Current Status of the CARO Malware Naming Scheme

Vesselin Bontchev, anti–virus researcher
FRISK Software International
Postholf 7180, 127 Reykjavik, ICELAND
E–mail: bontchev@complex.is

The Malware Naming Mess

- Glut
  - 250,000+ malware programs and rising
  - 5,000 new malware programs per month
- Lack of Time and Other Resources
  - We’re overloaded
  - Levels of (in)competence
  - Changing names is expensive
- Lack of a Common Virus Naming Standard
  - Sensible, Understandable, Usable

The Naming Mess - Continued

- Lack of Reliable Means for Automatic Malware Identification
  - `MyDoom.BQ` or `MyDoom.ED`?
  - Some tools exist:
    - F-VBACRC
    - SCIRD
    - PE-Info
  - Reference collection
    - Maintenance
    - Access

The Naming Mess - Continued

- Lack of Reliable Means for Automatic Malware Classification
  - MIRA
  - No such tool for binary viruses
  - And what about the packers?
- Inability to Enforce a Particular Naming Scheme
  - CARO is not an enforcement body
  - Willingness to do the job doesn’t imply competence

Alternate Naming Schemes

- Geographic Naming
  - impractical, leads to confusion
- Naming after the Infective Length
  - Sometimes it is variable
  - Sometimes it is meaningless
  - Different viruses can have the same length
- Descriptive Naming
  - Some malware doesn’t do anything visible
  - Different malware can have the same effects
  - The description is subjective
  - Requires time-consumming analysis

Alternate Naming Schemes - Cont.

- Naming after Some Text Found in the Virus
  - Not always present
  - Sometimes libelous and/or obscene
  - Boosts the malware author's ego
- Bezrukov's Naming Scheme
  - RCE-1800A, BP-EB
  - Difficult to remember
  - Different viruses can have similar names

Alternate Naming Schemes - Cont.

- Numeric Naming
  - Pretty much meaningless
  - Similar malware has very different names
  - Difficult to remember
- Enter the CARO Naming Scheme

History of the CARO Malware Naming Scheme

- Created in 1991 by Alan, Fridrik & Vess
- Malware grouped in families by code similarity
- Updated in 2002 and re-described by Nick
- Now and forever

http://www.people.frisk-software.com/~bontchev/papers/naming.html

The CARO Malware Naming Scheme

- General Format

  [<type>://][<platform>/]<family>[.<group>]
  [.<length>].<variant>[<modifiers>]
  [!<comment>]

  –The full names are unique

  –Only <family> and <variant> are mandatory

The CARO Naming Scheme - Cont.

- Malware Type
  - virus - recursive self-replication
  - dropper - drops malware
  - intended - wannabe virus
  - trojan - pretends to be benign but is malicious
  - pws - steals passwords
  - dialer - intercepts maliciously DUN connections
  - backdoor - provides unauthorized access
  - exploit - demonstrates security flaws (use CAN/CVE)
  - tool - including virus creation kits
  - garbage - self-explanatory

The CARO Naming Scheme - Cont.

- Platform
  - Short and long forms
  - Environment - not file type
  - See list
  - DOS is default
  - Multi-platform malware
    - virus://{W97M,X97M}/Foo.A
    - virus://O97M/Foo.A
    - virus://Multi/Foo.A
    - W97M/Foo.A & X97M/Foo.A

The CARO Naming Scheme - Cont.

- Family
  - General Format
    - charset **`[A-Za-z0-9_-]`**
    - Use “_And_” and “_Pct_” instead of ‘&’ and ‘%’
    - Use “_” instead of space
    - case insensitive
    - up to 20 characters
  - Rules for Constructing Proper Family Names

Constructing Proper Family Names - Don’ts

- No company names, brand names, people’s names
- No existing family, unless appropriate
- No new family, unless necessary
- No obscenities
- Don’t assume
- No numeric families
- No generic words

Constructing Proper Family Names - Do's

- Avoid the malware author's suggestion
- Avoid the file name
- Avoid the activation date
- Avoid geographic names
- If multiple acceptable names exist, select the one most commonly used already

Special Family Names

- HLLC - High Level Language Companion
- HLLO - High Level Language Overwriter
- HLLP - High Level Language Parasitic
- SillyB - Silly Boot Sector Virus
- SillyC - Silly COM-file infector
- SillyCE - Silly COM & EXE infector
- SillyCER - Memory-resident SillyCE

Special Family Names - Cont.

