

What's in a Name?

By Sarah Gordon, Senior Research Fellow Symantec Security Response

Originally published in SC Magazine, June 2002 SC On-Line SC Magazine www.scmagazine.com



In this article, issues of virus naming as they impact users will be discussed, and some of the ways that naming might affect the reviewing of products will be examined.

Contents

Introduction	3
Practical Problems	4
Implications for Tests and the Users of Tests	6
Conclusions	7

> Introduction

MONTAGUES, CAPULETS AND OTHER NAMING ISSUES...

At first sight, the issue of virus naming seems to be a moot point. After all, one would expect that every computer virus would have a "scientific" name that can be applied to it, just as every plant has a "correct" name. However, it is at this point that the problem rears its head: the traditional taxonomy of biological things relies upon samples.

One of the nicest explanations of proper scientific classification is given on the WildList Organization site (http://www.wildlist.org/naming.htm) and the interested reader is referred there for a more in depth discussion.

In summary, biologists use a sample-based naming scheme, so that a new plant is ultimately identified by comparing it to reference samples of known plants. Thus, in every case, it is theoretically possible to go back to the "reference," allowing a scientist to formally identify that the sample he or she is looking at really is a Primula veris, not something else entirely. Correctly implemented, the system is very thorough and extremely reliable.

The problem with applying this approach in the anti-virus world has been the lack of a reference collection or even a central naming body. Whereas new species of plant are not discovered on a daily basis (nor is there an urgent need to name such new plants before they propagate!), many viruses are discovered every day, and in some cases, information about these viruses needs to be gotten to the user community. This requires that the virus be called something in the interim. And, as the virus may be discovered at slightly different times by different vendors, until they can compare the sample, they can't coordinate the name.

Given each vendor rightly wants to get the solution to the customer as quickly as possible, waiting around while "the right name" is discussed seems rather counterproductive. At the same time, users can become confused by the slightly different names for the same virus listed on various vendor sites. What is to be done?

The lack of a central reference collection that is accessible to all anti-virus product developers has been a huge barrier to establishing a scientifically valid taxonomy of computer viruses. While loosely organized member-only based groups like CARO have done a great deal of valuable work in the area of structuring virus naming, their very nature disqualifies them for maintenance and administration of any real scientifically based naming scheme.

While the total number of viruses is huge, the vast majority of viruses have never been seen "in the wild"—that is, actually spreading on users' computers rather than simply existing in a laboratory somewhere. Therefore rather than approach the general problem, it is more approachable to deal with the most pressing problem, which is those viruses which are actually causing a real-world problem. To this end, the WildList Organization took two important steps which should help with naming issues, at least in the crucial in the- wild test sets.

Because all vendors do not discover a given virus at exactly the same time, but some may discover it roughly concurrently, it is inevitable that different virus names will be used at least initially. Thus the WildList alias function has provided a much needed, and unbiased reference of virus names/aliases. This can be used when your product detects a virus to help you "translate" between one product's naming scheme and another. For example, the virus J&M below is called by at least one other vendor "Jimi", and "Hasita" by another.

Name of Virus	[Alias(es)]	List Date	Reported by
AntiCMOS.A	[Lenart]	1/95	OzSkSmWs
AntiEXE.A	[D3	9/94	EwOzSmWsZz
Empire.Monkey.B	[Monkey 2]	7/94	WsZz
Form.A	[Form 18]	7/94	CsEwPbSkSmWsZz
J&M.A	[Jimi,Hasita]	6/95	MtPb

Second, the WildList Organization created the WildCore reference sample set which has been made available to anti-virus product developers. This set effectively became the set of "reference" samples that has to this point been missing.

In a perfect world, all vendors would use this type of set to create scanners which all use exactly the same name for the same virus. However, there is no way to force vendors to use identical names, even if (as we shall see is not always the case) we would always want to. Therefore, this sort of sample set allows for easy compilation of virus' aliases, as well as a method for implementing a more coordinated naming scheme if and when appropriate.

