Virus Naming.
The “Who’s who?” Dilemma

WHITEPAPER

bitdefender

Disclaimer

The information and data asserted in this document represent the current opinion of BitDefender® on the topics addressed as of the date of publication. This document and the information contained herein should not be interpreted in any way as a BitDefender’s commitment or agreement of any kind.

Although every precaution has been taken in the preparation of this document, the publisher, authors and contributors assume no responsibility for errors and/or omissions. Nor is any liability assumed for damages resulting from the use of the information contained herein. In addition, the information in this document is subject to change without prior notice. BitDefender, the publisher, authors and contributors cannot guarantee further related document issuance or any possible post -release information.

This document and the data contained herein are for information purposes only. BitDefender, the publisher, authors and contributors make no warranties, express, implied, or statutory, as to the information stated in this document.

The document content may not be suitable for every situation. If professional assistance is required, the services of a competent professional person should be sought. Neither BitDefender, the document publishers, authors nor the contributors shall be liable for damages arising here from.

The fact that an individual or organization, an individual or collective work, including printed materials, electronic documents, websites, etc., are referred in this document as a citation and/or source of current or further information does not imply that BitDefender, the document publisher, authors or contributors endorses the information or recommendations the individual, organization, independent or collective work, including printed materials, electronic documents, websites, etc. may provide. Readers should also be aware that BitDefender, the document publisher, authors or contributors cannot guarantee the accuracy of any information presented herein after the date of publication, including, but not limited to World Wide Web addresses and Internet links listed in this document which may have changed or disappeared between the time this work was written and released and the moment it is read.

The readers are entirely responsible to comply with all applicable international copyright laws arising from this document. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of BitDefender.

BitDefender may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from BitDefender, this document does not provide any license to these patents, trademarks, copyrights, or other intellectual property.

All other product and company names mentioned herein are for identification purposes only and are the property of, and may be trademarks of, their respective owners.

Authors

Sabina DATCU, Communication Specialist

Ioana JELEA, Communication Specialist

Table of Contents

Got new malware. What should we call it?

Anyone who has ever created something new is granted the right to baptize it. However, given that they are born under the sign of destruction and disruption, viruses are an exception to this rule.

Normally, you would not expect anything in the “John jr.” vein. Any hint as to the identity of virus creators would probably get them into trouble. Plus, in order to avoid adding to the glory of malware authors antimalware producers will probably re-name the malware samples they discover. And the naming trouble does not stop here. A scenario where several antimalware labs simultaneously conduct research on the same new malware sample is not that uncommon. In this case, the first to publicly announce the discovery gets to give it a name.

Aside from creativity and authorship, virus naming also raises the issue of utility. Confronted with an overwhelming malware population, researchers and antimalware producers have understood how important it is to approach the naming process systematically. All in all, simple logic calls for malware names that contain information the industry can recognize: the affected platform, the virus family name and its spreading method.

This whitepaper aims to summarize the efforts that have been invested into creating a coherent, unanimously accepted and, most of all, efficient malware naming system as well as to briefly dwell on how these regulatory attempts are reflected in practice.

Naming Conventions. Unifying Attempts.

A. The Caro System

In a 1991 meeting of Computer AntiVirus Researcher Organization (CARO), a New Virus Naming Convention[1] was agreed upon and it was supposed to provide a means of avoiding the confusion generated by the lack of uniform regulations in the virus naming process. According to this document, a full virus name should have the following format:

Family_Name.Group_Name.Major_Variant.Minor_Variant[:Modifier]

[1] Fridrik Skulason, Vesselin Bontchev, 1991, A New Virus Naming Convention, http://vx.netlux.org/lib/asb01.html

Here is an example of a virus name that complies with this model:

