Internal distribution code:

(A) [-] Publication in OJ
(B) [-] To Chairmen and Members
(C) [-] To Chairmen
(D) [X] No distribution

Datasheet for the decision of 27 April 2021

Case Number: T 2386/16 - 3.5.01

Application Number: 10828858.0

Publication Number: 2494446

IPC: G06F11/00, G06F21/56, H04L29/06

Language of the proceedings: EN

Title of invention:
USING FILE PREVALENCE TO INFORM AGRESSIVENESS OF BEHAVIORAL HEURISTICS

Applicant:
CA, Inc.

Headword:
MALWARE DETECTION USING FILE PREVALENCE TO INFORM AGRESSIVENESS OF BEHAVIORAL HEURISTICS/CA INC

Relevant legal provisions:
EPC Art. 56, 84

Keyword:
Inventive step - (no) - main request and auxiliary requests 1 and 2

Europäisches Patentamt
European Patent Office
Office européen des brevets

Beschwerdekammern

Boards of Appeal

Chambres de recours

Boards of Appeal of the
European Patent Office
Richard-Reitzner-Allee 8
85540 Haar
GERMANY
Tel. +49 (0)89 2399-0
Fax +49 (0)89 2399-4465

Case Number: T 2386/16 - 3.5.01

D E C I S I O N
of Technical Board of Appeal 3.5.01
of 27 April 2021

Appellant: (Applicant)	CA, Inc. 1320 Ridder Park Drive San Jose, CA 95131 (US)
Representative:	Bosch Jehle Patentanwaltsgesellschaft mbH Flüggenstraße 13 80639 München (DE)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 23 May 2016 refusing European patent application No. 10828858.0 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairwoman	A. Wahrenberg
Members:	M. Höhn
	P. Schmitz

Summary of Facts and Submissions

I. This appeal is against the decision of the examining division refusing European patent application No. 10828858.0 pursuant to Article 97(2) EPC on the ground of lack of novelty (Articles 52(1) and 54(1) and (2) EPC) with regard to prior-art publication:

D7: US 2007/0079379 A1.

II. In its statement setting out its grounds of appeal, the appellant requested that the appealed decision be set aside and that a patent be granted on the basis of the main request or auxiliary request 1 on which the decision under appeal was based, or auxiliary request 2 submitted with the grounds of appeal. Oral proceedings were requested as an auxiliary measure.

III. In a communication dated 8 June 2020, the Board raised an objection under Article 84 EPC for lack of clarity and expressed its preliminary opinion that all requests lacked an inventive step (Article 56 EPC).

IV. In a reply by letter dated 15 October 2020, the appellant submitted further auxiliary requests 3 and 4 together with arguments in support of clarity and inventive step.

V. Oral proceedings were held on 27 April 2021 in the course of which the appellant withdrew the main request and auxiliary requests 1 and 2 and made auxiliary requests 3 and 4 their main and auxiliary request 1, respectively. The appellant filed an auxiliary request 2 consisting only of the independent system claim 12.

Thus, the appellant's final request was that the decision under appeal be set aside and that a patent be granted based on the main request or auxiliary request 1 (former auxiliary requests 3 and 4) filed with letter dated 15 October 2020, or auxiliary request 2 filed during the oral proceedings before the Board.

After due consideration of the appellant's arguments the Chair announced the decision.

VI. Independent claim 1 according to the main request reads as follows:

"1. A computer implemented method for adjusting an aggressiveness level (303) to use in behavior based heuristics malware detection, based on target file prevalence rates, the method comprising the steps of:

determining, by a computer, a prevalence rate (309) of a file (305) subject to behavior based heuristics analysis to determine whether the file (305) comprises malware, wherein widely distributed files (305) have the highest prevalence rate (309), and singleton files (305), which have not seen by the computer before, have the lowest prevalence rate (309);

adjusting, by a computer (210), the aggressiveness level (303) with which to conduct the behavior based heuristics analysis on the file (305), responsive to the determined prevalence rate (309) of the file (305), comprising setting the aggressiveness level (303) to a higher level for files (305) having a lower prevalence rate (309), and to a lower level for files (305) having a higher prevalence rate (309), wherein the highest aggressiveness level (303) is set for singleton files (305);

applying, by a computer (210), behavior based heuristics analysis to the file (305), using the aggressiveness level (303) adjusted according to the prevalence rate (309) of the file (305); and

determining, by a computer (210), whether the file (305) comprises malware (321), based on the applied behavior based heuristics analysis."

