

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 February 2008 (14.02.2008)

PCT

(10) International Publication Number
WO 2008/019056 A2

(51) International Patent Classification:
H04L 12/66 (2006.01)

(74) Agent: **HUNT, Gregory, A.**; Jenkins, Wilson, Taylor & Hunt, P.A., Suite 1200, University Tower, 3100 Tower Boulevard, Durham, NC 27707 (US).

(21) International Application Number:
PCT/US2007/017329

(22) International Filing Date: 3 August 2007 (03.08.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/835,487 4 August 2006 (04.08.2006) US
11/510,284 25 August 2006 (25.08.2006) US

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(71) Applicant (for all designated States except US): **TEK-ELEC** [US/US]; 5200 Paramount Parkway, Morrisville, NC 27560 (US).

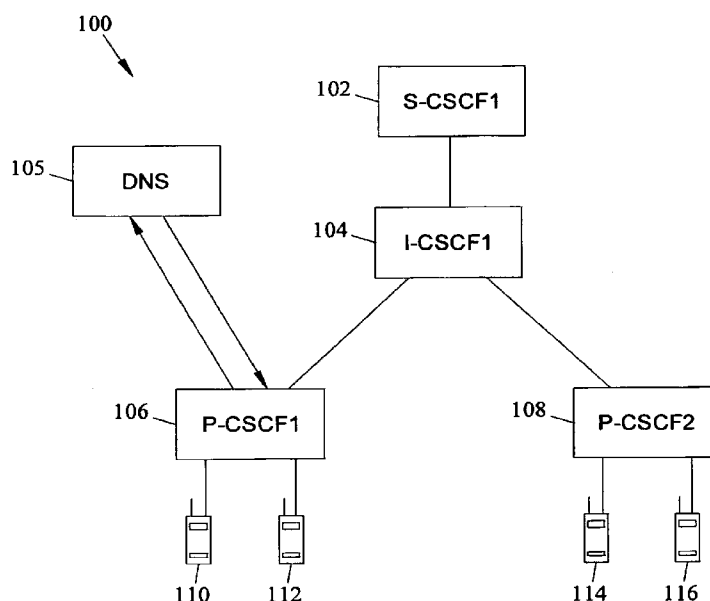
(72) Inventor; and

(75) Inventor/Applicant (for US only): **SPRAGUE, David, Michael** [US/US]; 2316 Heartley Drive, Raleigh, NC 27615 (US).

(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, MT, 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).

[Continued on next page]

(54) Title: METHOD, SYSTEMS, AND COMPUTER PROGRAM PRODUCTS FOR INHIBITING MESSAGE TRAFFIC TO AN UNAVAILABLE TERMINATING SIP SERVER



(57) Abstract: Methods, systems, and computer program products for inhibiting message traffic to an unavailable terminating SIP server are disclosed. According to one method, at least one list of SIP servers accessible by a call session control function (CSCF) is maintained at an originating CSCF. The list includes at least one of availability and unavailability status information for the terminating SIP servers. A request is received to contact a destination via a one of the terminating SIP servers. The list is indexed and a first terminating SIP server is identified. If the first terminating SIP server is determined to be unavailable based on information in the list, a second terminating SIP server is identified using the list.

WO 2008/019056 A2



Published:

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

DESCRIPTION
METHODS, SYSTEMS, AND COMPUTER PROGRAM PRODUCTS FOR
INHIBITING MESSAGE TRAFFIC TO AN
UNAVAILABLE TERMINATING SIP SERVER

5

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/835,487 filed August 4, 2006 and U.S. Patent Application Serial No. 11/510,284 filed August 25, 2006; the disclosures of which are
10 incorporated herein by reference in their entirety.

TECHNICAL FIELD

The subject matter described herein relates to management of message traffic distribution and processing among a plurality of terminating servers.
15 More particularly, the subject matter described herein relates to methods, systems, and computer program products for forwarding or inhibiting message traffic to an unavailable terminating SIP server.