Stoned. Michelangelo.A

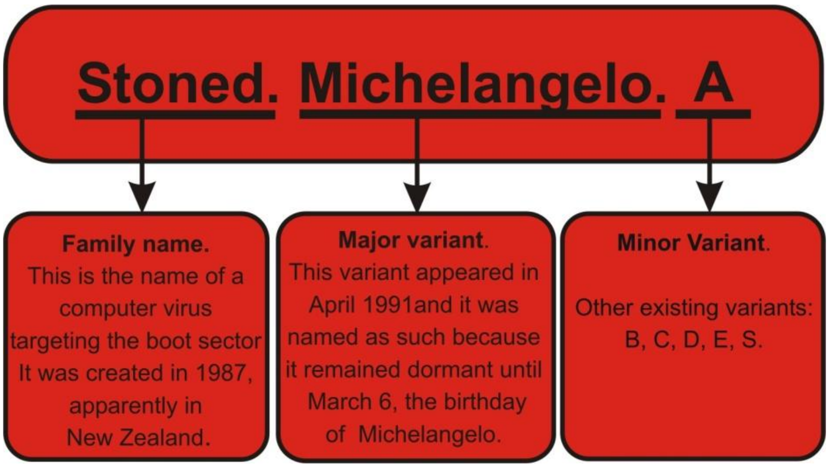
Stoned. Michelangelo. A
Family name.
This is the name of a computer virus targeting the boot sector It was created in 1987, apparently in New Zealand.
Major variant.
This variant appeared in April 1991and it was named as such because it remained dormant until March 6, the birthday of Michelangelo.
Minor Variant.
Other existing variants: B, C, D, E, S.

Figure 1: Virus Name based on the Caro Model (1991)

Though it appears to provide a clear solution to the naming problem, this format is likely to raise uniformity- related issues as well. A first grey area that the authors of the convention admit to is the “family name” section: “*Every attempt is made to group the existing viruses into families, depending on the structural similarities of the viruses, but we understand that a formal definition of a family is impossible*.”[2]

Starting from this inherent fallacy of the system, the authors provide a few guidelines on how to choose a relevant family name:

- the use of brand, company or individual’s names is forbidden (unless there is proof that the individual actually created the virus),
- existing virus family names should be considered carefully to avoid confusion (does the virus belong to that family? is the sample actually new or does it belong to an existing family?)
- dates, geographic and numeric names should be avoided because they can be misleading

The principles of agreed authorship and of utility are clearly stated as a viable solution: “*If multiple acceptable names exist, select the original one, the one used by the majority of existing anti-virus programs or the more descriptive one*.”[3]

An updated version[4] of these rules was created in 1999, as a private initiative, and it was offered as a suggestion to be adopted by the entire antivirus industry. This update was intended to accommodate into the CARO naming system malware types that affected other platforms than MS-DOS. As stated by the

[2] Ibid.

[3] Ibid.

[4] Gerald Scheidl, 1999, Virus Naming Convention 1999 (VNC99), http://members.chello.at/erikajo/vnc99b2.txt

author of the document, this change was triggered by the appearance of WM/Concept.A, the first macro virus to spread through Microsoft Word. Therefore, a proposal was made for the adoption of an extended form of the Caro standard: platform.type/caro-name [message].

In an attempt to further reflect the diversity of the malware population, the document also suggested considering the term “virus” as a default type and including other malware denominations in the Caro system: Trojan, dropper, worm, Joke, germ, etc.

Other elements intended to make the malware name as clearly descriptive as possible were the language identifiers and the short message that was supposed to clarify to the end user the malicious nature of the program.

Here is an example of a malware name that follows this model: **Win32.MSNWorm.Rachel.A**

