(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification:
G06F 15/16 (2006.01)

(21) International Application Number:
PCT/US2006/023425

(22) International Filing Date: 15 June 2006 (15.06.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/152,890   15 June 2005 (15.06.2005)   US

(71) Applicant (for all designated States except US): INTER-NATIONAL BUSINESS MACHINES CORPORA-TION [US/US]; New Orchard Road, Armonk, New York 10504 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): GEORGE, David [US/US]; 1 Arthur Court, Somers, New York 10589 (US). JENNINGS, Raymond B., III [US/US]; 31 Barnes Road, Ossining, New York 10562 (US). LAVOIE, Jason D. [US/US]; 75 Chestnut Ridge Road, Mahopac, New York 10541 (US).

(74) Agent: TONG, Kin-Wah; Patterson & Sheridan, L.L.P., 595 Shrewsbury Avenue, Shrewsbury, New Jersey 07702 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): 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, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
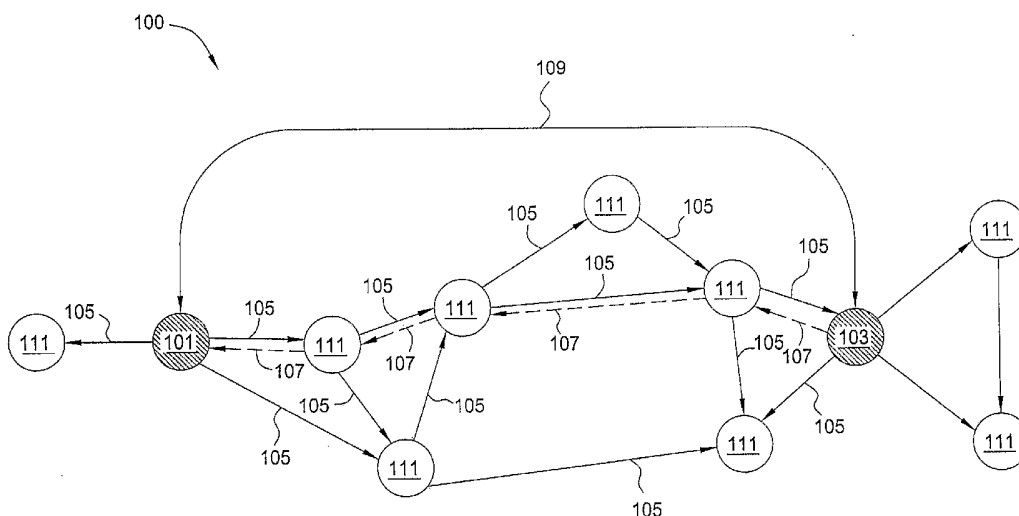
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (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: METHOD AND APPARATUS FOR REDUCING SPAM ON PEER-TO-PEER NETWORKS

(57) Abstract: One embodiment of the present method and apparatus for reducing spam in peer-to-peer networks includes forming a search message including at least one decoy word and sending the search request message, including the decoy word(s), to one or more nodes in the network. Embodiments of the present invention make it possible to weed out nodes in the network that send spam in response to every search message (e.g., regardless of the search message's content).

METHOD AND APPARATUS FOR REDUCING SPAM ON PEER-TO-PEER NETWORKS

BACKGROUND

[0001] The present invention relates generally to computing networks and relates more particularly to the propagation of spam (*e.g.*, unsolicited or spoofed data) over peer-to-peer data transfer networks.

[0002] Figure 1 is a schematic diagram of a network 100 of nodes (*e.g.*, computing devices) interacting in a peer-to-peer (P2P) manner. Generally, a requesting node 101 sends a search message 105 (*e.g.*, containing keywords relating to data that the requesting node 101 wishes to locate) to at least one intermediate node 111 in communication with the requesting node 101 via a peer connection. The intermediate node 111 receives the search message 105 and forwards the search message 105 to at least one additional node 111. Eventually, the search message 105 reaches at least one responding node 103 having the requested data (in some cases, the first intermediate node 111 to which the search message 105 is forwarded will also be a responding node 103). At least one responding node 103 then sends a response message 107 back to the requesting node 101, *e.g.*, via the intermediate nodes 111. The requesting node 101 then requests the relevant data from a responding node 103 by connecting directly to the responding node 103, *e.g.*, via direct connection 109.

