Internal distribution code:

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

Datasheet for the decision of 7 September 2021

Case Number: T 2147/16 - 3.4.03

Application Number: 13194781.4

Publication Number: 2811441

IPC: G06Q10/10, H04L12/58

Language of the proceedings: EN

Title of invention:
System and method for detecting spam using clustering and rating of e-mails

Applicant:
Kaspersky Lab, ZAO

Headword:

Relevant legal provisions:
EPC Art. 56

Keyword:
Inventive step - (no) - effect not made credible within the whole scope of claim - improvement not credible - obvious combination of known features - mixture of technical and non-technical features - problem and solution approach - obvious solution

This datasheet is not part of the Decisior
It can be changed at any time and without notic

Decisions cited:

G 0001/19, G 0003/08, T 1741/08, T 1316/09, T 0022/12, T 1849/17, T 1028/06, T 1179/14, T 1028/14, T 0340/14, T 2535/16, T 1670/07, T 1358/09, T 1784/06, T 0258/03, T 0154/04, T 0641/00, Decision of national courts cited: German Federal Patent Court (Bundespatentgericht) 20 W (pat) 13/09

Catchword:

The mere assumption that an algorithm is optimised for the computer hardware and may have a technical contribution is not sufficient. The implementation of an algorithm in a method for filtering spam messages must have a proved further technical effect or specific technical considerations; such further technical effect must be specifically and sufficiently documented in the disclosure of the invention and be reflected in the claim wording; the algorithm must serve a technical purpose.

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 2147/16 - 3.4.03

DECISION
of Technical Board of Appeal 3.4.03
of 7 September 2021

Appellant: (Applicant)	Kaspersky Lab, ZAO 39A/3 Leningradskoe Shosse Moscow 125212 (RU)
Representative:	Sloboshanin, Sergej V. Füner, Ebbinghaus, Finck, Hano Mariahilfplatz 3 81541 München (DE)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 3 May 2016 refusing European patent application No. 13194781.4 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman	G. Eliasson
Members:	A. Böhm-Pélissier
	G. Decker

Summary of Facts and Submissions

I. The appeal is against the decision of the Examining Division to refuse European patent application No. 13 194 781 on the basis of lack of inventive step (Article 56 EPC).

II. Reference is made to the following documents:
D1 = US 2012/215853 A1
D5 = US 2012/030293 A1
D6 = US 2004/260922 A1
D7 = US 2010/191819 A1

III. The Appellant (Applicant) requests that the decision under appeal be set aside and that a patent be granted on the basis of the Main Request or the Auxiliary Request as filed with the statement of grounds of appeal. Oral Proceedings were requested in case the Board should come to the conclusion that it is not possible to grant a patent on the basis of the Main Request.

IV. With the letter of 25 March 2021 the Appellant informed the Board that it did not intend to attend the oral proceedings scheduled on 23 April 2021. Furthermore, it requested that a decision be made on the basis of the filed documents. Subsequently the Board cancelled the date for the oral proceedings.

V. Highlighting (Additions, ~~deletions~~, **bold**) and labeling in citations were added by the Board.

VI. Claim 1 of the **Main Request** reads:
(A) A computer-implemented method for classification of electronic messages being e-mail messages as spam or

legitimate on a client mail server (200), the method comprising:
(B) receiving (510), by a computer processor (15) of the client mail server (200), all electronic messages (180) directed to different users who have email addresses registered on the client mail server (200);
(C) classifying (515) the received electronic messages (180) as legitimate or spam messages;
(D) identifying (535) unknown messages that could not be classified as legitimate or spam;
(E) obtaining (550) metadata (860) of the unknown messages,
(F) the metadata (860) including at least a set of hash sums (830) for an unknown message and an IP address of a sender of the unknown message;
(G) placing (553) the metadata (860) of the unknown messages into one cluster of a plurality of clusters forming a cluster index tree based on a degree of similarity between a newly-arrived set of hash-sums $\{h_1; h_2 \ldots h_k\}$ from the metadata (860) and existing sets in the cluster index tree
(H) in accordance with: $D = \boldsymbol{\Sigma}\, w_j /k$,
(I) wherein the sum is formed over all weighting factors w_j corresponding to hash sums in the cluster index tree that are matched by hash sums in the newly-arrived set $\{h_1; h_2 \ldots h_k\}$;
(J) rating (590) each unknown message in accordance with a rating of the cluster where the metadata (860) of the unknown messages was assigned to,
(K) wherein the rating of the cluster is based, at least, on a number of similar hash sums (830) of unknown messages received from different IP addresses of message senders contained in said cluster; and
(L) classifying (560, 565) unknown messages as legitimate or spam based on message ratings.