Claim 1 of auxiliary request 1 adds the following as the penultimate feature: "determining, by a computer (210), which file attributes to measure during the behavior based heuristics analysis of the file (305), responsive to the aggressiveness level (303) adjusted according to the prevalence rate (309) of the file (305);".

Claim 1 of the second auxiliary corresponds to the independent system claim 12 of auxiliary request 1 and reads as follows:

1. A computer system (210), for adjusting an aggressiveness level (303) to use in behavior based heuristics malware detection, based on target file prevalence rates, the computer system comprising:

a processor (214);

computer memory (217);

means (307) for determining a prevalence rate (309) of a file (305) subject to behavior based heuristics analysis to determine whether the file (305) comprises malware, wherein widely distributed files (305) have

the highest prevalence rate (309), and singleton files (305), which have not seen by the computer before, have the lowest prevalence rate (309);

means (301) for adjusting the aggressiveness level (303) with which to conduct the behavior based heuristics analysis on the file (305), responsive to the determined prevalence rate of the file (305), comprising setting the aggressiveness level (303) to a higher level for files (305) having a lower prevalence rate (309), and to a lower level for higher prevalence files (305) having a higher prevalence rate (309), wherein the highest aggressiveness level (303) is set for singleton files (305);

means for applying behavior based heuristics analysis to the file (305), using the aggressiveness level (303) adjusted according to the prevalence rate (309) of the file (305);

means for determining which file attributes to measure during the behavior based heuristics analysis of the file (305), responsive to the aggressiveness level (303) adjusted according to the prevalence rate (309) of the file (305); and

means for determining whether the file (305) comprises malware (321), based on the applied behavior based heuristics analysis.

VII. The appellant argued that the virus score value in D7 did not correspond to the prevalence rate of a file in claim 1. While the virus score depended on the probability of a virus, it did not contain any information about the prevalence of a specific file. Figure 8 of D7 was a hypothetical example of how the

number of infected machines decreased over time, and merely demonstrated when an updated version of the virus signature was published. The new signature could be used to disinfect machines which could lead to a decrease in the number of infected machines.

By contrast, in claim 1, the prevalence rate was determined not just for files that were infecting machines or known to include viruses, but for every file to be tested to determine whether it contained a virus. D7 did not disclose or suggest "determining...a prevalence rate of a file" without knowing whether or not the file included a virus or malware, and then adjusting the test to determine if malware was included in the file based on the prevalence rate.

The examining division had wrongly understood the "prevalence rate" as applying only to the prevalence of a virus, whereas the prevalence rate in claim 1 was related to the distribution of a file where it was unknown whether that file had a virus/malware or not. In D7, the "prevalence rate" applied to all files, not just files that were known to contain computer viruses.

The appellant further argued that the claims differed from D7 by the feature "setting the aggressiveness level to a higher level for lower prevalence files and to a lower level for higher prevalence files."

In claim 1 the "prevalence rate" was determined for files without knowing whether the files contained a virus/malware or not, and then the test used to determine whether the files were infected was adjusted based on their prevalence.

Paragraphs [0227] to [0229] of D7 cited by the examining division related to the detection of viruses using a heuristic approach. However, the aggressiveness level of the heuristics did not depend on the prevalence rate of a file. Paragraph [0248] of D7, disclosing that the rules may be adapted later to better scope the virus, was missing that the rule did not relate to the prevalence rate of a file.

The subject-matter of claim 1 was therefore novel over D7. It was furthermore inventive since D7 gave no hint to use the prevalence rate of a file to adjust the behaviour-based heuristics analysis of files in order to improve the virus detection.

The additional features of claim 1 according to auxiliary request 1 further clarified the claimed "aggressiveness level".

