

A Virus by Any Other Name: Virus Naming Practices

Costin Riau 2002-06-03

A Virus by Any Other Name: Virus Naming Practices

by *Costin Riau, Kaspersky Labs Romania*

last updated June 3, 2002

Introduction: a Little History, a Lot of Questions

When the "VBS/VBSWG.J" virus appeared, the media decided to call it by a more appealing name, "AnnaKournikova", which was derived from the JPEG file that the virus claimed to be. However, none of the anti-virus products included in the excellent virus names cross reference tool [VGrep](#) currently lists this virus as "AnnaKournikova", "Kournikova", or any other variation based on the name of the charismatic tennis player. On the other hand, a considerable number of AV programs detect it as "SST", while a very small number don't call it "VBSWG" or "SST".

One virus, so many names. Why? Why didn't the media call it "SST" in the first place? Why aren't all AV products detecting it as "VBSWG"? And why aren't all the AV products at least referring to it by a single name. These are all questions this article intends to cover.

NVNC '91: the 1991 New Virus Naming Convention

Before discussing why a relatively trivial virus such as "VBSWG.J" is known by at least five different names, it may help to discuss current standards for naming conventions. Eleven years ago, a group of security experts known as the Computer AntiVirus Researcher Organization (CARO), first attempted to develop a standard virus naming scheme in the form of the [1991 New Virus Naming Convention](#) (NVNC '91). The '91 NVNC promoted the now commonly used 'Family_Name.Group_Name.Variant' form, through the following more complex formulae:

Family_Name.Group_Name.Major_Variant.Minor_Variant[:Modifier].

Besides that, it also covered some other important aspects such as the role of "family names" in virus names, what is and is not suitable for inclusion in the virus name and the usage of various modifiers in order to specify common run-time packing information or polymorphic engines linked into the viral code. As an example of the type of issues that the original document attempted to address, part of the guidelines regarding virus naming included the following prohibition:

1. Do not use company names, brand names, or names of living people, except where the virus is provably written by the person. Common first names are permissible, but be careful - avoid if possible. In particular, avoid names associated with the anti-virus world. If a virus claims to be written by a particular person or company do not believe it without further proof. (NVNC '91, Sec. 1.1)

While this stipulation may seem to be commonsensical, its inclusion in the Naming Convention was intended to prevent possible legal issues by associating someone's name or image with a piece of malware code. By extension, it also serves as an indication that even if the media doesn't seem to care much about virus naming conventions, the AV community does. It may also cast some light on why AV producers have not followed the media lead in naming viruses after charismatic tennis players. (Unfortunately, while the AV community may be bound by industry-developed convention, there is nothing to hold the media to the same standards, a sad reality that promises more confusion around malware nomenclature in the future.)

The Need to Update NVNC '91

Unfortunately, since NVNC '91 was released, a lot of things have changed. New types of malware have appeared, macro viruses, Windows viruses or mass mailers being just a few examples. The very same people who designed the NVNC '91, have been working on an updated naming convention for almost two years. The new convention would need to incorporate new criteria that simply were not relevant in 1991, into the so-called "CARO name".

For instance, current anti-virus products detect much more than just viruses. Backdoors, trojans, joke programs, droppers are just a few examples of the malicious code which require special naming prefixes, but which weren't covered by the NVNC '91. Thus, the improvements to the NVNC '91 being discussed currently would include a "malware class" part that indicates the type of malware in question. For example, "trojan://" in the following name: trojan://X97M/Johar.A.

In addition to the types of malware being developed, the platforms on which infection can occur has also diversified since the development of the initial NVNC. As a result, the inclusion of a "virus platform" indicator must be another component of the new naming convention. In the example cited above, the virus platform is indicated by the "X97M" substring. In short, this indicates that the trojan works in Excel97, and is written in VBA5 or 6, thus providing additional

useful information to the user. Currently many other platform strings exist, and more are likely to appear in the future. This is one reason why a naming convention will never be complete, but will continue to be a work in progress, one that will require constant updates as new virus types appear.