BACKGROUND

20 The IP multimedia subsystem (IMS) is a new type of network through which multimedia communications can be established between users and between users and applications. In the IMS network, session initiation protocol or SIP is the signaling protocol used to establish communication sessions. The entities involved in establishing a communication session in an IMS network
25 include call session control functions (CSCFs) that communicate with each other using the SIP protocol.

One problem with the conventional call setup scenarios in an IMS network occurs when a terminating SIP server or call session control function becomes unavailable. When an originating SIP server or call session control
30 function receives a request to establish a communication, the originating SIP server may attempt to contact a terminating SIP server. If the terminating SIP server is unavailable, the originating SIP server may be required to wait for a timeout period defined by SIP call processing before attempting to contact an

alternate server. The originating SIP server typically does not store state information for terminating SIP servers. As a result, when a second communication setup request arrives at the originating SIP server, the originating SIP server may again attempt to contact the failed terminating SIP
5 server. Repeated attempts to contact a failed terminating SIP server and the associated time out periods waste resources of the originating SIP server and delay call establishment.

Accordingly, in light of these difficulties, there exists a need for methods, systems, and computer program products for inhibiting message traffic to an
10 unavailable terminating SIP server.

SUMMARY

Methods, systems, and computer program products for inhibiting message traffic to an unavailable terminating SIP server are disclosed.
15 According to one method, at least one list of SIP servers accessible by a call session control function (CSCF) is maintained at an originating CSCF. The list includes at least one of availability and unavailability status information for the terminating SIP servers. A request is received to contact a destination accessible via a one of the terminating SIP servers. The list is indexed and a
20 first terminating SIP server is identified. If the first terminating SIP server is determined to be unavailable based on information in the list, a second terminating SIP server is identified using the list.

As used herein, the term "available server" refers to a message server with sufficient operating capacity to accept and process new messages. As
25 used herein, the term "unavailable server" refers to a message server that has either failed or has entered a processing overload state and consequently is not capable of accepting and processing a new message.

As used herein, the term "normal status" refers to the operating status of an available terminating SIP message server.

30 As used herein, the term "congested status" refers to the operating status of an unavailable terminating SIP message server that, while operational, does not have sufficient internal resources to accept and process a new message.

As user herein, the term "failed status" refers to the operating status of an unavailable terminating SIP message server that has failed to properly receive, process, and provide a valid status response in response to a received message.

5 As used herein, the term "S-CSCF" refers to an IMS serving call session control function. The S-CSCF function may maintain a list of registered user equipment, associated status information, and associated capabilities.

10 As user herein, the term "I-CSCF" refers to an IMS interrogating call session control function. The I-CSCF function may provide a terminating subscriber location function for a message received from a P-CSCF server and forward the received message to either a S-CSCF server in the network or to an message server in an adjacent network through a network interconnect function.

15 As used herein, the term "P-CSCF" refers to an IMS proxy call session control function. The P-CSCF function may process a message received from a SIP device and forward the processed message to an I-CSCF server in the network. Processing the message may include compressing and/or encrypting the received message.

20 As used herein, the term "CSCF" refers to an IMS call session control function that implements any one or more of the above referenced S-CSCF, I-CSCF, or P-CSCF functions.

25 As used herein, the term "server list" refers to a list of terminating SIP message servers to which an originating SIP message server may forward message traffic. The server list may be stored at the originating SIP message server in any format suitable to the originating server.

30 As used herein, the term "server exception list" refers to a list including one or more unavailable terminating SIP message servers, as detected by an originating SIP message server in response to forwarding a first message. The server exception list may be stored at the originating SIP message server in any format suitable to the originating server.

As used herein, the term "originating SIP server" refers to any SIP server capable of performing an IMS function. An originating SIP server may be an S-CSCF, an I-CSCF, or a P-CSCF.

As used herein, the term "terminating SIP server" refers to any SIP server capable of performing an IMS function, including any of an S-CSCF, an I-CSCF, or a P-CSCF.