Reasons for the Decision

1. The Board's interpretation of claim 1:

The invention concerns behavior based heuristics malware detection. Instead of using a static approach, the invention suggests to have varying dynamic measures to be applied to a file dependent on how likely the file is considered to be infected. The more likely a file is to contain malware, the stronger will be the measures. The "aggressiveness level" of the malware detection is dynamically adjusted based on the "prevalence rate" of a file such that lower prevalence files are treated more aggressively than higher prevalence files.

The expressions "prevalence rate" and "aggressiveness level" do not have a specific technical meaning in the art and the claim does not provide its own definition. Both expressions were objected to under Article 84 EPC for lack of clarity, not only in the Board's communication, but also during the first instance proceedings (see the European Search Opinion).

The Board interprets "prevalence rate" in view of Figure 3 and paragraphs 22 to 24 of the application as published as the number of copies of a file in a network indicating how widely the file is distributed.

The term "aggressiveness level" is interpreted as a level with which to conduct behaviour based heuristics analysis on the file, including various measures for isolating the file or stopping execution of the file in order to further analyse it according to the set level.

Main request

Article 56 EPC - Inventive step

2. The Board agrees with the contested decision that D7 discloses a computer implemented method for adjusting an aggressiveness level to use in behavior based heuristics malware detection (paragraph 226), based on target file prevalence rates, with the steps of:

- determining, by a computer, a prevalence rate of a file subject to behavior based heuristics analysis to determine whether the file comprises malware ("virus score value for the message based upon one or more rules that specify attributes of messages that are known to contain computer viruses, wherein the

attributes comprise a type of file attachment to the message, a size of the file attachment, and one or more heuristics based on the message sender, subject or body and other than file attachment signatures", paragraph 45), wherein highly prevalent files are widely distributed (see Figure 8, the horizontal axis 814 represents time and vertical axis 812 represents a number of infected machines. Point 806 represents a time at which an updated virus definition that will detect a malicious message and prevent further infection on machines in networks protected by messaging gateways 107 that are using that anti-virus software, paragraph 228. The fact that a virus signature is available corresponds to a prevalence rate as shown in the figure. The passage "virus score value for the message based upon one or more rules that specify attributes of messages that are known to contain computer viruses" of paragraph 45 discloses this feature, because viruses that are identified and known necessarily have a prevalence in contrast to unknown viruses which have not yet been seen and thus have a lower prevalence),

- adjusting, by a computer, the aggressiveness level to use in the behavior based heuristics analysis of the file ("quarantine", paragraph 227 to 228 in particular also early quarantine at an early point 810 of the outbreak as being particularly aggressive), responsive to the determined prevalence rate of the file (Figure 8, discussion of outbreak in paragraph 228), comprising setting the aggressiveness level to a higher level for lower prevalence files and to a lower level for higher prevalence files ("A message can be placed in quarantine storage, because it may contain a virus, based on the results of heuristic operations rather

than or in addition to definitions of virus outbreaks", paragraph 227);

- applying, by a computer, behavior based heuristics analysis to the file, using the aggressiveness level ("different actions can be taken on quarantined messages when those messages leave the quarantine based on the threat level associated with the messages when they leave the quarantine", paragraph 243); and

- determining, by a computer, whether the file comprises malware, based on the applied behavior based heuristics analysis (releasing of files from the outbreak buffer means scanning the files, see e.g paragraph 105 and 245).

D7 also discloses using stricter scanning rules at later times of the outbreak, which also discloses adjusting the aggressiveness level up for lower prevalence files (see paragraph 248) since that number declines during the outbreak (see also Figure 8).

3. In contrast to the appellant's arguments, D7 further discloses that one of the criteria for determining a virus rate is related to the number of messages distributed in the network during a time period, see e.g.

- [0063]: generates counts of received messages; creating a count of suspicious messages received in a particular time period; providing the count to virus information processor;
- [0064]: generate counts of messages;
- [0073]: increased mailing list activity; suspicious volume patterns;

- [0074]: the current rate of the number of messages containing attachments;

In particular paragraph [0073] discloses that a virus score is determined based on the percentage of dynamic network addresses or of computerized hosts associated with sources of received messages. The Board considers this to be information related to how widely a message is distributed in the network and therefore as equivalent to a prevalence rate in general.