VII. Claim 1 of the **Auxiliary Request** reads:
[Features (A)-(L)],
(M) wherein the rating of the cluster is reported to the client mail server (200) which makes a decision as to the message based on the received rating.

VIII. The Examining Division argued essentially as follows:
(a) a spam filter did not serve a technical purpose;
(b) it did not produce a relevant technical effect;
(c) it did not imply technical considerations.

IX. The Appellant argued essentially as follows:
(a) the method facilitated an improved performance of a computer and computer network;
(b) the method decreased on the end user computer the memory volume of the hard disk required for the legitimate electronic messages due to decreasing an amount of spam;
(c) the method decreased a volume of the email traffic to the end user computer;
(d) the method decreased a load for the processor at the end user computer.

Reasons for the Decision

1. The appeal is **admissible**.

2. <u>**The invention as claimed**</u>

2.1 The present invention has as object to provide an efficient spam filter. Spam messages present the inconvenience of "cluttering" the user's email box, such that the user may easily overlook an important message in the flow of numerous spam messages. Spam messages take up a substantial volume of the email traffic. The share of spam includes more than 70% of global mail traffic. The spam messages accumulate in user mail boxes and take memory space and time to clean out (page 1, line 10ff of the description as filed).

2.2 Among the spam messages there are messages with computer viruses leading to harm of computer hardware/ software and messages which lead to phishing resources that can be a cause of theft of passwords and personal data. Prior methods of spam detection, which typically involve grouping of messages by various criteria, often result in false positives, i.e., when the spam message ends up in the group of legitimate messages or when the legitimate message is grouped as spam.

2.3 In order to overcome these drawbacks the present invention proposes an algorithm comprising the steps of:

- hashing: creating metadata comprising hash sums and IP addresses;
- clustering: creating clusters and a cluster index tree based on a degree of similarity;
- weighting: creating weights for the hash sums;
- summing up the hash sums;
- rating each message according to a rating of the cluster to which the metadata of the message is assigned.

3. Admissibility of the new requests

The claims of the Main Request and the First Auxiliary Request, which were filed for the first time with the statement of grounds of appeal, contain new features having basis only in the description and which could have been filed before the Examining Division. However, as the amendments are not complex and can be dealt with without considerable additional effort, the Board decided not to exercise its discretion under Article 12(4) RPBA 2007 to hold the new requests inadmissible.

4. **Clarity**

In claim 1 of both requests the variable *k* is not defined in the expression $D = \boldsymbol{\Sigma}\ w_j\ /k$.
This formula is an essential feature of the invention. *k* appears to be the number of weights being summed up. However, the description does not provide a direct and unambiguous disclosure for this fact. The corresponding passage in the description merely discloses that *"[t]he sum here is formed over all wj corresponding to the hash sums in the tree of clusters that are matched by hash sums in the arriving set* $\{h_1;\ h_2...h_k\}$ (see the last two lines of page 17 of the application as filed). Consequently, the Board is of the opinion that the requests do not comply with the requirements of Article 84 EPC.

5. **Main Request - inventive step**

5.1 **Closest prior art**

D5 was cited as prior art in the US procedure. D5 teaches in addition to the teachings of D1 weighting and taking IP addresses into account for the spam analysis. The Board considers D5 a better springboard for the problem and solution approach than D1 used by the Examining Division. D6 and D7 were cited by the Board and provide evidence of some concepts defined in the independent claims.