5 The subject matter described herein may be implemented using a computer program product comprising computer executable instructions embodied in a computer-readable medium. Exemplary computer-readable media suitable for implementing the subject matter described herein include chip memory devices, disk memory devices, programmable logic devices, application specific integrated circuits, and downloadable electrical signals. In
10 addition, a computer-readable medium that implements the subject matter described herein may be located on a single device or computing platform or may be distributed across multiple physical devices and/or computing platforms.

15 BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the subject matter described herein will now be explained with reference to the accompanying drawings of which:

Figure 1 is a block diagram of an exemplary SIP/IMS call processing architecture in which embodiments of the subject matter described herein may
20 be implemented;

Figure 2 is a block diagram of an exemplary P-CSCF server cluster and an exemplary I-CSCF server cluster according to an embodiment of the subject matter described herein;

Figure 3A is a diagram of an exemplary global server list that may be
25 maintained by a DNS server in a SIP/IMS call processing architecture according to an embodiment of the subject matter described herein;

Figure 3B is a diagram of an exemplary server list that may be maintained by a CSCF according to an embodiment of the subject matter described herein;

30 Figure 3C is a diagram of an exemplary server exception list that may be maintained by a CSCF according to an embodiment of the subject matter described herein;

Figure 4 is a flow chart of an exemplary process for inhibiting message traffic to an unavailable terminating SIP message server according to an embodiment of the subject matter described herein;

Figure 5 is a block diagram illustrating an exemplary method for inhibiting routing of messages to unavailable terminating SIP servers using a server exception list according embodiment of the subject matter described herein;

Figures 6 and 7 are flow charts illustrating exemplary steps for maintaining a server exception list according to an embodiment of the subject matter described herein;

Figure 8 is a block diagram illustrating an exemplary method for load balancing messages among terminating SIP servers using DNS according to an embodiment of the subject matter described herein; and

Figure 9 is a block diagram illustrating an exemplary P-CSCF server that includes a terminating SIP server database according to an embodiment of the subject matter described herein.

DETAILED DESCRIPTION

In view of the problems described above, the subject matter described herein provides methods for forwarding or inhibiting message traffic to a terminating SIP server. Figure 1 illustrates an SIP/IMS call processing architecture 100 in which embodiments of the subject matter described herein may be implemented. In Figure 1, SIP/IMS call processing architecture 100 includes a plurality of SIP/IMS call control elements, including a serving call session control function (S-CSCF) server cluster S-CSCF1 102, an interrogating call session control function (I-CSCF) server cluster I-CSCF1 104, a DNS server 105 and a plurality of proxy call session control function (P-CSCF) server clusters P-CSCF1 106 and P-CSCF2 108. P-CSCF1 106 and P-CSCF2 108 may provide control interface and message transfer operations for a plurality of SIP devices 110 - 116. SIP devices may be any user equipment capable of establishing multimedia sessions using SIP, including mobile and fixed terminals.

In order to establish a communication, an originating SIP device **110** may send a request message to P-CSCF1 **106**. P-CSCF1 **106** may query DNS server **105** to identify the IP address corresponding to an I-CSCF, which represents the point of contact into the destination subscriber's network. DNS server **105** may return the identifier or IP address of more than one I-CSCF server in an I-CSCF cluster. Rather than blindly contacting each server until an available server is located, a P-CSCF server according to an embodiment of the subject matter described herein may store at least one of availability and unavailability information for SIP servers that it is capable of contacting and may use this information to inhibit the forwarding of message traffic to unavailable servers.

Figure 2 illustrates an exemplary P-CSCF server cluster **106** and an exemplary I-CSCF server cluster **104** according to an embodiment of the subject matter described herein. In Figure 2, P-CSCF1 server cluster **106** may include a network identifier **202** and a plurality of P-CSCF servers P1 **204**, P2 **206**, P3 **208**, and P4 **210**. Each server may be operable to implement P-CSCF functions for SIP/IMS call processing architecture **100**, including message compression and/or message security encryption.