> Practical Problems

Aside from the preceding scientific issues regarding exact naming, there are several practical reasons why exact identification of a virus sample is not always beneficial.

One of the most important reasons for not carrying out exact identification is illustrated by the Pareto Principal, which essentially talks about the diminishing rewards from increased effort. Thus, there is an overhead in terms of analysis and code required to perfectly identify a virus, but the return to the user for this extra information is limited.

Consider the case of a class of viruses (for the sake of this article, we will discuss the fictitious class WM.Theora) which are all slightly different, but do not differ in terms of behavior. Should we require each of these viruses to be exactly identified (e.g. WM.Theora.A, WM.Theora.B...), or is it sufficient to do this: WM.Theora.Generic, or even this: WM.Harmless? The answer to this question is a question of perspective.

For a scientist interested in building a formal taxonomy of samples, clearly the exact identification is better. However, from a user perspective (and it is, ultimately, the user who pays for virus analysis and classification), the primary interest is in detection and removal of the virus; within reason, the naming is irrelevant! Thus, one can argue that the perfect naming of viruses is a "nice to have" from a theoretical perspective, but not necessarily really an advantage for users.

Aside from the small returns on the extra work, there is also a speed overhead involved in exactly identifying a virus, as well as larger update files and memory footprints. Thus, the cost of perfect identification is not zero— and this cost is likely to rise as the number of viruses and worms continues to skyrocket.

Another issue faced by product manufacturers is that of response time. When a new, network-aware virus is discovered, speed of response is often critical. The recent W32.Nimda.A@mm outbreak only goes to show how quickly a new virus can spread in a largely inter-networked environment. Thus, drivers are often released for a new virus before the industry has come to anything even approaching consensus for its name.

This is quite as it should be, as the more important issue is protection, not naming. While later revisions of the product may well standardize a name that is given post facto to the virus, it is certainly not reasonable to expect a new driver to necessarily have the "right" name.

Moreover, as we see more and more blended threats, it is becoming critically important that information about the vulnerability exploited by such a threat is readily available, and able to be easily referenced. The common vulnerabilities and exposures (CVE) dictionary (www.cve.mitre.org) provides an excellent resource for this type of cross reference. Whether or not vendors may adapt this reference information into their products, or into virus analysis remains to be seen.

In any case, whether or not a vendor should go back at some date and reconcile the name to the name given on The WildList, or used by the majority of vendors, or to coincide with a CVE entry, is debatable. Professor Klaus Brunnstein, from the University of Hamburg's Virus Test Center states: "Renaming seems adequate (and has often been done) if some critter is just 'in the zoo' but if a virus or worm has become widespread, then the only mechanism is to name/update known ALIASES on the website. This implies that AV websites are maintained also in more professional manner, which must be further developed. The best solution (even with its inherent risks) is to work in a professional manner, rather than search for impossible (even not academic) solutions.

Costs for such efforts are inevitably going to be passed to the user. Given the alias information provided throughout the anti-virus industry—providing a ready method to translate between products—the wisdom of investing much time and effort to exactly match a name is questioned by some.

> Implications for Tests and the Users of Tests

None of the preceding discussion should be taken as an effort to minimize the importance of identifying those viruses found in the wild. Indeed, as discussed, the WildList Organization has gone to great lengths to model a framework by which it is possible to scientifically establish the correspondence of one sample to another (a critical link in the naming issue—the creation of a reference collection).

However, it is important to place the relative importance of compliance with the naming schemes in their proper perspective.

As testers look for ways to improve their testing protocols, one of the most obvious (and easiest) pieces of analysis which can be carried out is comparing the given name of a virus with that reported by the product which is being tested. The problem with this is that the focus is moved off the most important part of the problem (preventing and repairing infection) and shifted to an issue that is cloudy at best.

One of the enduring debates in the anti-virus community is the perennial debate over the "right" name for a computer virus. While sparking a great deal of academic interest, this debate also has the potential to impact users more directly when they attempt to interpret reviews of anti-virus products.