[0003] In conventional P2P systems, it has become common for some responding nodes 103 to disguise "spam" content (*e.g.*, unsolicited or spoofed data, such as advertisements) inside of transferred files. Some responding nodes may even send spam content in response to substantially any search message that is received, regardless of the keywords included in the search message. For example, in response to a search request message 105 including the search terms "Joe's poetry", a responding node 103 may indicate that it has a file labeled "Joes_poetry.mp3" or "JoesPoetry.wmv". However, instead of containing content related to Joe's poetry, the file in fact contains an advertisement for a product completely unrelated to Joe or

1

poetry. If the user at the requesting node does not immediately check the content of the transferred file, he or she may be unaware that the transferred file contains spam and may unknowingly share the spam content with other users on the P2P network.

[0004] Thus, there is a need in the art for a method and apparatus for reducing spam on a P2P network.

SUMMARY OF THE INVENTION

[0005] One embodiment of the present method and apparatus for reducing spam in peer-to-peer networks includes forming a search message including at least one decoy word and sending the search request message, including the decoy word(s), to one or more nodes in the network. Embodiments of the present invention make it possible to weed out nodes in the network that send spam in response to every search message (*e.g.*, regardless of the search message's content).

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] So that the manner in which the above recited embodiments of the invention are attained and can be understood in detail, a more particular description of the invention, briefly summarized above, may be obtained by reference to the embodiments thereof which are illustrated in the appended drawings. It is to be noted, however, that the appended drawings illustrate only typical embodiments of this invention and are therefore not to be considered limiting of its scope, for the invention may admit to other equally effective embodiments.

[0007] Figure 1 is a schematic diagram of a network of nodes interacting in a peer-to-peer manner;

[0008] Figure 2 is a flow diagram illustrating one embodiment of a method for reducing spam on a P2P network, according to the present invention;

[0009] Figure 3 is a flow diagram illustrating a second embodiment of a method for reducing spam on a P2P network, according to the present invention; and

[0010] Figure 4 is a high level block diagram of the spam reduction method that is implemented using a general purpose computing device.

[0011] To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

DETAILED DESCRIPTION

[0012] In one embodiment, the present invention is a method and apparatus for reducing spam in P2P networks. Embodiments of the present invention make it possible for a user to identify disguised spam sent in response to search messages before downloading the associated data, so that the user does not download and subsequently potentially share the spam with other users. Embodiments of the present invention are especially effective in identifying nodes that send spam in response to substantially any search message that they receive, regardless of the content of the search message.

[0013] For the purposes of the present invention, the term "node" refers to a computing device such as a personal computing device (e.g., a desk top computer, a laptop computer, a cellular telephone, a personal digital assistant, etc.), a server, a router or the like that is connected to a P2P network.

[0014] Figure 2 is a flow diagram illustrating one embodiment of a method 200 for reducing spam on a P2P network, according to the present invention. Although the invention is described within the context of a P2P network, it will be appreciated by those skilled in the art that the present invention may have utility in substantially any data transfer network in which users are sharing data with unknown parties. The method 200 may be implemented at, for example, a node in a P2P network.

[0015] The method 200 is initialized at step 202 and proceeds to step 204, where the method 200 receives a search message, e.g., from a user of the node that is the source of the search message. The search message comprises, for example, one or more search keywords relating to data that a user wishes to search for in the network. For instance, the search message might comprise the words "Joe Smith" and "Happy

Birthday" to indicate that the user wishes to search for media files related to Joe Smith's "Happy Birthday" album or song.

[0016] In step 206, the method 200 adds at least one "decoy" word to the search message received in step 204. In one embodiment, the decoy word is a natural word that is unrelated to the keywords in the search message. In one embodiment, decoy words are selected arbitrarily from a dictionary for insertion in the search message. In another embodiment, decoy words are fanciful or made up words. For instance, in the example above, the method 200 might add the words "elephant", "giraffe" and "dinosaur" to the search message containing the words "Joe Smith" and "Happy Birthday".

[0017] In one embodiment, the number of decoy words that are added to the search message is predefined. For example, a user may specify that a fixed number of decoy words automatically be added to each search message, or that added decoy words comprise a fixed percentage of the total number of words or characters in the original search message. In another embodiment, the number of decoy words added to the search message is based on the lengths of previous search messages sent by the user (*e.g.*, a number of decoy words may be added so that the length of the search message, including the decoy words, substantially matches the average length of previous search messages).