P-CSCF1 server cluster **106** may include one or more servers **204**, **206**, **208**, and **210** that are operable to perform IMS proxy CSCF functions, such as serving as the point of contact for user equipment in an IMS network. A server in P-CSCF1 server cluster **106** not designated as active may be placed in a standby state, ready to process one or more SIP messages in the future. For example, servers P1 **204**, P2 **206**, and P3 **208** may be configured as active, with server P4 **208** configured in a standby state. P-CSCF1 server cluster **106** may transition server P4 **210** to the active state in response to one of the active servers P1 **204**, P2 **206**, or P3 **208** becomes unavailable.

I-CSCF1 server cluster **104** may include a network identifier **212** and a plurality of servers I1 **214**, I2 **216**, and I3 **218**. Each server may be operable to implement I-CSCF functions for SIP/IMS call processing architecture **100**, including providing a terminating subscriber location function.

Servers **214**, **216**, and **218** may be active and operable to process SIP messages received from either P-CSCF1 server cluster **106** or S-CSCF1 server

cluster **102**. A server in I-CSCF1 server cluster **104** not designated to be active may be placed in a standby state, ready to process one or more SIP messages in the future. I-CSCF1 server cluster **104** may distribute message processing tasks among the active servers in the cluster using a suitable load balancing
5 algorithm.

Figure 3A illustrates an exemplary global server list **300** that may be maintained by DNS server **105** for the servers illustrated in Figure 2. In Figure 3A, global server list **300** may include a list of server clusters in SIP/IMS call processing architecture **100** with a list of servers configured in each cluster.
10 For example, list **300** may include one or more P-CSCF server clusters, one or more I-CSCF server clusters, and one or more S-CSCF server clusters, each identified with a suitable domain name identifier. A copy of global list **300** may be stored at each server cluster and/or in a central server in SIP/IMS call processing architecture **100**, using any format compatible with the server
15 storing the table.

According to another aspect, each P-CSCF server illustrated in Figure 1 may maintain a list of terminating SIP servers through which call destinations may be reached. Figure 3B illustrates an exemplary terminating server list **302** that may be maintained by a P-CSCF according to an embodiment of the
20 subject matter described herein. In Figure 3B, server list **302** may include an entry for each terminating server that the originating server is capable of accessing in SIP/IMS call processing architecture **100** and may further include an IP address or other network identifier suitable to SIP/IMS call processing architecture **100** for each terminating server in the server list. For example, P-
25 CSCF server P1 **204** may store a server list **302** including I-CSCF servers I1 **214**, I2 **216**, and I3 **218** in I-CSCF1 server cluster **104**.

According to yet another aspect of the subject matter described herein, each P-CSCF server may maintain an exception list including servers accessible by the P-CSCF server that are currently unavailable. Figure 3C
30 illustrates an exemplary server exception list **304** that may be maintained by a P-CSCF according to an embodiment of the subject matter described herein. In Figure 3C, server exception list **304** may include a list of accessible servers identified in server list **302** that are currently unavailable. Each entry in server

exception list 304 may include an identifier for the unavailable server, a status indication, and a retry interval value. The status indication stored in server exception list 304 may be of any form suitable to the originating SIP message server. For example, the status indication may be SIP status response code 503, representing a congested terminating server, or SIP status response code 408, representing a failed terminating server. The retry interval may be a time period after which an originating server can reattempt to contact the failed terminating server. After the retry interval, the entry may age out of server exception list 304. Server exception list 304 may be stored in a database maintained by the originating P-CSCF.