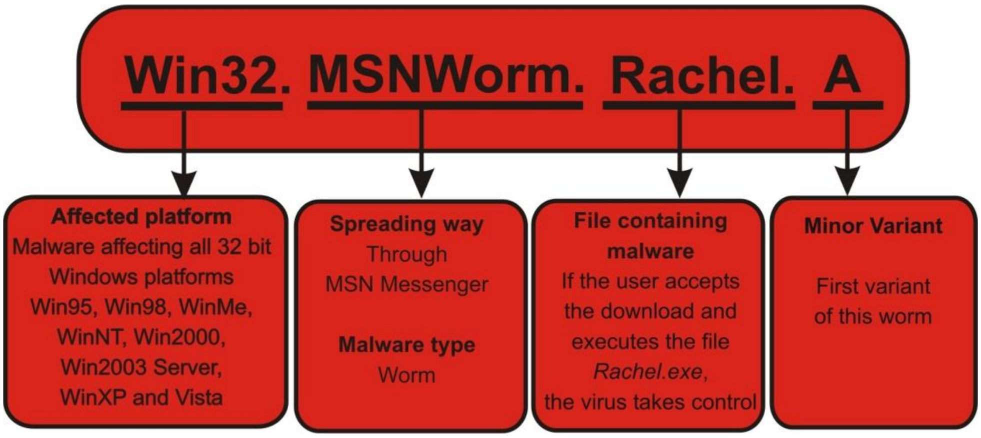
Win32. MSNWorm. Rachel. A
Affected platform
Malware affecting all 32 bit Windows platforms Win95, Win98, WinMe, WinNT, Win2000, Win2003 Server, WinXP and Vista
Spreading way
Through MSN Messenger
Malware type
Worm
File containing malware
If the user accepts the download and executes the file *Rachel.exe*, the virus takes control
Minor Variant
First variant of this worm

Figure 2: Virus name based on the updated Caro model (1999)

B. The Wildlist Approach

In his statement on How Scientific Naming Works[5], Joe Wells, CEO of Wildlist Organization International approaches the inconvenients of virus naming from a very practical point of view. In the absence of a scientific name giving system, such as in biology, and of a unified collection of virus samples that any researcher in this domain can access, a virus name should not be viewed as correct/wrong and all the existing names of a virus should be considered to be equally valid.

He points out an extremely important aspect that tends to be disregarded in this debate: the ultimate purpose is to warn the end-users of the threat, no matter what the name it is presented under. As the accuracy of virus identification (is it new? is it a variant of an existing one?, etc.) becomes the main focus, naming remains a secondary issue. To put it simply, any malware sample

[5] Joe Wells, 1999, How Scientific Naming Works, http://www.wildlist.org/naming.htm

should be identified by its Caro name, if not, by what the majority calls it, if not, by what the first person to discover it called it.

C. Towards a Common Malware Denomination

In 2005, during a Virus Bulletin Conference, a new attempt was made to bring order into the malware denomination system. This is when the CME initiative was born and brought together several major players in the data security industry that aimed “[…] *to provide a common name for high profile threats in the hope that customers will be able to protect their computers from malware attacks more effectively.”*[6]

The organizations that signed up to the CME agreed on a common malware identifier format, namely: CME- N, where N is an integer between 1 and 999. As illustrated by the CME list[7], one CME-N identifier corresponds to several aliases of the same malware sample. For instance CME-416 is the same as Trojan.Downloader.AOW (BitDefender), Email-Worm.Win32.Warezov.dc (Kaspersky), W32/Stration.dr (Mcafee), W32/Stratio-AW(Sophos), etc. In addition to that, in keeping with its encyclopedic aim, the list provides a description of the malware sample and the date of its activation.

Despite its capacity to bring more clarity into the matter of malware classification, some voices were skeptical about this system’s ability to keep up with the tremendous speed at which the antimalware industry works. The need to deliver a solution to counter each threat as soon as possible will most likely prevail over this new naming requirement, which will probably only be applied post factum. In other words, in the identification stage, there will be just as many malware sample aliases, but in the classification stage, there will be a way for several aliases to be reunited under a distinct CME-N identifier.

Although efforts have been made towards reaching a consensus on virus naming rules, diversity seems to hold the upper hand for the moment. Therefore, when trying to figure out the principles behind virus naming, sheer inspiration appears to be the answer.

Overview of Naming Trends

Besides the purely technical denomination it is given through conventions, a piece of malware can also have a more “familiar” name, which makes it more accessible to the general public. A non-technical audience will be capable of reading more into such a “nickname” rather than into the technical one. In this way, the public’s awareness of e-threats increases, which is consistent with the ultimate aim of the data security industry.