5.2 **Difference**

5.2.1 D5 discloses (wording of claim 1, comments added by the Board, references according to D5)
(A) a computer-implemented method for classification of electronic messages being email messages as spam or legitimate on a client mail server (*paragraphs [0004], [0007]*), the method comprising:
(B) receiving, by a computer processor (*MADC 116, Figs. 1 and 3*) of the client mail server, all electronic messages directed to different users who have email addresses registered on the client mail server *(paragraphs [0004], [0007]);*
(C) classifying the received electronic messages as legitimate or spam messages *(paragraphs [0002], [0009], [0108], [0110]*);
(D) identifying unknown messages that could not be classified as legitimate or spam *(paragraphs [0009], [0010], [0108], [0110]);*
(E) obtaining metadata of the unknown messages (*paragraph [0179]*),
(F) the metadata including at least a set of hash ~~sums~~ values for an unknown message ("*mobile message content*") and an IP address (*paragraph [0106]* mentions *an "originating address"* in an *"IP based network"* and therefore an *"IP address"*) of a sender of the unknown message;
(L) classifying (*paragraph [0108]*) unknown messages as legitimate or spam based on message ratings (*abstract, paragraph [0183] discloses different kind of filter criteria, e.g. messages suitable for a certain age or an address-specific filter in order to classify messages into legitimate and illegitimate messages*).

5.2.2 D5 discloses in paragraph [0020], [0232] weighting and in paragraph [0183] rating the content of a message according to a content rating parameter value (classifying into legitimate and illegitimate messages, illegitimate messages are filtered out). D5 does not disclose the combination of Features (F) (part) to (K) relating to the algorithm of the spam filter.

5.3 **Technical effect**

5.3.1 The Examining Division argued that the subject-matter of Claim 1 defined essentially a method for classifying electronic messages in terms of a computer-implemented mathematical algorithm. With reference to T 1784/06, reasons 3.1.1, it was argued that a mathematical algorithm contributed to the technical character of a computer implemented method only in so far as it served a technical purpose. In the present case, the algorithm served the general purpose of classifying emails as legitimate or spam. Classification of emails was certainly useful for locating emails with a relevant cognitive content, but in the Examining Division's view it did not qualify as a technical purpose.

5.3.2 Reference was also made to T 1316/09, reasons 2, where it was held that methods of text classification *per se* did not produce a relevant technical effect or provided a technical solution to any technical problem.

5.3.3 The Examining Division further determined whether the algorithm provided a technical contribution. The Examining Division argued that according to T 0258/03, reasons 5.8, an algorithm might be considered to provide a technical contribution to the invention, if it was particularly suitable for being performed on a computer in that its design was motivated by technical considerations of the internal functioning of the computer. However, following G 3/08, reasons 13.5 and

13.5.1, such technical considerations had to go beyond merely finding a computer algorithm to carry out some procedure.

5.3.4 The Examining Division considered that no such technical considerations were present. The algorithm underlying the method of claim 1 did not go beyond a particular mathematical formulation of the task of classifying electronic messages. The aim of this formulation was to enable a computer to carry out this task, but no further consideration of the internal functioning of a computer could be recognised.

5.3.5 The Examining Division did not contest that the claimed classification method might provide more reliable results, but this was an inherent property of deterministic algorithms. The mere fact that an algorithm lead to reproducible results did not imply that it made a technical contribution. Since the mathematical algorithm did not contribute to the technical character of the claimed method, an inventive step could be present only in its technical implementation. The only implementation features specified in the claim were references to the method being "computer-implemented", i.e. being executed by a computer processor and the text documents being "electronic messages".

5.3.6 The Appellant argued that in the proposed algorithm rather than processing all received electronic messages on each individual user computing device, the invention used a client mail server. Design and implementation of innovative algorithms and data structures went beyond a particular mathematical formulation of the task of classifying electronic messages. The invention utilised a cluster rating system to obtain and analyse metadata of the unknown messages and classify these unknown messages using a cluster index tree data structure. The cluster rating was dynamically changing during the course of the filling of all clusters with various metadata of incoming electronic messages. The invention analysed mass messages sent by various sources over time and updated cluster ratings and therefore allowed to classify unknown messages more accurately. The method therefore facilitated an improved performance of a computer and computer network. Therefore, the method achieved the technical effects mentioned in section VIII above.