For example, should a product be penalized for using the "wrong" name for a virus? Should we expect a product to use the "exact" name, and what precisely does this mean anyway?

One of the most important maxims which testers should consider when deciding how to approach expanding tests is to test the product in the way in which it is most critical to the user, not to test the product in ways which are easy, or ways which produce mounds of scientific-looking charts and forms. Reliability of disinfection and detection are clearly paramount, as is ease and (most recently) speed of updating. Improvements in any of these areas are far more useful for users than somewhat esoteric naming issues.

When examining an independent test, it is therefore important to understand clearly how products detection rates and reliability have been calculated. If overall scores take into account naming "errors," the test should be reinterpreted in this light. Furthermore, it is probably beneficial in the longer term to contact the tester, and let them know if the accuracy of naming results should even be included in overall scores.

> Conclusions

Vendors clearly need guidelines to enable them to name viruses using a standard method, with open, publishable guidelines that are accessible by everyone. It remains unavoidable that names provided by different vendors will be different, at least at first.

With the advent of blended threats, a database of vulnerabilities can greatly add in understanding the role of the vulnerability in the virus attack. At the end of the day, the reference number used by a vendor to identify a vulnerability doesn't have to be a sexy designator designed for media appeal; the CVE database provides this type of information in a readily accessible form.

After all, the most important feature of any anti-virus product is how well it reduces the overall cost of the computer virus problem. The best product for your company is one that, in the final analysis, has the lowest total cost of ownership. How much the cost of ownership is reduced by "exact" naming is not clear, but it would seem that it is possible that the cost may actually be raised by such a feature.

As we have seen, there are many factors to examine when we consider how we should evaluate how "well" products name viruses—and even how (if at all!) such an evaluation should be used to decide how effective a product is at protecting the enterprise from computer viruses.

Perhaps Shakespeare had it right, when he penned the following:

What's in a name? That which we call a rose

By any other name would smell as sweet...

SYMANTEC, THE WORLD LEADER IN INTERNET SECURITY TECHNOLOGY, PROVIDES A BROAD RANGE OF CONTENT AND NETWORK SECURITY SOFTWARE AND APPLIANCE SOLUTIONS TO INDIVIDUALS, ENTERPRISES AND SERVICE PROVIDERS. THE COMPANY IS A LEADING PROVIDER OF VIRUS PROTECTION, FIREWALL AND VIRTUAL PRIVATE NETWORK, VULNERABILITY ASSESSMENT, INTRUSION PREVENTION, INTERNET CONTENT AND EMAIL FILTERING, AND REMOTE MANAGEMENT TECHNOLOGIES AND SECURITY SERVICES TO ENTERPRISES AND SERVICE PROVIDERS AROUND THE WORLD. SYMANTEC'S NORTON BRAND OF CONSUMER SECURITY PRODUCTS IS A LEADER IN WORLDWIDE RETAIL SALES AND INDUSTRY AWARDS. HEADQUARTERED IN CUPERTINO, CALIF., SYMANTEC HAS WORLDWIDE OPERATIONS IN 38 COUNTRIES.

FOR MORE INFORMATION, PLEASE VISIT WWW.SYMANTEC.COM

WORLD HEADQUARTERS

20330 Stevens Creek Blvd. Cupertino, CA 95014 U.S.A. 408.517.8000 800.721.3934

www.symantec.com

For Product Information In the U.S., call toll-free 800.745.6054.

Symantec has worldwide operations in 38 countries. For specific country offices and contact numbers please visit our Web site.

The information in this document is subject to change without notice and must not be construed as a commitment on the part of Symantec Corporation. Symantec assumes no responsibility for any errors that may appear in this document. Symantec and the Symantec logo are U.S. registered trademarks of Symantec Corporation. Other brands and products are trademarks of their respective holders. Copyright © 2003 Symantec Corporation. All rights reserved. Printed in the U.S. A 02/03