[6] Virus Bulletin Conference: Industry unveils unified naming for virus threats, 2005, http://www.pcpro.co.uk/news/security/78425/virus-bulletin-conference-industry-unveils-unified-naming-for-virus-threats

[7] http://cme.mitre.org/data/list.html

There is no single and unanimously accepted scientific point of view on the classification of the common names that viruses wear in time. In this whitepaper, the following criteria have been used for categorization purposes:

a) author’s name,
b) spreading method,
c) baits
d) messages in the code,
e) date when the virus becomes active
f) technical aspects
g) personal touch

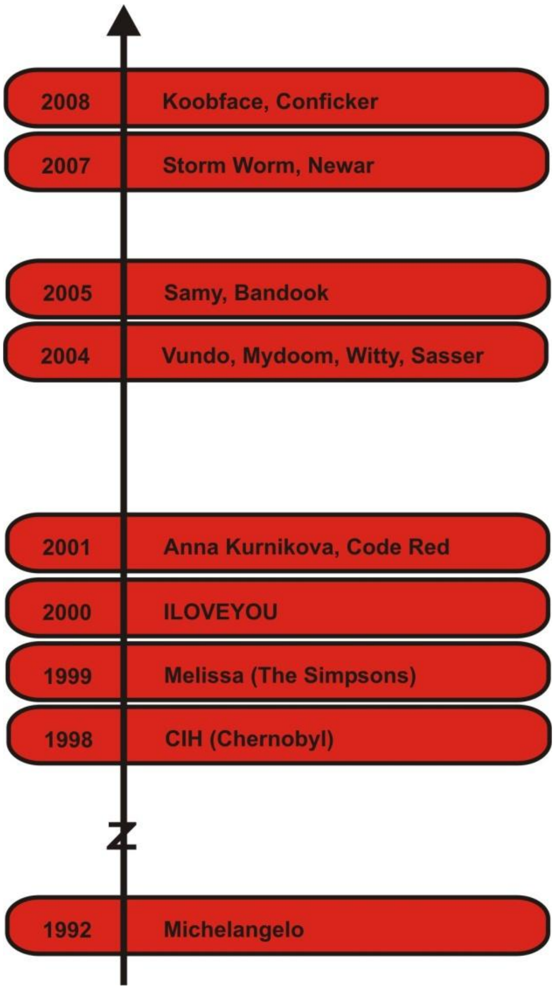

Figure 3: Analyzed virus names vs time frame

a) Author’s Name

Even though CARO’s *A New Virus Naming Convention* (1991) states that the creator’s name can be used when naming an e-threat, provided that authorship is proven, this cannot be considered common practice.

Because malware authors usually protect themselves and want to remain unknown, just a few viruses have been named after their “parents” in the history of e-threats.

The Samy worm and the CIH virus are two exceptional cases in which the names of malware are clearly connected to their creators’ identities. The Samy

Worm[8] was developed in 2005 by Samy Kamkar and it took over its father’s first name, while CIH[9], the virus that caused massive data losses beginning with 1998, got its name from its author’s initials: **C**hen **I**ng **H**au.

b) Spreading Method

There are lots of ways in which malware can spread, such as social networks, links placed on very popular web sites or using as hooks different personalities’ names, pictures of events, enticing messages, etc.

In this category, the Koobface worm appears as an interesting case because its name is the anagram of a very well known social network, the users of which it actually targets. Another example of the kind is the Trojan name Vundo, which is short for Virtual Mundo (Virtual world), a form of online community in which users interact which each other through their avatars in a virtual world.

c) Baits

Malware naming also finds a source of inspiration in the type of messages authors concoct in order to trick the victim into downloading the malicious code.

The messages used as baits may appeal to the potential victim’s curiosity. One such example is the promise to reveal famous persons’ pictures. The VBS.SST virus, for instance, was designed to trick e-mail users into opening an e-mail message purportedly containing a picture of tennis player Anna Kournikova, but which was actually carrying a malicious program. With a simple message - "Hi: Check This!"- the Kournikova virus[10], as it came to be called, tempted users into accessing what appeared to be a picture file labeled "AnnaKournikova.jpg.vbs". If set off, the malicious program behind it would plunder the user’s address book in Microsoft Outlook and it would attempt to send itself to all the contacts listed there.