5.3.7 The Board agrees in so far with the arguments of the Appellant as the combination of comparing digital text content by similarity preserving hashing and dynamic cluster rating may be considered an algorithm optimised for the computer hardware and may have a technical contribution. However, this mere assumption is not sufficient. The Board is of the opinion that

(a) the implementation of an algorithm in a method for filtering spam messages must have a proved further technical effect or specific technical considerations;

(b) such further technical effect must be specifically and sufficiently documented in the disclosure of the invention and be reflected in the claim wording;

(c) the algorithm must serve a technical purpose.

ad (a)

5.3.8 Case law: Formulating an algorithm is a cognitive exercise (see G 1/19, reasons 112). The definition of an algorithm does not necessarily involve technical considerations (see G 3/08, Reasons 13.5.1). According to T 1358/09 (reasons 5.2 to 5.7) an algorithm may be particularly suitable to be run on a computer in that its design was motivated by technical considerations relating to the internal functioning of the computer. It was further concluded that not all efficiency aspects of an algorithm are by definition without relevance for the question of whether the algorithm provides a technical contribution.

5.3.9 In G 1/19, reasons 115, it was confirmed that a computer software - including the underlying algorithms - may contribute to the technical character of a computer-implemented invention in that it is adapted to the internal functioning of the computer or computer system/network.

5.3.10 Decisions T 0022/12 (reasons 2.2), T 1849/17 (reasons 9 to 9.3), T 1028/06 (reasons 9, 10), T 1179/14 (reasons 5.1), and decision 20W (pat) 13/09 (reasons II, 3.2 a) and c)) of the German Federal Patent Court (*Bundespatentgericht)* address spam filters and consider the implementation of a spam filter algorithm to be an administrative act or underline the importance of a further technical effect / technical considerations. However, in T 1028/14 (reasons 1.1.2 to 1.1.4) the features of an algorithm for identifying a message as an undesired message were considered to be technical and a supplementary search for these features was ordered.

5.3.11 The <u>present invention</u>, page 19, lines 1 to 8, discloses that the cluster size has to be optimised in order to reduce the load on the computer. However, further details are not provided, for example, the range of the optimal cluster size, relevant parameters, the amount of memory saved or the ratio of increased speed.

5.3.12 The Board is of the opinion that specific details as to how an algorithm is implemented in practice and how the load is reduced must be provided in order to give evidence that the algorithm has any further technical effect with respect to known algorithms and that it provides an improvement over the prior art.

ad (b)

5.3.13 Case law: According to T 0154/04, reasons 5, under (E) and (F), for examining patentability of an invention in respect of a claim, the claim must be construed to determine the technical features of the invention, i.e. the features which contribute to the technical character of the invention.

5.3.14 However, the present invention does not provide sufficient and specific disclosure, such as parameters, how the algorithm is optimised for the computer, nor is this reflected in the claim wording.

5.3.15 The Board is of the opinion that any further technical effect has to be specified and sufficiently disclosed in the invention and that the claims must comprise the specific features which contribute to the further technical effect of the invention.

ad (c)

5.3.16 Case law: In T 1358/09, reasons 5, it was decided that whether two electronic messages in respect of their textual content belonged to the same "class" of documents (spam or legitimate) was not a technical issue. Furthermore, it was decided that algorithms contribute to the technical character of a computer-implemented method for classifying text documents only if they serve a technical purpose.

5.3.17 The present invention, on page 21, lines 24 and 25, discloses that a human malware expert may be employed to make conclusions as to whether messages are spam before placing the metadata of the messages into clusters. Such conclusions influence the forming of the rating of the clusters. The rating and clustering is an essential feature of the present invention.

5.3.18 The Board is of the opinion that, if the algorithm claimed in the present invention depends on the preferences of the user, the purpose of the claimed method may be considered non-technical. Spam filters in general need a human conditioning for training an algorithm which kind of emails should to be classified as spam messages and which kind of emails should be kept in the mail box. This depends on the individual user preferences, because some users want to receive certain types of advertising or emails that other users would consider spam. The parameters and structure of the algorithm may have to be adapted to the user preferences. Therefore, a spam filter algorithm cannot have a purely technical purpose if the classification depends on the personal preference of a user. As held in T 1670/07, if the chain of a technical process is broken by the intervention of a user who trains the algorithm, the whole spam filter process may be considered non-technical (cf. "broken technical chain fallacy", T 1670/07, reasons 11, referring to T 1741/08).