[0018] In step 208, the method 200 sends the search message, including the decoy words, *e.g.*, in accordance with typical P2P protocols. The method 200 then receives at least one response message responding to the search message in step 210. In one embodiment, the method 200 may specify that response messages stop being accepted after a specified period of time has elapsed or after a specified number of response messages has been received, in order to prevent response messages from being received indefinitely.

[0019] Once the method 200 starts to receive response messages, the method 200 proceeds to step 212 and examines a received response message. Specifically, the method 200 examines the content of the response message. In step 214, the method

200 determines whether the response message contains any of the decoy words that were added to the search message in step 206.

[0020] If the response message includes enough decoy words (*e.g.*, as specified by predefined criteria), the method 200 proceeds to step 216 and categorizes the response message as spam. Alternatively, if the response message does not include enough decoy words, the method 200 proceeds to step 218 and categorizes the response message as a legitimate response. In one embodiment, a response message contains "enough" decoy words to be categorized as spam if the response message includes at least a predefined threshold number of decoy words or if decoy words comprise at least a predefined threshold percentage of the total number of words or characters in the response message. In one embodiment, responses that are categorized as spam are discarded.

[0021] Those skilled in the art will appreciate that categorizations of response messages as spam or legitimate responses are not absolute. That is, an occasional response message that is identified as spam based on the predefined criteria may in fact be a legitimate response. Likewise, an occasional spam message may not include enough decoy words to be identified as spam based on the predefined criteria. Thus, the predefined criteria may be adjusted by the user to tune the sensitivity of the method 200. Moreover, in one embodiment, the method 200 may present the categorization results (*e.g.*, from steps 216 and 218) to the user as a visual display, where the visual display represents an initial filtering of received response messages by the method 200, and the user may review the initial filtering results in order to verify the proper categorization of the received response messages (*e.g.*, in a manner similar to the filtering of electronic mail by many commercial electronic mail providers). For example, the method 200 may present "spam" and "non-spam" response messages in separate windows or panes on the user's display, or may flag response messages as "spam" or "non-spam" (*e.g.*, using one or more icons).

[0022] Once the method 200 has categorized a received response message as either spam or a legitimate response (*e.g.*, in step 216 or 218), the method 200 proceeds to step 220 and determines whether any unexamined response messages remain. If the

method 200 determines that there is at least one unexamined response message, the method 200 returns to step 212 and proceeds as described above to examine and categorize the next response message. Alternatively, if the method 200 determines that no unexamined response messages remain, the method 200 terminates in step 222.

[0023] The method 200 thereby enables a user on a P2P network to quickly identify spammers within the network. By inserting arbitrary decoy words into a user's search message, the method 200 is able to weed out response messages from nodes that respond to substantially any search message they receive, which is a typical approach that spammers take. Moreover, by identifying spammers prior to downloading data from them, a user can avoid not only downloading spam, but sharing the spam with other users who subsequently download data from the user. In addition, if the user is directly connected to the spammer, the user can choose, based on the information gleaned via application of the method 200, to sever the connection to the spammer, thereby potentially isolating the spammer and preventing him or her from distributing any more spam. Similarly, the user may decide simply to ignore all messages sent from the spammer for a predefined period of time. For example, in one embodiment, the method 200 may add a network identifier (*e.g.*, an IP address) of the spammer to a list of known spammer nodes, where all messages received from nodes on the list are to be ignored for a predefined period of time.

[0024] Figure 3 is a flow diagram illustrating a second embodiment of a method 300 for reducing spam on a P2P network, according to the present invention. Namely, the method 300 is tailored to data transfer networks in which search messages are considered to be the logical AND (*e.g.*, only data matching all keywords in the search message will generate a search response). The method 300 may be implemented at, for example, a node in a P2P network.

[0025] The method 300 is initialized at step 302 and proceeds to step 304, where the method 300 constructs a search message using only decoy words. Similar to the decoy words used in the method 200, the decoy words used to construct the search message in step 304 comprise arbitrary, individually un-related natural words, for

example selected from a dictionary. Thus, for example, a search message constructed in step 304 might contain the words "elephant", "orange" and "hockey".

[0026] In step 306, the method 300 sends the search message *e.g.*, in accordance with typical P2P protocols. The method 300 then receives at least one response message responding to the search message in step 308. In one embodiment, the method 300 may specify that response messages stop being accepted after a specified period of time has elapsed or after a specified number of response messages has been received, in order to prevent response messages from being received indefinitely.

