

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
22 July 2004 (22.07.2004)

PCT

(10) International Publication Number
WO 2004/062155 A2

- (51) International Patent Classification⁷: **H04L**
- (21) International Application Number:
PCT/IL2003/001047
- (22) International Filing Date:
10 December 2003 (10.12.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
10/335,871 3 January 2003 (03.01.2003) US
- (71) Applicant: **ALADDIN KNOWLEDGE SYSTEMS LTD.** [IL/IL]; 15 Beit Oved St., 61110 Tel Aviv (IL).
- (72) Inventors: **COHEN, Oded**; 12 Hertzl st., 36091 Kiriati Tivon (IL). **MEIR, Inbal**; 56 Harofe St., 34367 Haifa (IL). **MARGALIT, Yanki**; 6 Carmeli St., 52223 Ramat Gan (IL). **MARGALIT, Dany**; 10 Kiriatiy St., Ramat Chen, 52223 Ramat Gan (IL).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

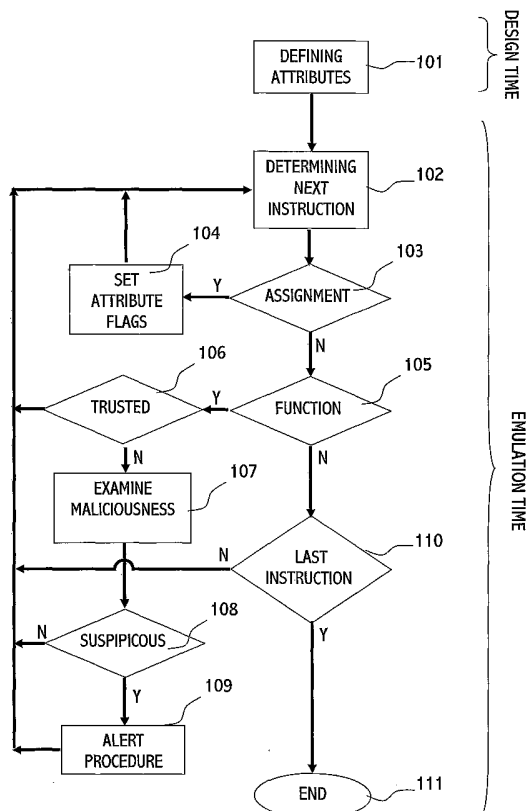
(84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: A METHOD FOR EMULATING AN EXECUTABLE CODE IN ORDER TO DETECT MALICIOUSNESS



(57) Abstract: A method for emulating an executable code, whether it is a human-readable code (e.g., macro and script) or a compiled code (e.g. Windows executable). At the design time, one or more "content attributes" (a property with relevance to maliciousness, e.g. a random value) are defined for the variables of the code. Also defined at the design time, is a list of malicious states (e.g., the combination of a call to a certain system function with certain content). When emulating an assignment instruction, the attributes of the assigned variable are set according to the assigned content. When emulating a mathematical operator, a content mathematics is also applied. When emulating a function call, the current state (the function identity and the calling content) is compared with the pre-defined malicious states, and if at least one malicious state corresponds, then the maliciousness of the code is determined.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A METHOD FOR EMULATING AN EXECUTABLE CODE
IN ORDER TO DETECT MALICIOUSNESS

Field of the Invention

5 The present invention relates to the field of detecting presence of virus and other kinds of malicious code by emulation.

Background of the Invention

Emulation is one of the methods of detecting viral activity. The emulator
10 creates a virtual environment under which the code of the tested sample is examined instruction-by-instruction. The viral activity is detected by correlating the changes with a viral behavior.

Emulation has several advantages in comparison to other methods of
15 detecting and/or preventing viral damages, since it helps to detect polymorphic viruses, encrypted viruses and unknown viruses, which cannot be detected by other methods.

A programmer can indicate the maliciousness of a source code just by
20 reviewing the code. For example, if he identifies a loop that gets the name of the files of a directory and deletes them, he can assume that the code is malicious. Contrary to a human programmer, an emulator, which is a machine, cannot indicate viral code just by reviewing the code, but it has to execute the code while comparing the caused changes with known

patterns. In other words, the emulator, being a machine, is absent of human intelligence.

It is an object of the present invention to provide a method for emulating
5 an executable code in order to detect maliciousness, which is more
“intelligent” than known in the prior art.

Other objects and advantages of the invention will become apparent as the description proceeds.