Recent high-impact events or potential catastrophes can also arouse curiosity. The Storm Worm[11] began infecting systems from Europe and United States on Friday, January 19, 2007, using an e-mail message with a subject line about a recent weather disaster: "230 dead as storm batters Europe”. Similar tactics were used for Newar, a malicious code inserted in a message that referred to a new war, the Third World War.

The appeal to human emotions seems to work as well. The ILOVEYOU worm, which attacked computers in 2000, spread through an e-mail message with the text "ILOVEYOU" in the subject line and an attachment entitled "LOVE-LETTER-FOR-YOU.TXT.vbs". Once the attached file was opened, the worm it contained would send a copy of itself to everyone in the victim’s Windows Address Book, using the victim’s address in the sender line. It also made a number of malicious changes to the user's system.

[8] Justin Mann,, 2007, MySpace speaks about Samy Kamkar's sentencing, http://www.techspot.com/news/24226-myspace-speaks-about-samy-kamkars-sentencing.html

[9] http://www.economicexpert.com/a/Chen:Ing:Hau.html

[10] ***, 2001, Kournikova computer virus hits hard, http://news.bbc.co.uk/2/hi/science/nature/1167453.stm

[11] ***, 2007, Storm chaos prompts virus surge, http://news.bbc.co.uk/2/hi/technology/6278079.stm

d) Messages in the Code

It is also customary for some cybercriminals to insert phrases that will be displayed on the infected systems or that only represent their “signature” inside the code. As a consequence of that, some pieces of malware have been named after these “signatures”. Here are just a few examples:

- **Witty**[12]: the name of this worm comes from the phrase "(^.^) insert witty message here (^.^)" that appeared in the destructive payload it carried.
- **Mydoom**[13]: Craig Schmugar named it this way when he observed the text “my domain” in a line of the program's code. He shortened this phrase to “mydom” and then thought of making it more appealing by turning it into “mydoom”. He noted: *“It was evident early on that this would be very big. I thought having 'doom' in the name would be appropriate.”*
- **Santy**[14]: this worm got this name *because it caused writable files (in the .php and .html formats) on the infected server to display the message* "This site is defaced!!! This site is defaced!!! NeverEverNoSanity WebWorm generation X".

e) Date when the virus becomes active

Viruses do not necessarily become active immediately after the system has been infected. They can remain in abeyance until a specified date, when they are set to activate. Sometimes, the date chosen by the virus’ author is not an “ordinary” one, but it carries a very well-defined significance.

In this case, the virus will be named after the event that occurred on that specific date. One example of this would be the Chernobyl virus (CIH), activated on April 26, which is when the Chernobyl nuclear accident took place. Similarly, the Michelangelo virus remained latent until March 6, the Renaissance artist’s birthday.

f) Technical Aspects

The technical vulnerabilities exploited also come in handy when researchers name malicious codes. This is the case of the Sasser worm, which exploited a buffer overflow in the LSASS component (Local Security Authority Subsystem Service) of the affected operating systems. In the same vein, the name of Bandook Rat, a secure remote control software or Trojan that enables its user to work on a remote computer as if he/she were sitting in front of it, is actually short for Bandook Remote Administration Tool.

[12] Colleen Shannon and David Moore, 2004, The Spread of the Witty Worm, http://www.caida.org/research/security/witty/

[13] ***, 2004, More Doom?, http://www.newsweek.com/id/52912/page/1

[14] John Leyden, 2004, Santy worm defaces thousands of sites, http://www.theregister.co.uk/2004/12/21/santy_worm/

g) Personal Touch

Perhaps some of the most interesting explanations regarding malware names are those which cover etymological descriptions.

From a linguistic point of view, Conficker[15] is an excellent illustration of this tendency in malware naming. Conficker seems to combine the German verb "ficken," which means to fornicate, with "con," which, in Latin, means "with." In this way, the damage the malicious code can cause is more obvious. This is not the only example in this category, as lots of e-threats wear names containing “bad” words, in different languages.