[0027] Once the method 300 starts to receive response messages, the method 300 proceeds to step 310 and examines a received response message. Specifically, the method 300 examines the response message to determine the sender of the response message (*e.g.*, as identified by an IP address associated with the response message).

[0028] In step 312, the method 300 categorizes the sender of the response message as a spammer (*e.g.*, a node that is known to send spam). The method 300 may then decide to ignore all messages sent from the sender for a predefined period of time. In one embodiment, the method 300 adds a network identifier (*e.g.*, an IP address) of the sender to a list of known spammer nodes, where all messages received from nodes on the list are to be ignored for a predefined period of time.

[0029] Once the method 300 has categorized the sender of the received response message as a spammer, the method 300 proceeds to step 314 and determines whether any unexamined response messages remain. If the method 300 determines that there is at least one unexamined response message, the method 300 returns to step 310 and proceeds as described above to examine and categorize the sender of the next response message. Alternatively, if the method 300 determines that no unexamined response messages remain, the method 300 terminates in step 316.

[0030] Figure 4 is a high level block diagram of the spam reduction method that is implemented using a general purpose computing device 400. In one embodiment, a general purpose computing device 400 comprises a processor 402, a memory 404, a spam reduction module 405 and various input/output (I/O) devices 406 such as a

display, a keyboard, a mouse, a modem, and the like. In one embodiment, at least one I/O device is a storage device (*e.g.*, a disk drive, an optical disk drive, a floppy disk drive). It should be understood that the spam reduction module 405 can be implemented as a physical device or subsystem that is coupled to a processor through a communication channel.

[0031] Alternatively, the spam reduction module 405 can be represented by one or more software applications (or even a combination of software and hardware, *e.g.*, using Application Specific Integrated Circuits (ASIC)), where the software is loaded from a storage medium (*e.g.*, I/O devices 406) and operated by the processor 402 in the memory 404 of the general purpose computing device 400. Thus, in one embodiment, the spam reduction module 405 for reducing spam in a P2P network described herein with reference to the preceding Figures can be stored on a computer readable medium or carrier (*e.g.*, RAM, magnetic or optical drive or diskette, and the like).

[0032] Thus, the present invention represents a significant advancement in the field of data transfer networks. A method and apparatus are provided that make it possible for a user to identify disguised spam sent in response to search messages before downloading the associated data, so that the user does not download and subsequently potentially share the spam with other users. Thus, the spreading of spam through a network is substantially reduced.

[0033] While foregoing is directed to the preferred embodiment of the present invention, other and further embodiments of the invention may be devised without departing from the basic scope thereof, and the scope thereof is determined by the claims that follow.

CLAIMS

1.   A method for reducing spam in a network, said method comprising:
      forming a search message comprising at least one decoy word; and
      sending said search request message to one or more nodes in said network.

2.   The method of claim 1, wherein said at least one decoy word is an arbitrarily selected natural word.

3.   The method of claim 1, wherein said at least one decoy word is a made up word.

4.   The method of claim 1, wherein said at least one decoy word is unrelated to at least one search keyword contained within said search message.

5.   The method of claim 1, wherein said forming comprises:
      receiving a search message from a user; and
      adding said at least one decoy word to said search message.

6.   The method of claim 5, wherein said at least one decoy word comprises a predefined percentage of a total number of words or characters contained in said received search message.

7.   The method of claim 1, wherein said at least one decoy word comprises a predefined fixed number of decoy words.

8.   The method of claim 1, wherein said at least one decoy word comprises a number of decoy words, said number being based on one or more previous search messages sent by a user.

9.   The method of claim 1, further comprising:

receiving at least one response message in response to said search message; and

determining whether said at least one response message contains enough decoy words to be categorized as spam.

10. The method of claim 9, wherein said at least one response message contains enough decoy words to be categorized as spam if a number of words in said at least one response message that match said at least one decoy word meets or exceeds a predefined threshold number.

11. The method of claim 9, wherein said at least one response message contains enough decoy words to be categorized as spam if at least a predefined threshold percentage of a total number of words or characters in said response message match said at least one decoy word.

12. The method of claim 9, further comprising:

identifying, to a user, response messages that are categorized as spam and response messages that are categorized as legitimate responses.

13. The method of claim 12, wherein said identification is made by presenting said spam messages and said legitimate responses in separate windows on said user's display.