10

Summary of the Invention

The present invention is directed to a method for emulating an executable code, whether it is a human-readable code (e.g., macro and script) or a compiled code (e.g. Windows executable). At the design time, one or more
15 content attributes are defined for the variables of the code. A content attribute indicates a property with relevance to maliciousness, e.g. Windows directory, a random value, “.EXE” at the right of a string, etc. A content attribute may be implemented, for example, by a flag. Also defined at the design time, is a list of malicious states, where a malicious state
20 comprises at least the combination of a call to a certain system function with certain content, as the calling parameter(s). When emulating an assignment instruction, the attributes of the assigned variable are set according to the assigned content. When emulating a mathematical operator, a content mathematics is also applied. When emulating a

function call, the current state (i.e. the function identity and the calling content and values) is compared with the pre-defined malicious states, and if at least one malicious state corresponds, then the maliciousness of the code is determined.

5

The system function can be, for example, a part of the programming language of the executable, a function provided by the operating system under which the code is to be executed, a function provided by a linked package of routines.

10

Brief Description of the Drawings

The present invention may be better understood in conjunction with the following figure:

15 Fig. 1 schematically illustrates a method for emulating an executable, in order to detect malicious code.

Detailed Description of Preferred Embodiments

20 The term Macro in conjunction with computers, refers herein to an executable code embedded within another entity, like a document or executable code.

The term Script in conjunction with computers, refers in the art to a program built into the HTML code of a Web page that work on almost all

Internet applications. Scripts can also be used with a variety of applications that support their scripting syntax. VBScript and JavaScript are examples of scripting languages.

- 5 Macros and scripts are “non-compiled executables”, i.e. programs wherein their executable code is the source code, contrary to “compiled executables” which are converted to machine-code before execution. ASP and XML are other examples of non-compiled executables.
- 10 Due to their human readability, non-compiled executables are exploited by vandals (malicious executables written into scripts and macros). They can do almost anything - infect a system with Trojans, modify files, cause DoS attacks, or spread malicious code using Microsoft Networks or Microsoft Outlook. Macro-viruses that attack Microsoft Word work by copying
- 15 themselves into the global environment - usually by modifying NORMAL.DOT. Unfortunately, the enhanced power of software applications also gives virus writers the potential to cause more damage.

It should be noted that the examples herein are presented in a human-

20 readable programming code, although they apply to both compiled and non-compiled code. Moreover, the examples are presented in a schematic programming language, rather than an existing programming language.

There are some programming tools and computer resources commonly used by malicious code. For example, the random generator, which is a programming tool provided by almost any operating system and almost any programming language, is commonly used by viruses to prevent the repetition of the same code when duplicating itself, thereby hardening the detection of the malicious code by bit-pattern (also known as "virus signature") scanning. A human-readable code, as a program written in a programming language, makes use of variables. Since it is plain text, the mechanism that duplicates the malicious code can generate random names for its variables, to add non-effective code comments between its effective code lines, etc. The non-effective code may also be generated randomly, thus using the random generator.

The real time clock, which is a programming-tool provided by any operating system and almost any programming language, is another example of a virus-commonly used programming tool, since it enables a late-trigger of the virus code, and also because it can be used as a substitute to the random generator. Moreover, some of the viruses have a late trigger, thereby enabling them to be widely propagated before causing the damage, and consequently their discovery. The late trigger is tuned to a future date, which involves use of the real time clock of the host machine.

Another example of a content used by malicious code in Windows operating system is the Windows folder, which stores Windows system files. A virus may try to change one of the existing files in the Windows folder, create a file with a random name, etc.

5

A combination of some of the contents described above may also indicate the presence of malicious code within an executable. For example, creating a new file having a random name in the Windows directory can be due to a virus. However, a more accurate indication of a malicious code can be
10 obtained if this is an executable file. Thus, if the created file has an EXE extension, which is the extension of Windows executables, or VBS, which is the extension of Visual Basic Script, the executable can be classified as malicious.

15 According to a preferred embodiment of the present invention, a “malicious state” comprises at least executing a certain system function with certain content as its parameter(s). Thus, at least two elements are involved – a certain system function and certain content as the parameter(s).

20

It should be noted that the term “content” of a variable refers herein also to its value. Nevertheless, in some cases the value has been also mentioned.

