuring Computer Virus Prevalence. Jeffrey O. Kephart and Steve R. White. Proceedings of the Second International Virus Bulletin Conference. Edinburgh, Scotland, September 2-3, 1992, pp. 9-28.
7. In [1]
8. *The WildList*. Joe Wells. <http://www.av.ibm.com>
9. In [6]

10. Real World Anti-virus Reviews and Evaluations - the Current State of Affairs. Presentation. Sarah Gordon. 19th National Information Systems Security Conference. National Institute of Standards and Technology, National Computer Security Center. Baltimore Maryland. October 1996.
11. In [1]
12. A Perspective of Evaluation in the UK versus the US. Alan Borrett. Proceedings 18th National Information Systems Security Conference. Baltimore, Maryland. 1995. pp.
13. What A Winword Concept. Sarah Gordon. Virus Bulletin. September 1995.
14. Excel Yourself. Sarah Gordon. Virus Bulletin. September 1996.
15. In [8]
16. In [11]
17. The Administrators' Guide to Macro Viruses. International Virus Prevention Conference. Presentation. Richard Ford. Arlington, Virginia. 1997.
18. Virus Exchange BBS: A Legal Crime? Sarah Gordon. American Association for the Advancement of Science. Conference on Computer and Network Use and Abuse. Irvine, California. 1993.
19. Technologically Enabled Crime: Shifting Paradigms for the Year 2000. Sarah Gordon. Computers and Security. October 1995. pp391-402.
20. Computer Viruses: A Brief Overview. Carrie France. August 1996.
<http://www.academic.marist.edu/papers/france/paper.htm>
21. David Chess. Private Communication. Used with Permission.
22. Biologically Inspired Defenses Against Computer Viruses. Jeffrey Kephart, Gregory Sorkin, William Arnold, David Chess, Gerald Tesauro and Steve White. Proc. Int'l J. Conf. on AI (IJCAI-95), Morgan Kaufmann, San Francisco, 1995, pp. 985-996.
23. ITSEC Anti-Virus Working Group. A-V Product Standard Functionality Class F-AVIR. Draft Issue 10. March 1997.
24. Ethical Implications and Impacts of Anti-virus Research. Sarah Gordon. In Progress.
25. A Biologically Inspired Immune System for Computers. Jeffrey O. Kephart . High Integrity Computing Laboratory . IBM Thomas J. Watson Research Center. Published in Artificial Life IV. Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems. MIT Press. Cambridge, Massachusetts. 1994. pp. 130-139.

What is Wild?

Sarah Gordon
IBM TJ Watson Research Center

P.O. Box 704
Yorktown Heights, NY 10598

Prepared for the 20th National Information Systems
Security Conference