14. The method of claim 12, wherein said identification is made by flagging said spam messages and said legitimate responses with different icons.

15. The method of claim 9, wherein said at least one response message is discarded if said at least one response message is categorized as spam.

16. The method of claim 1, wherein said forming comprises:

constructing said search message from only said at least one decoy word.

17.   The method of claim 1, further comprising:

receiving at least one response message in response to said search message; and

categorizing a sender of said at least one response message as a spammer.

18.   The method of claim 17, wherein said categorizing comprises:

adding said sender to a list of known spammers.

19.   The method of claim 17, wherein said categorizing comprises:

ignoring subsequent messages sent by said sender for a predefined period of time.

20.   A computer readable medium containing an executable program for reducing spam in a network, where the program performs the steps of:

forming a search message comprising at least one decoy word; and

sending said search request message to one or more nodes in said network.

21.   The computer readable medium of claim 20, wherein said at least one decoy word is an arbitrarily selected natural word.

22.   The computer readable medium of claim 20, wherein said at least one decoy word is a made up word.

23.   The computer readable medium of claim 20, wherein said at least one decoy word is unrelated to search keywords contained within said search message.

24.   The computer readable medium of claim 20, wherein said forming comprises:

receiving a search message from a user; and

adding said at least one decoy word to said search message.

25.   The computer readable medium of claim 24, wherein said at least one decoy word comprises a predefined percentage of a total number of words or characters contained in said received search message.

26.   The computer readable medium of claim 20, wherein said at least one decoy word comprises a predefined fixed number of decoy words.

27.   The computer readable medium of claim 20, wherein said at least one decoy word comprises a number of decoy words, said number being based on one or more previous search messages sent by a user.

28.   The computer readable medium of claim 20, further comprising:
          receiving at least one response message in response to said search message; and
          determining whether said at least one response message contains enough decoy words to be categorized as spam.

29.   The computer readable medium of claim 28, wherein said at least one response message contains enough decoy words to be categorized as spam if a number of words in said at least one response message that match said at least one decoy word meets or exceeds a predefined threshold number.

30.   The computer readable medium of claim 28, wherein said at least one response message contains enough decoy words to be categorized as spam if at least a predefined threshold percentage of a total number of words or characters in said response message match said at least one decoy word.

31.   The computer readable medium of claim 28, further comprising:
          identifying, to a user, response messages that are categorized as spam and response messages that are categorized as legitimate responses.

32.   The computer readable medium of claim 20, wherein said forming comprises:
constructing said search message from only said at least one decoy word.

33.   The computer readable medium of claim 20, further comprising:
receiving at least one response message in response to said search message;
and
categorizing a sender of said at least one response message as a spammer.

34.   The computer readable medium of claim 33, wherein said categorizing
comprises:
adding said sender to a list of known spammers.