During oral proceedings the appellant further argued that in contrast to the claimed invention D7 did not disclose an adjustment of the analysis of the file based on the prevalence rate. The Board does not agree, because D7 discloses the use of stricter scanning rules at later times of the outbreak (see last paragraph of point 2 above). D7 explicitly discloses a re-scanning step of messages with an adapted set of rules for a virus outbreak that are initially broad and of which the scope is later narrowed down (see paragraph 248 of D7). As a result, messages that matched the earlier rule set may not match the revised rules, and become known false positives. The approach in D7 attempts to release known false positives automatically in response to a rule update, without intervention. This corresponds to the present invention as described in paragraphs 25 and 26 of the description, which is covered by claim 1.

4. Furthermore, D7 suggests varying measures for isolating files or stopping the execution of files such that they can be further analysed. The Board regards this as an aggressiveness level to be used in a behaviour based heuristics malware detection.

4.1 D7 discloses that when the virus score value is greater than or equal to a specified threshold, the message is stored in a quarantine queue without immediately delivering it to the recipient account (see [0045]). In addition, a variable quarantine time is suggested (see [0229]). A minimum quarantine delay is applied to messages that are less likely to contain a virus. Thus, the quarantine time is coupled to the probability that a message contains a virus. The Board considers this to be at least equivalent to the aggressiveness level in the context of claim 1.

4.2 Claim 1 further specifies "singleton files", which have not been seen by the computer before and have the lowest prevalence rate. Those singleton files are treated with the highest aggressiveness level.

Such "singleton files" are not explicitly disclosed in D7. The Board is however not convinced that treating files having the lowest prevalence rate (singleton files) more aggressively involves an inventive step. In the Board's view, this does not necessarily improve the virus detection over D7, which already suggests that the aggressiveness level is coupled to the probability that a message contains a virus.

Defining singleton files as the files most likely to contain malware is considered to be a mere heuristic, which does not reliably provide a technical effect. Also D7 uses heuristics. Depending on probabilities and subjective experiences, the skilled person would consider many heuristic approaches to define such files within the general idea of applying the highest aggressiveness for files most likely to contain malware.

In view of this the Board considers the feature of treating singleton files the most aggressively to be an obvious choice for the skilled person.

4.3 For the reasons given above, the Board judges that the subject-matter of claim 1 does not involve an inventive step (Article 56 EPC) over the teaching of D7 in view of the skilled person's common general knowledge.

Auxiliary request 1

5. The further features according to claim 1 of this request are at least rendered obvious in view of the re-scanning step in combination with paragraphs 136 and 137 in D7, where it is disclosed that the message characteristics of an incoming message are identified and it is determined which rules of the rule set are matched based on the message characteristics for the message.

Therefore, the subject-matter of claim 1 of the first auxiliary request does not involve an inventive step (Article 56 EPC) over the teaching of D7 in view of the skilled person's common general knowledge.

Auxiliary request 2

6. Claim 1 according to the second auxiliary request is directed to a computer system with apparatus features corresponding to the method of claim 1 of the first auxiliary request.

The appellant argued that in contrast to the disclosure of D7, which was based on a distributed system

comprising a client-server-architecture, the claimed system was directed to a stand-alone computer, which did not require any external access or network connection.

6.1 The Board is not persuaded by the appellant's arguments. Indeed, the appellant's interpretation of claim 1 is contradicted by the published application where it is described that the prevalence rate is based on input from a source (see paragraph 24), which may include an external source. Thus, the Board considers that the claim 1 of the second auxiliary request does not differ in substance from claim 1 of the first auxiliary request. Consequently, the same reasons apply.

6.2 In any case, the Board does not see any technical benefit with regard to malware detection that could be achieved from having a stand-alone system or any technical hurdle to be overcome with the need for inventive skills.

6.3 The subject-matter of claim 1 according to auxiliary request 2 therefore does not involve an inventive step (Article 56 EPC).

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chair:

T. Buschek A. Wahrenberg

Decision electronically authenticated