A state can be indicated as malicious if it corresponds to at least one pre-defined malicious state. Two states "correspond" if their elements (i.e. function identity and the calling content) are equal or at least "similar" (e.g., at least certain percentage of their content / value is equal).

- 5 Maliciousness can be defined also by tracking the history of the called functions, e.g., if certain functions have been called with certain content, or even without regard-ness to their content.

Identifying a call to a certain function is quite obvious, and almost any
10 compiler parser performs it. However, the concept of dealing with the content of a parameter rather than its value is novel.

For example, the parameter when calling the following function is a random number:

15

```
MyVariable = Random()  
MyVariable = MyVariable + 1  
Call MyFunction(My_Variable)
```

- 20 Where *Random()* is a random number generator.

The first time *MyFunction* is performed, its parameter may be, e.g., 0.443213, and the next time it can be, e.g., 0.332455. Thus, from analyzing

the text it is not possible to indicate that the parameter has a random nature.

According to a preferred embodiment of the invention, the content of a
5 variable can be indicated as follows:

Each variable of the executable is characterized by several attributes (e.g. flags), which points out on its content (referred herein as content attributes). For example:

- 10 - a flag to indicate if the variable has been assigned with the Windows folder name;
- a flag to indicate if the variable has been assigned with a random value;
- a flag to indicate if the variable has been assigned with the real
15 time;
- a flag to indicate if the variable has been assigned with the value
 ".EXE", which is the extension of a type of executable file under the
 Windows operating system;

and so forth.

20

While emulating the execution of an executable, the following instructions (for example) are performed:

File_name = Windows_directory() + Random() + 0.2 + ".EXE"

Create_file(File_name)

Where: *Windows_directory()* is a system function that returns a string of the Windows directory, and *Random* returns a numeric random integer.

5

In this example, the Windows folder flag, the Random flag and the EXE flag of the variable *File_name* are turned on.

It should be noted that

10 *File_name = Windows_directory() + Random() + 0.2 + ".EXE"*

actually comprises several instructions:

- Executing the *Random* function;
- Executing the *Windows_directory* function;
- Mathematical addition;
- 15 - String addition.

It should also be noted that while some programming languages convert a numeric value to a string value and vice versa when it is required, some compilers have no such ability. Thus, assuming the function *Str* converts a numeric value to a string, in this case the example should look like:

20

File_name = Windows_directory() + Str(Random() + 0.2) + ".EXE"

There is no point in considering assigning a value to variables as malicious operation; however without tracking the history of the

assignments to a variable, its content cannot be determined. Thus, the examining stage should be at the point where specific system functions which may cause damage to the computer are called, e.g. a function that creates a file, a function that changes the content of a file, and so forth.

5

The following instruction also creates an executable file with a random name in the Windows directory:

Create_file(Windows_directory() + Random() + ".EXE")

10

However, instead of using a variable as the parameter of the function *Create_file*, the name of the file is the parameter. Actually, this is equivalent to creating a temporary variable, and assigning a value to the temporary variable just before the execution of said function:

15

Temp_parameter = Windows_directory() + Random() + ".EXE"

Call Create_file(Temp_parameter)

As mentioned above, according to the present invention, the variables
20 comprise content attributes in addition to the value. Thus, there is a question as to how the content is influenced by mathematical operations.

For example:

MyValue = 123 + First_var + Second_var

Where, for example, the Real time flag of the variable *First_var* is turned on and the EXE flag of the variable *Second_var* is turned on. It is obvious that in this case the content result of the assignment is turning on the Real time flag and the EXE flag of the variable *MyValue*. Thus, the
5 emulator should be provided with “content mathematics” rules. Such rules, which for example comprise a commutative law of addition, an associative law of addition and so forth, should also take into consideration comparison statements and other features supported by the emulated programming language. Of course it is desired to emulate the
10 values as well, in addition to emulating the content.

Another example:

Var_A = GetAddressBook()

Var_B = Var_A->Count()

15 Where *Var_A* has the content type “address book object”.

When the operator “*->Count()*” is operated on an “address book object”, it returns not just a numeric value, but a numeric value with the content attribute “number of entries in address book”, thus adding the “number of
20 entries in address book” to the variable *Var_B*.

Another example:

Var_A = GetAddressBook()

Var_C = Var_A[2]

After the assignment to *Var_C*, its value will be the second entry of the address book which has been assigned to *Var_A*. The content attributes of *Var_A* may be something like "address book database", and the content attribute of *Var_C* may be something like "address book entry".