35.   Apparatus for reducing spam in a network, said apparatus comprising:
means for forming a search message comprising at least one decoy word; and
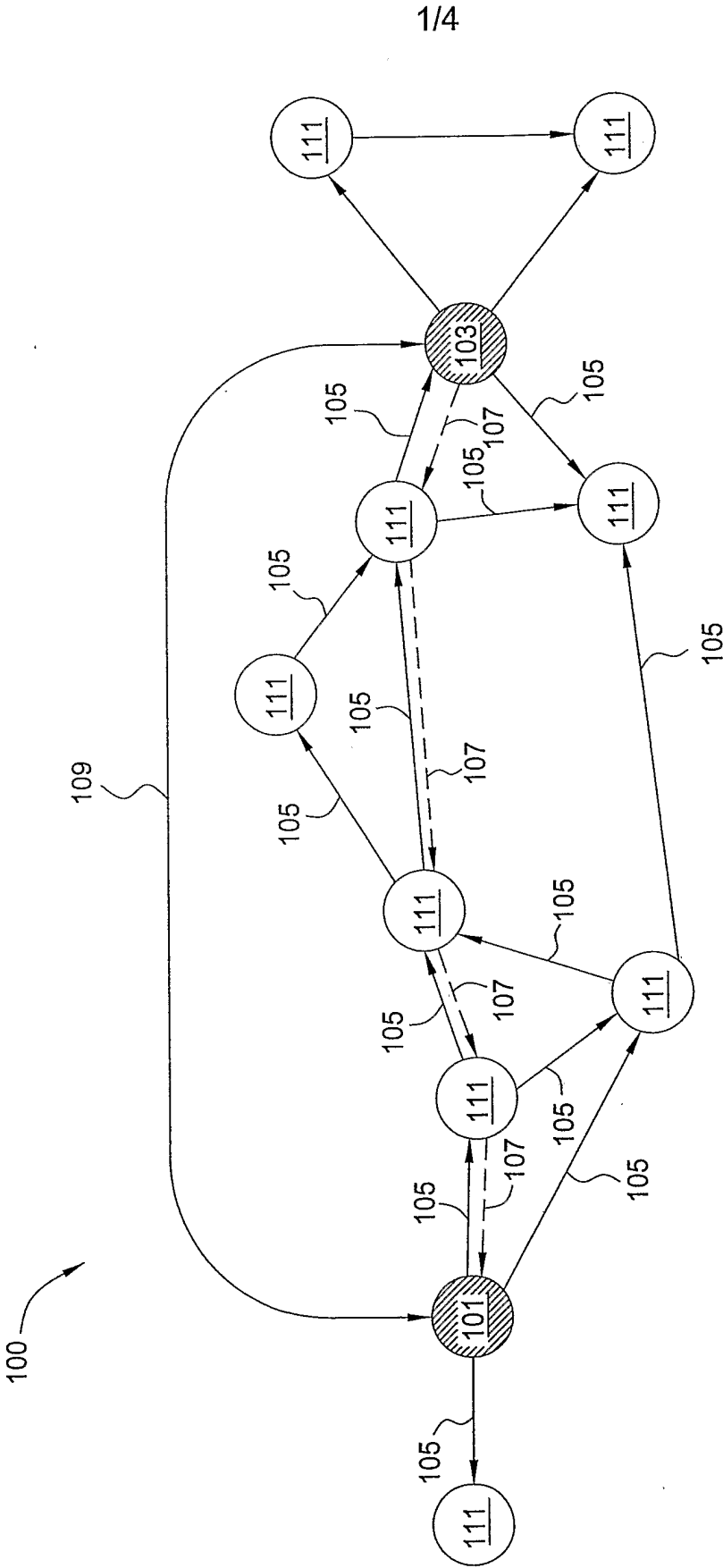