Alternate Approaches to Naming Conventions

While this discussion has thus far privileged NVNC '91 as the primary standard for virus naming, it's interesting, and important, to note that since the release of the NVNC '91, some other efforts towards improving, updating or creating new naming conventions have been made. For instance, [one such attempt, hereafter referred to as GSNC '99](#), was imagined by Gerald Scheidl, then of Ikarus Software, who took some steps towards the inclusion of the "virus platform" and "malware class" (the latter is referred to as "type" in the GSNC '99) into the virus name syntax. Overall, GSNC maintains the compatibility with the initial CARO convention, improving it to meet the new changes that were recorded in the AV field for the past years, but still following the initial recommendations of the NVNC '91 documents towards the selection of new virus names.

A different approach was described by Frank W. Felzmann, Klaus-Dieter Moeller and Guenter Musstopf in a paper called "Naming of viruses, worms, trojans and related malware.", which promotes the idea of simplifying the virus name as much as possible with the purpose of reducing confusion between the users. This second proposal was discussed between a number of AV researchers at the Virus Bulletin 2001 Conference in Prague; however, it didn't garner the support required for adoption by the majority or the AV developers.

One last thing that should be noted regarding the NVNC '91 is that it didn't make any attempts to unify the virus family names between all the anti-virus products, a job which in the early days of the AV community used to be done directly between the various AV researchers participating in CARO. However, an attempt to unify the names of macro and script viruses is currently underway. This is known as VMacro and is based on VTC Hamburg's lists of known [macro](#) and [script](#) viruses. Together, these two lists have provided a common knowledge base and the tools needed for the classification of new viruses since 1996. Unfortunately, no similar efforts exist for other types of viruses, but more about that later.

A Virus By Any Other Name – Why Naming Conventions are Important

Although this article has discussed the evolution of naming conventions, it still may not be clear to the reader why such a convention is necessary. After all, does it really matter if the AV community refers to AnnaKournikova by the same name as the media?

[Kaspersky Labs](#) receives hundreds, sometimes up to a couple of thousand, new viruses each month. Some of these are new versions of known viruses, some are just using code from already known viruses, and some don't bear any similarity to any known viruses. It is difficult for virus analysts to differentiate one group from another. Fortunately, tools exist that can help with the identification of similar variants, especially for macro and script viruses, thus simplifying the naming process and reducing redundancy and confusion. However, even after weeding out the new versions of known viruses, a lot of completely new viruses remain, all of which need to receive new names. As a result, the total number of virus names reported by most AV increases every month at various rates, depending on how generic the detection algorithms of the products are and how many viruses are received and processed by the lab.

In the course of researching this article, I made a small attempt at counting the total number of virus names from the most known antivirus products. The main problem of this was of course the fact that not all AV products provide users with a list of known viruses, and some products, even if they can show the list to the user, do not offer the possibility of saving it to an external file. In the end, I managed to collect a list of roughly 400,000 virus names from 14 different anti-virus vendors. Parsing the virus names, removing the platform and malware classes in all their forms, I've come up with a list of around 300,000 unique names. On average, that means that around 7000 virus names are common between all the 14 tested AV producers. Looked at in a different way, this means that roughly one in four virus names is shared between the majority of AV products. That doesn't necessarily mean exactly the same virus is called the same way by all the products, but it provides a good estimate of the average number of names a unique virus has: four. Now, back to our initial example, counting the different names available for the "VBSWG.J", we get the same number, four. Counting the name developed by the media, six. Obviously, having an industry-wide standard for the nomenclature of the vast and increasing numbers of malicious code can only help in the analysis, categorization, detection and subsequent management of malware, both in the zoo and in the wild.

Where Do the Names Come From?