Sometimes, malware names represent more or less funny situations in which those who discovered and analyzed the codes found themselves at that moment.

For example, the Code Red[16] worm, which appeared in July 2001, was named in this way because the researchers who discovered it had been drinking Pepsi's Mountain Dew Code Red when analyzing it and because of the phrase "Hacked by Chinese!" with which the worm defaced websites.

Melissa[17], on the other hand, is a romance-inspired virus name. Its creator was a 30 year old single computer programmer, whose love life was far from happy. Somewhere in his past there had been a special girl called Melissa and that’s how the virus got its name.

Conclusions

International standardization and coordination of the malware nomenclature is increasingly needed as more malware types are discovered at global scale. A consistent numeric identification is also increasingly required in order to make possible functional and analytical studies of specific malicious software.

Although it might be extremely unpractical to set a unique nomenclature system mandatory for all antivirus vendors, an attempt to standardize the malware naming process is in order, as the situation is confusing for analysts, and, most importantly, for computers users at large.

One thing is certain: the unification procedure should be flexible enough to cover all the e-threats that could appear and it should not place antivirus vendors under unreasonable restrictions.

Encouraging the adoption of a specific model from biology/medicine will most likely increase confusion, as the IT researchers should first be biologists or doctors in order to understand the system applied. Only then would they be able to accurately replicate this system in their own domain.

[15] Diane Prange, 2009, Conficker Naming: A Virus Named to... Screw with Your Computer, http://www.namedevelopment.com/blog/archives/2009/04/conficker_namin_1.html

[16] Moore, David, Colleen Shannon, 2001, The Spread of the Code-Red Worm (CRv2), http://www.caida.org/research/security/code-red/coderedv2_analysis.xml

[17] William Langley, 2000, A plague on all your mouses, http://www.fortunecity.com/emachines/e11/86/melissa.html

However, these sciences may have an answer to this problem. They might not provide the naming system proper, but an example of how to tackle the matter. For instance, for standardization purposes, biologists have created web-ontologies in which they can include the names they are used to, and which are automatically correlated with other classification systems.

References and selected bibliography

1. Langley William, 2000, A plague on all your mouses, http://www.fortunecity.com/emachines/e11/86/melissa.html
2. Leyden John , 2004, Santy worm defaces thousands of sites, http://www.theregister.co.uk/2004/12/21/santy_worm/
3. Mann Justin , 2007, MySpace speaks about Samy Kamkar's sentencing, http://www.techspot.com/news/24226-myspace-speaks-about-samy-kamkars-sentencing.html
4. Moore David, Colleen Shannon, 2001, The Spread of the Code-Red Worm (CRv2), http://www.caida.org/research/security/code-red/coderedv2_analysis.xml
5. Prange Diane, 2009, Conficker Naming: A Virus Named to... Screw with Your Computer, http://www.namedevelopment.com/blog/archives/2009/04/conficker_namin_1.html
6. Scheidl Gerald, 1999, Virus Naming Convention 1999 (VNC99), http://members.chello.at/erikajo/vnc99b2.txt
7. Shannon Colleen and Moore David, 2004, The Spread of the Witty Worm, http://www.caida.org/research/security/witty/
8. Skulason Fridrik, Bontchev Vesselin, 1991, A New Virus Naming Convention, http://vx.netlux.org/lib/asb01.html
9. Wells Joe, 1999, How Scientific Naming Works, http://www.wildlist.org/naming.htm
10. Virus Bulletin Conference: Industry unveils unified naming for virus threats, 2005, http://www.pcpro.co.uk/news/security/78425/virus-bulletin-conference-industry-unveils-unified-naming-for-virus-threats
11. ***, 2001, Kournikova computer virus hits hard, http://news.bbc.co.uk/2/hi/science/nature/1167453.stm
12. ***, 2007, Storm chaos prompts virus surge, http://news.bbc.co.uk/2/hi/technology/6278079.stm
13. ***, 2004, More Doom?, http://www.newsweek.com/id/52912/page/1
14. http://cme.mitre.org/data/list.html
15. http://www.economicexpert.com/a/Chen:Ing:Hau.html