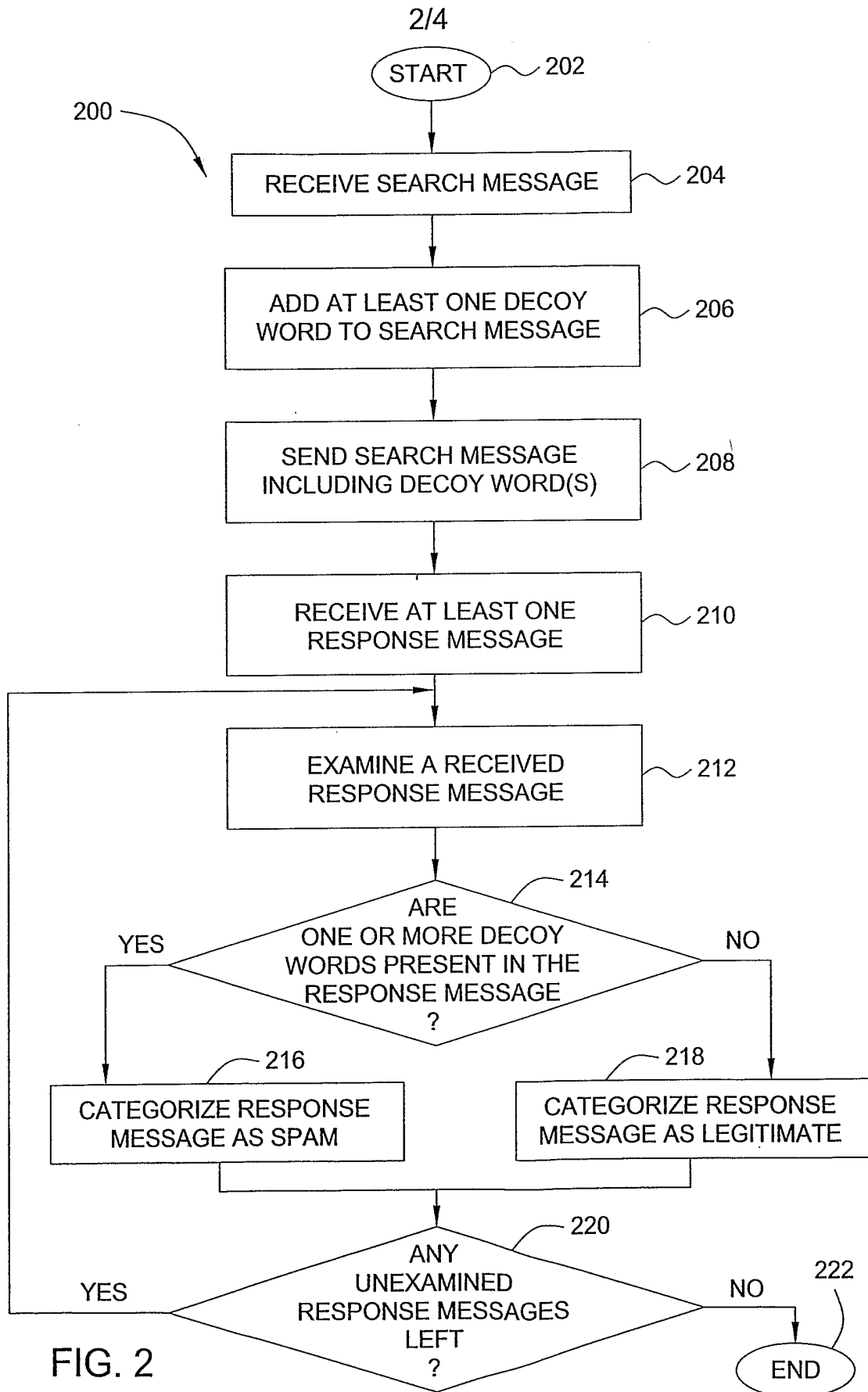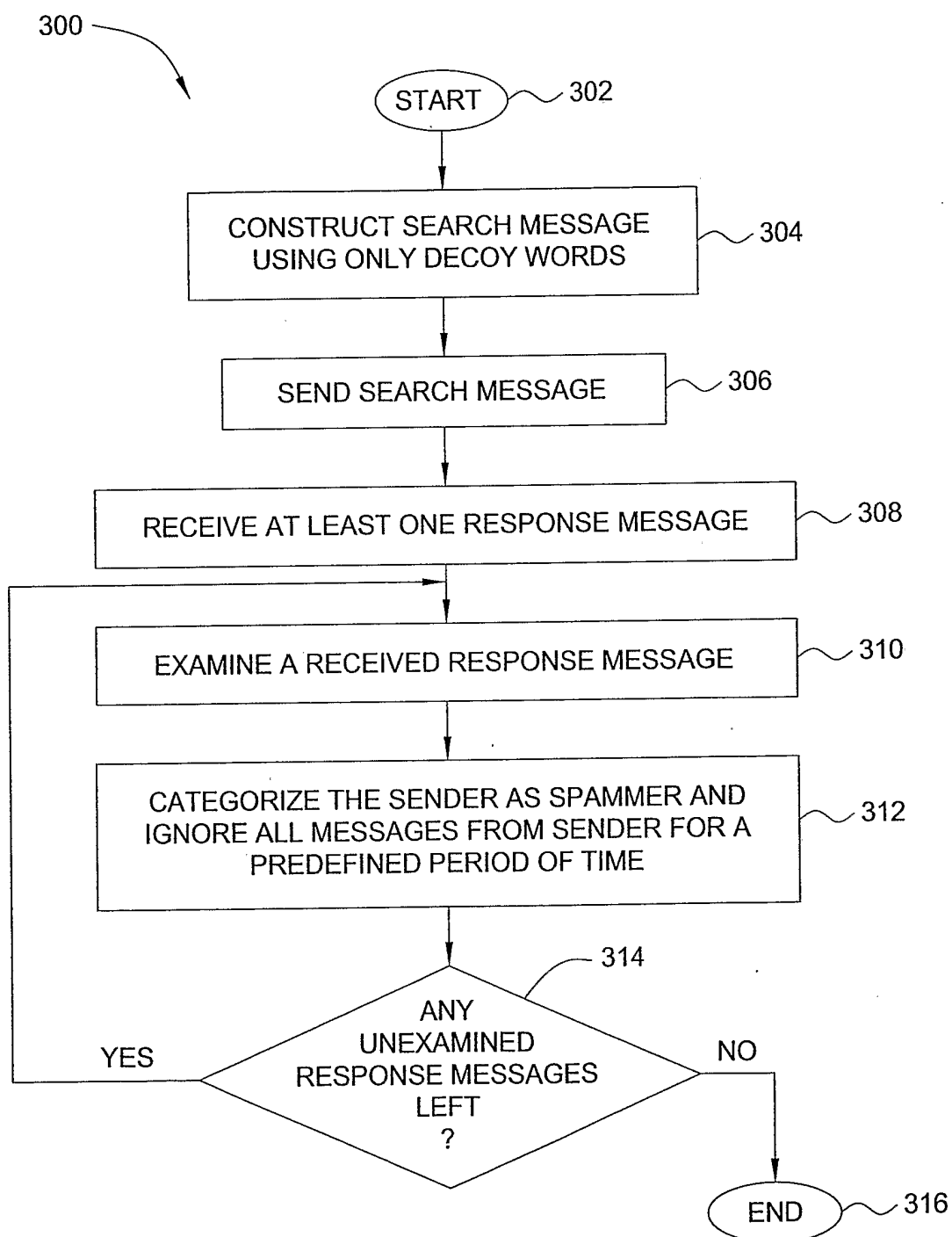means for sending said search request message to one or more nodes in said
network.