Figure 4 is a flow chart illustrating an exemplary process for inhibiting message traffic to an unavailable terminating SIP server using a server exception list according to an embodiment of the subject matter described herein. Referring to Figure 4, in step 400, at least one list of terminating SIP servers accessible by a P-CSCF is maintained. The list indicates at least one of availability and unavailability information for each server in the list. Step 400 may be performed by a P-CSCF server, such as P-CSCF server P1 204 illustrated in Figure 2. In response to call setup requests during normal call processing, P-CSCF server P1 204 may send SIP request messages to each terminating server, such as an I-CSCF, that it is capable of contacting. Exemplary SIP request messages include an INVITE message, an ACK message, a BYE message, a CANCEL message, an OPTIONS message, and a REGISTER message. If the destination server responds with a SIP 503 or SIP 408 retry message, P-CSCF server P1 204 may determine that the server is unavailable and may mark the status of the server as unavailable in its exception list. As stated above, entries may age out of exception list based on the retry interval associated with an exception list. Once an entry ages out of the exception list, the corresponding server can be retried for normal communications. The request messages used to build the exception list may be generated in response to normal communications between originating and terminating parties. In an alternate implementation, each terminating SIP server may communicate heartbeat messages with the P-CSCF at predetermined time intervals. Failure to receive a heartbeat message within a

predetermined time interval may cause the P-CSCF mark a terminating SIP server as unavailable.

In step **402**, the originating P-CSCF server receives a request to contact a destination accessible via one or more destination SIP servers. In step **404**,
5 P-CSCF P1 **204** indexes the SIP server list and locates a destination SIP server to which the communication should be directed. Step **404** may be performed using any suitable indexing method. In one example, step **404** may include randomly indexing a SIP server list corresponding to a destination. In an alternate implementation, SIP servers may be contacted according to a
10 predetermined order of preference, as will be described in more detail below.

In step **406**, P-CSCF P1 **204** determines whether the server is available. Step **406** may be accomplished by reading the status information for the SIP server in the list. If the server is not available, control proceeds to step **408** where a new index is computed, and steps **404** and **406** are repeated to
15 contact the new server. If the server is available, control proceeds to step **410** where the message is routed to the server.

Thus, in the example illustrated in Figure 4, rather than blindly sending messages to unavailable SIP servers and waiting for timeouts, server availability or unavailability status is stored and used to immediately determine
20 whether a proposed terminating server is unavailable. If the terminating server is unavailable, an alternate server can be selected without requiring communication with the unavailable server. These steps are believed to significantly reduce connection establishment time over implementations where a communication timeout is required to trigger reindexing of a SIP server list.

25 Figure 5 is a network diagram illustrating a message routing example according to an embodiment of the subject matter described herein. Referring to Figure 5, P-CSCF server P1 **204** receives a SIP request message. The SIP request message may originate from another domain or from a terminal associated with P-CSCS P1 **204**. The SIP request message may include the
30 domain EAST-POP.I-CSCF.OPERATOR-A.COM. In this example, it is assumed that server **204** does not cache the corresponding IP address information. Accordingly, server **204** queries DNS server **105** to resolve the

domain name. DNS server **105** resolves the domain name into IP addresses for I-CSCF servers **I1 214**, **I2 216**, and **I3 218**.

In this example, it is assumed that server **204** did not previously store status information for servers **I1 214**, **I2 216**, and **I3 218**. Accordingly, server **P1 204** sends request messages to servers **I1 214**, **I2 216**, and **I3 218** to determine their respective statuses. Server **I1 214** is unavailable and does not respond. Server **I2 216** responds with a retry message indicating that it is congested. Server **I3 218** responds with a 200 OK message indicating that it is available. Accordingly, server **P1 204** updates exception list **304** to indicate the unavailable of servers **I1 214** and **I2 216**.

Server **P1 204** then computes an index into its server list, which includes servers **I1 214**, server **I2 216**, and server **I3 218**. In this example, it is assumed that the first index corresponds to server **I1 214**. Server **P1 204** then determines whether terminating server **I1 214** is in the exception list. Because server **I1 214** is in the exception list, server **P1 214** may refrain from sending message traffic to server **I1 214** during its retry interval and compute a new index. In this example, it is assumed that the new index corresponds to server **I3 218**. Since server **I3 218** is not in the exception list, the SIP request message gets routed to server **I3 218**, as indicated by message 1i in Figure 5.