5

Fig. 1 schematically illustrates a method for emulating an executable, in order to detect malicious code.

- At step 101, which is the design step, a group of malicious states is defined. A malicious state is the combination of a call to a certain system function with certain content. This is a preliminary stage, which is performed at the anti-virus firm, as a part of the anti-virus emulator design. Of course the list of the content attributes and/or functions can reside on a file, and modified later (whether at the anti-virus firm residence or at the installation site).
- At the next step 102, which is performed at the emulation time (i.e. run time), the next instruction to be performed is determined.
- At the next step 103, the type of the instruction is examined. If it is an assignment instruction, then the flow continues with step 104, otherwise the flow continues with step 105.

- At step 104, the instruction is analyzed in order to determine the content attribute(s) of the assigned variable to be set, and the corresponding values are set, and from there, the flow returns to step 102.

5

- At step 105, if the type of the instruction is a function call, then the flow continues with step 106; otherwise the flow continues with step 110.

10

- From step 106, if the function is pre-determined as trusted, then the flow continues to step 102; otherwise flow continues to step 107.

15

- At step 107 the maliciousness is examined by comparing the current state (i.e. the calling function with the content attributes of the parameters) with the malicious states that have been pre-defined in step 101. Determining the maliciousness of a code can be indicated if the current state corresponds to at least one of the pre-defined malicious states.

20

- From step 108, if the instruction seems to be suspicious according to the examination carried out at step 107, then the flow continues with step 109, where an alert procedure is performed, and from there to step 102.

- From step 110, if the last tested instruction is the last instruction (e.g. a Stop instruction), then the emulation stops, otherwise the flow continues with step 102.

5 Usually, the emulation continues until the last instruction is performed, and then the results of the emulation are examined. This way, more details are gathered and the maliciousness of the code can be determined more precisely. However, the alert procedure can suspend further execution, thereby shortening the emulation time.

10

As known to a person of ordinary skill in the art, a function can be declared such that a parameter is transferred by its address or by its value. In case of transferring a value, a temporary parameter is created and its address is transferred as a parameter, like in the other case. In
15 both cases when emulating the code, the variables, whether permanent or temporary, also comprise the content attributes.

It should be noted that the term "function" also refers herein to procedures, software interrupts and so forth.

20

The emulation can be executed at the user's machine, at the gateway to a network and so forth. For example, while a user may install an emulator to check every incoming HTML file while browsing the Internet / Intranet,

an organization may install such an emulator at the gateway to its network, at the mail server and so forth.

- 5 Those skilled in the art will appreciate that the invention can be embodied by other forms and ways, without losing the scope of the invention. The embodiments described herein should be considered as illustrative and not restrictive.

CLAIMS

1. A method for emulating an executable code, said method comprising:

- pre-defining at least one content attribute having relevance to
5 maliciousness, for at least one variable of said executable code;
- pre-defining at least one malicious state, being a call to a certain
system function with certain content and/or value;
- upon emulating an assignment instruction, setting the content
attributes of the assigned variable according to the assigned
10 content, and the value of the assigned variable;
- upon emulating a mathematical operator, employing content
mathematics; and
- upon emulating a function call, checking that the current state,
being the combination of the function identity with the content
15 and/or value of the calling parameters, corresponds to at least one
of the pre-defined malicious states, thereby indicating the
maliciousness of said code.

2. A method according to claim 1, wherein said checking is selected from a

- 20 group comprising: testing the equality between the function identity
and the content and/or value of the calling parameters, testing the
similarity of the function identity and the content and/or value of the
calling parameters, testing if certain functions have already been

called, testing if certain functions have been already called with certain parameters.

3. A method according to claim 1, wherein said executable is selected from
5 a group comprising: compiled code and human-readable code.

4. A method according to claim 1, wherein said at least one content
attribute is selected from a group comprising: a programming flag, a
Boolean value, a numeric value, and a string value.

10

5. A method according to claim 1, wherein said at least one content
attribute is selected from a group comprising: a random value, a time
value, a certain folder name, a certain file extension, the number of the
files in a folder.

15

6. A method according to claim 1, wherein said system function is selected
from a group comprising: the programming language of executable, the
operating system under which said code is to be executed.