1/4

FIG. 1

2/4

```
                              ┌─────────┐
                              │  START  │────╲ 202
                              └─────────┘
                                   │
  200 ──╲                          ▼
           ╲          ┌──────────────────────────┐
                      │  RECEIVE SEARCH MESSAGE   │────╲ 204
                      └──────────────────────────┘
                                   │
                                   ▼
                      ┌──────────────────────────┐
                      │   ADD AT LEAST ONE DECOY  │────╲ 206
                      │   WORD TO SEARCH MESSAGE  │
                      └──────────────────────────┘
                                   │
                                   ▼
                      ┌──────────────────────────┐
                      │   SEND SEARCH MESSAGE     │────╲ 208
                      │   INCLUDING DECOY WORD(S) │
                      └──────────────────────────┘
                                   │
                                   ▼
                      ┌──────────────────────────┐
                      │   RECEIVE AT LEAST ONE    │────╲ 210
                      │   RESPONSE MESSAGE        │
                      └──────────────────────────┘
                                   │
                                   ▼
                      ┌──────────────────────────┐
                      │   EXAMINE A RECEIVED      │────╲ 212
                      │   RESPONSE MESSAGE        │
                      └──────────────────────────┘
                                   │
                                   ▼
                              ╱─ 214
                          ╱    ARE    ╲
              YES    ╱   ONE OR MORE DECOY  ╲    NO
            ┌──────<   WORDS PRESENT IN THE   >──────┐
            │        ╲  RESPONSE MESSAGE   ╱          │
            │            ╲      ?       ╱             │
            ▼                 ╲──╱                    ▼
      ┌─ 216                                    ┌─ 218
   ┌──────────────────┐                  ┌──────────────────────┐
   │ CATEGORIZE RESPONSE │              │ CATEGORIZE RESPONSE    │
   │ MESSAGE AS SPAM     │              │ MESSAGE AS LEGITIMATE  │
   └──────────────────┘                  └──────────────────────┘
            │                                        │
            └────────────────┬───────────────────────┘
                             ▼
                         ╱─ 220
                     ╱    ANY    ╲
       YES      ╱   UNEXAMINED     ╲    NO        222
     ┌────────< RESPONSE MESSAGES   >──────┐   ╱
     │          ╲     LEFT        ╱         ▼  ╱
     │              ╲    ?     ╱      ┌─────────┐
  FIG. 2               ╲──╱          │   END   │
                                     └─────────┘
```

300

302 — START

304 — CONSTRUCT SEARCH MESSAGE USING ONLY DECOY WORDS

306 — SEND SEARCH MESSAGE

308 — RECEIVE AT LEAST ONE RESPONSE MESSAGE

310 — EXAMINE A RECEIVED RESPONSE MESSAGE

312 — CATEGORIZE THE SENDER AS SPAMMER AND IGNORE ALL MESSAGES FROM SENDER FOR A PREDEFINED PERIOD OF TIME

314 — ANY UNEXAMINED RESPONSE MESSAGES LEFT ?

YES

NO

316 — END

FIG. 3

FIG. 4