So, now that we have established the current naming conventions, as well as the need for those

conventions, we can discuss the ways in which some names are currently derived. Some guidelines for choosing virus names are available along with the NVNC '91. Besides the guideline mentioned previously in this article, another interesting one is to avoid naming the virus based on its payload. It is relatively tempting to want to name malicious code based on its date of activation, this can create confusing duplication of names. For instance, if we were to name every new virus with some word derived from its payload, like "March6", "January Friday 13th" or "CrashWindows" the fictional exchange illustrated below could become commonplace:

(A1 - Analyst1, works for the respectable AV company C1)

(A2 - Analyst2, works for the most respectable AV company C2)

(A3 - Analyst3, works for the (even more) respectable AV company C3)

A1: "Hey A2, have you seen that new beast, the 'Newyork' virus?"

A2: "You mean the one which fills all the files on disk with 'New York'?"

A1: "No, that's the 'NYFiller' virus, I mean the one which shows a message box with the text 'New York New York'"

A2: "Could be, I remember having seen two of them, one was a macro virus and the other one infecting Linux ELF files"

A1: "Hm, the 'Newyork' I was thinking of actually infects Windows PE files"

A2: "Ah, but I think I know what you mean, however, the one I've seen shows a message box stating 'New Orleans New Orleans'. We are calling it 'NewOrleans', of course."

A1: "Hm, that must be a new version of our 'NewYork' virus with a modified message. I think you should rename your 'NewOrleans' virus to something like 'NewYork(version:Orleans)'."

A2: "Hey, wait a minute, why not rename `_your_` virus to 'NewOrleans(York)'?"

A3: "Hey guys, have you seen the new virus which fills all the files on disk with 'New Delhi'? We're calling it 'NewDelhi', of course."

A1: "Arghhh..."

A2: "Who designed this stupid payload-based naming scheme anyway...?"

While the preceding is a fictional example, it gives a brief illustration of why it's better not to use parts of the payload of a virus in its name. So, if not according to the payload, how else should malware be named? Eliminate the names of real living persons (as explained earlier), well-known diseases, software products, companies and the list of potential, relevant names shrinks quickly. However, there are ways around these pragmatic prohibitions. For instance, one common practice which became more and more popular in recent years is the reversal of some obvious name derived from the virus or various transpositions of it. Examples include "Arual" (reverse of "Laura"), "Golni" ("Winlog" reversed and truncated) or the very well known

"Nimda" ("Admin" reversed).

Experience also shows that it may be tempting to go to the other extreme, to choose names that are hard to pronounce, let alone remember. This was evident in the initial example that we discussed in this article. The preceding statement explains the main reason that some AV vendors refer to the virus as "SST" and not "VBSWG". It was argued that "VBSWG" is too hard to remember (and pronounce) for the average anti-virus user; thus, some products went with the shorter "SST" variant name. Of course, the developers of other AV producers argued that "VBSWG" is not that hard to remember for their users, and decided to keep it.

Now for the Good News...

The good news is that increasing cooperation and coordination between AV developers is starting to show in the unanimity of names of massively distributed malware, such as "Klez", "Magistr" or "SirCam". For these three viruses, a general agreement towards their names exists, so cases similar to "VBSWG.J" hopefully have less chances of occurring again in the future. More encouragingly, the media has also used these industry-supported names. This goes a long way to reducing confusion amongst computer users. On the other hand, viruses are still named by humans, and humans are bound to make errors. That's why cooperation should be backed up by methods to correct the errors, and above all, the understanding that the standardization of virus names will not only be helpful to the users, but to the whole AV industry as well.

Costin Raiu, has researched and developed anti-virus programs since 1992. Currently, Costin is working for [Kaspersky Labs](#), in the AVP4 (Prague) development team. Costin has written for publications like [VirusBulletin](#) and presented at many computer security conferences. He is a member of the [WildList Organization International](#) and a participant in the [Computer AntiVirus Research Organization](#).

Relevant Links

[A New Virus Naming Convention - \(NVNC 1991\)](#)

[Kournikova virus smashes the Net](#)

ZDNet News

[Virus Naming Convention 1999 \(GSNC '99\)](#)

Gerald Scheidl

[List of Known Macro Viruses](#)

Virus Test Centre

[List of Known Script Viruses](#)

Virus Test Centre

[Privacy Statement](#)

Copyright 2006, SecurityFocus