- SillyCR - Memory-resident SillyC
- SillyE - Silly EXE-file infector
- SillyER - Memory-resident SillyE
- SillyOR - Memory-resident overwriter
- SillyP - Silly MBR infector
- Trivial - Silly overwriter
- _<Number> - awaiting proper naming

Malware Relationship

- If packed or encrypted - unpack and decrypt
- Ignore non-code
- Fundamental differences - different families
- IF Related (A, B) THEN A and B are in the same family
- IF Related (A, X) AND Related (B, X) THEN A, B and X are in the same family
- IFF (A’ and B’ are in the same family) AND Related (A, A’) AND Related (B, B’) THEN A and B belong to the same family

Related (X, Y)

- Related (X, Y) ::=
  Average (Substrings (X, Y, N) / (Length (Y) - N + 1),
  Substrings (Y, X, N) / (Length (X) - N + 1)) > L;
- *Substrings (u, v, t)* is the number of all substrings of *u* of length *t* found within *v*
- *L* is ≈ 0.5-0.6

The CARO Naming Scheme - Cont.

- Group
  - Like a sub-family
  - Constructed the same way
  - Mainly for historical purposes; avoid
- Length
  - Number
  - No longer significant
  - Use only when meaningful

The CARO Naming Scheme - Cont.

- Variant
  - Variant Naming
    - A, B, … Z, AA, AB, … AZ, BA, BB, … BZ, CA, CB, … CZ, … ZZ, AAA, AAB, … ZZZ, AAAA, … etc.
    - In order of discovery - *not* in order of creation
  - Variant Reporting
    - Only when properly identified
    - Fuzzy variant reporting - Foo.{A-C,E}
  - Devolutions
    - Numbers appended to the variant name
    - Only for macro viruses
    - Report only when properly identified

The CARO Naming Scheme - Cont.

- Modifiers
  - General Format

    [:<locale>][{@<at_modifier>}]
  - Locale
    - Only for macro malware
    - Only the required Locale - not the supported one
    - English is the default
    - Platform major locales - not country or language
    - See list
    - Multiple locales

The CARO Naming Scheme - Cont.

- Modifiers - continued
  - At Modifiers
    - Specify important properties - e.g., @mm
    - Use only if the property is really present
    - See list (exp, i, irc, m, mm, p2p, s)
    - List multiple in alphabetical order
- Comment
  - free text devoid of white space

The CARO Naming Scheme - Cont.

- Conclusion
- Questions?

The Problems of the CME Initiative

Vesselin Bontchev, anti–virus researcher
FRISK Software International
Postholf 7180, 127 Reykjavik, ICELAND
E–mail: bontchev@complex.is

The Problems

- CAN/CVE - 73% of the vulnerabilities in the SANS @RISK bulletin have no CAN/CVE numbers
- Zotob.E has 2 different CME numbers, according to Symantec's site
- Two-hours timeout doesn't solve anything. What if the same malware is submitted again the next month?

The Problems - Continued

- What does MS03-039 do?
  - Hint: "No Blasters!”
- A scanner that doesn't identify exactly can report a CME number for the wrong threat
- Who is going to assert that something is a threat?
- How will names be changed or revoked if a mistake occurs?
- How are cross-references going to be made? (AV testing)

What will happen

- Viruses will appear without CME numbers
- CME numbers will appear without viruses
- The CME numbers will be late (overload)
- The same virus will get multiple CME IDs
- Different AV products will report different CME numbers for the same virus
- Nobody will remember which is the CME-nnn virus

What will happen - Continued

- Everybody will claim to be using CME
- MITRE will slap itself on the back
- The Anti–Virus industry will slap itself on the back
- US government bureaucracy will increase
  - But that will happen anyway
- Confusion will increase

= The users will lose

What is needed

- Exact identification
- Manual and competent malware analysis
- Competent collection maintainers
- Competent Anti-Virus testers
- Lots of the above

= A **very** good Anti-Virus company

  – and not an outsourced one

- But that ain’t gonna happen

Conclusion

- It’s not going to work
- But everybody will be claiming that it is
  - In other words - another WildList problem
- The users will be left bewildered
- In other words - the same old story
- Questions?