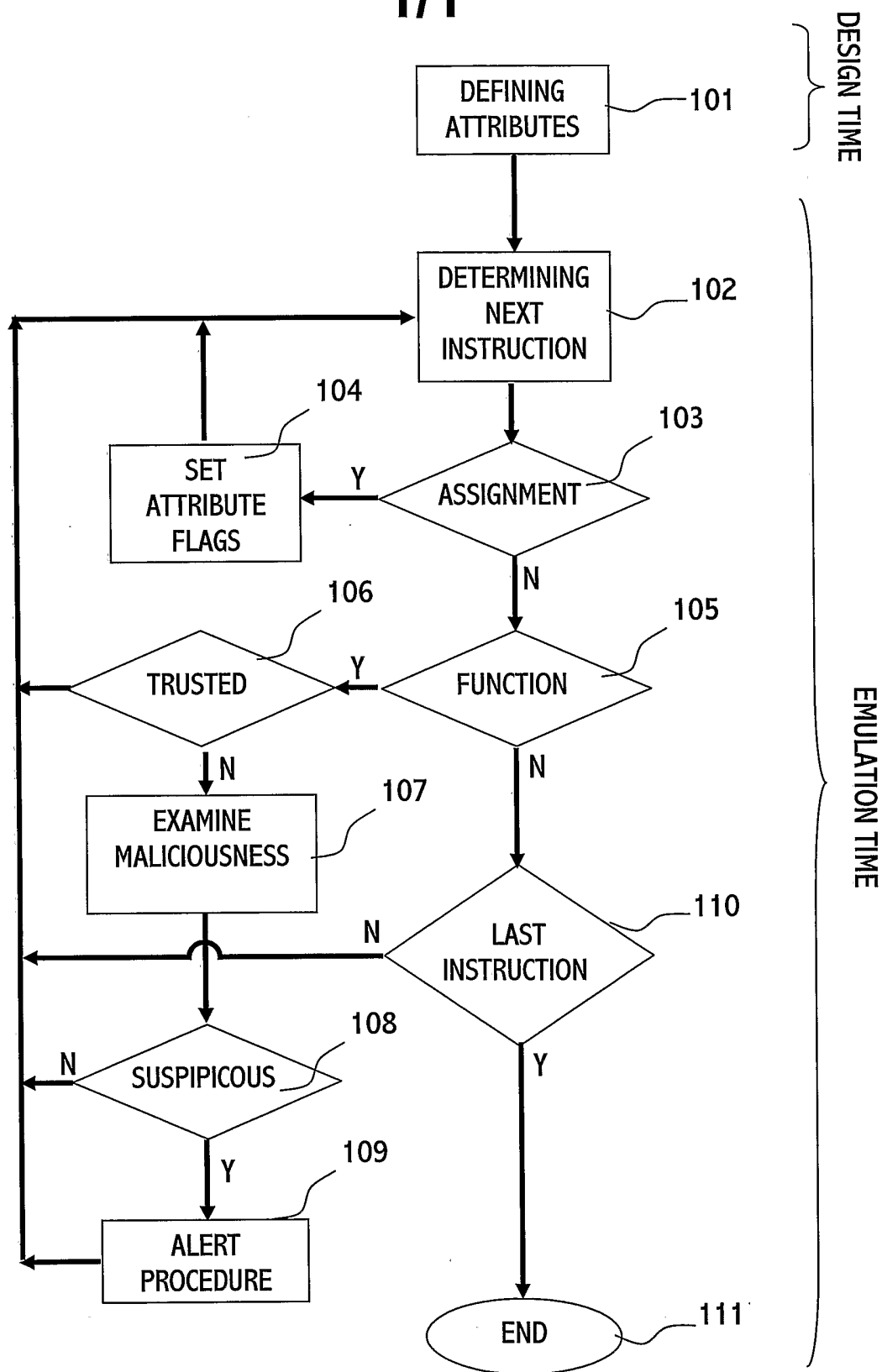
20 7. A method according to claim 1, wherein said parameter(s) are selected
from a group comprising: permanent parameters, and temporary
parameters.

8. A method according to claim 1, wherein said parameter(s) are passed by their content or by their address.

9. A method according to claim 2, wherein said compiled code is selected
5 from a group comprising: Windows EXE, Windows DLL, Windows OCX, Linux executables, Solaris executables.

10. A method according to claim 2, wherein said human-readable code is selected from a group comprising: macro, script, VBScript, JavaScript,
10 BAT, PHP, and ASP.

1/1

**Fig. 1**