5.3.19 *Antivirus* software has to be distinguished from filtering undesired spam messages, because *antivirus* software does not depend on the preferences of a user. Furthermore, a computer virus has a direct impact on the computer hardware. T 0340/14 and T 2535/16, e.g., concern *antivirus* software and do not challenge the technicity of a method related to detecting malware.

5.3.20 **In summary**, the Board is of the opinion that any technical effects going beyond merely finding and implementing an algorithm to carry out the algorithm on a computer are not sufficiently documented in the application and are not reflected in the claim wording.

5.4 **Problem**

5.4.1 The problem could thus be formulated as improving the method of classifying email content disclosed in D5 such that spam emails are filtered out efficiently and therefore the load on the computer systems is further decreased.

5.4.2 According to G 1/19 (reasons 121) algorithms first of all define (non-technical) constraints to be considered in the context of the COMVIK approach (T 0641/00). Depending on whether they contribute to any technical effect achieved by the claimed invention, they may or may not in fact be taken into account in the inventive step assessment.

5.5 **Obviousness**

5.5.1 As discussed above a special technical effect related to the algorithm as defined in the independent claim is not sufficiently documented in the present application and is not reflected in the claim wording. The claim therefore defines a mere implementation of an algorithm without any additional special technical effect. This mere implementation of an algorithm cannot be considered involving an inventive step.

5.5.2 In addition, the algorithm of the present application is not inventive over the disclosure of D5 in combination with the teachings of D6 and D7. As discussed above the difference with respect to D5 is a combination of clustering, weighting hashes, summing up weights and rating according to the sum of weights. These differing features are taught by D6 and D7:

5.5.3 D5 itself discloses creating metadata by hashing (item 306 in Fig. 3, paragraph [0114] ff). It is an obvious option that the hash value of D5 is a hash sum. D5 fails to teach normalisation by summing up the weights and clustering. Clustering is done in the present application for determining a degree of similarity.

5.5.4 D6 teaches (paragraphs [0016], [0033], [0077], [0078]) weighting using the IP address and "smoothing" by summing up the weights. Clustering in the context of spam filters is a known option in the art. D7 teaches (paragraphs [0045] to [0054], Figs. 4 to 7) dynamic clustering for assessing the similarity of text in emails. It is obvious that the components of a cluster have to be indexed. In view of the problem to be solved the skilled person would apply the concepts taught in D6 (smoothing the weights) and D7 (clustering) without any technical difficulties to the algorithm disclosed in D5.

5.5.5 **In summary**, the Board is of the opinion that Features (F) (part) to (K) are obvious techniques in the field of spam filters / classifying emails. These features represent a combination of well-known options and concepts which the skilled person would consider in view of the problem to be solved. Different known concepts of spam filters or methods for classifying documents are combined to a novel spam filtering algorithm, which however does not produce any additional unexpected technical effect. Consequently, the subject-matter of claim 1 does not involve an inventive step within the meaning of Article 56 EPC.

6. Auxiliary Request - inventive step

6.1 Effect and problem

Additional Feature (M) relates to filtering out emails from the server and not only in a user application. This reduces further the load on the network and computer resources.

6.2 Obviousness

6.2.1 D5 (paragraphs [0040], [0108]; the Message Abuse Detector Component (MADC) is part of the core network and hence has to be located on a server) and D7 (Fig. 2 discloses that an "E-mail server program" filters and classifies emails) teach that emails rated as spam are filtered out in the email server. Reporting the rating to the client mail server which makes a decision based on the received rating is therefore obvious to the skilled person. Therefore, the subject-matter of claim 1 of the Auxiliary Request is not inventive within the meaning of Article 56 EPC.

6.2.2 **To summarise**, since none of the requests on file meets the requirements of the EPC, the Examining Division's decision refusing the application is confirmed. Consequently the appeal has to be dismissed.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chairman:

Beschwerdekammern des Europäischen Patentamts
Boards of Appeal of the European Patent Office
Chambres de recours de l'Office européen des brevets

S. Sánchez Chiquero G. Eliasson

Decision electronically authenticated