Thus, as illustrated in Figure 5, server unavailability information may be used to reduce the time for completing communications when one or more servers are unavailable. In the example illustrated in Figure 5, the availability status information of the destination servers is determined in response to a request message received by an originating server. As stated above, in an alternate example, the originating server may store this information in advance of receiving a request, eliminating the need for determining availability status before completing a call. In addition, the timeout period for non-receipt of a response to a status request may be set to less than that of conventional SIP call processing to reduce the time required to determine unavailability of a destination server.

Figure 6 and 7 are flow charts illustrating exemplary processes for maintaining a server exception list according to an embodiment of the subject matter described herein. More particularly, Figure 6 illustrates an exemplary

process by which status information is obtained for a terminating SIP server, and Figure 7 illustrates an exemplary process by which entries are removed from a server exception list according to an embodiment of the subject matter described herein. In Figure 6, it is assumed that an originating server

5 periodically tests terminating servers with which it has contact. Referring to Figure 6, in step 600, a message is sent to a terminating SIP server. The message query may be any suitable SIP message, such as a SIP request message. The request message may be a "dummy" message that prompts the terminating SIP server to return its status. In step 602, it is determined whether

10 a response to the message has been received within a timeout period for the query. The timeout period may be set to less than that of conventional SIP call processing to reduce call processing delay. If the timeout period expires before a response is received, control proceeds to step 604 where the entry is added to the server exception list. Control then proceeds to step 606 where the next

15 server maintained in the originating SIP server is exception list is tested. Steps 600-604 may be repeated for the next server.

In step 602, if a response is received within the timeout period, control proceeds to step 608 where it is determined whether the status code in the message indicates that the destination server is unavailable. If the status code

20 indicates that the server is unavailable, control proceeds to step 604 where the server is added to the exception list. Steps 606-608 may be repeated for the next server.

In step 608, if the status code indicates that the server is available, the server is not added to the exception list. Control then proceeds to step 610

25 where the next server in the server list is determined. Steps 600-608 may then be repeated for the next server.

Referring to Figure 7, a separate process may execute to determine when to remove entries from the exception list. Referring to Figure 7, in step 700, an entry in the exception list is examined. In steps 702 and 704, it is

30 determined whether a retry timeout for the entry has expired. The retry timeout may define the time period during which an exception list entry is valid. Once an entry is no longer valid, it may be removed from the exception list. Accordingly, in step 704, if it is determined that the timeout has expired, control

proceeds to step **706** where the entry is removed from the exception list. In step **708**, the next entry is accessed. Steps **700-706** may be repeated for the next entry.

In step **704**, if it is determined that the retry timeout has not expired,
5 control proceeds to step **710**, where the entry is maintained in the exception list. Control then proceeds to step **712** where the next entry is accessed. Steps **700-710** may be repeated for the next entry. Thus, using the process of Figure 7, entries in the exception list are aged out. When considered in combination with the process illustrated in Figure 6, entries may be added and
10 aged out to maintain a current exception list and to prevent servers from being maintained perpetually in the exception list.

Although the examples described above illustrate maintaining a server list containing all servers and maintaining a separate exception list, the subject matter described herein is not limited to maintaining two separate lists. A
15 single list that includes both available and unavailable servers may be maintained without departing from the scope of the subject matter described herein.

According to another aspect, the subject matter described herein may include a method for using DNS to load balance among terminating SIP
20 servers. Figure 8 illustrates this concept. Referring to Figure 8, a SIP server **204** may initiate a DNS request to resolve the domain name EAST-POP.I-CSCF.OPERATOR-A.COM. In response to the request, DNS server **105** may return IP addresses in the order of I1, I2, I3, representing servers **214**, **216**, and **218** in order of preference. Server **204** may attempt to contact servers **214**,
25 **216**, and **218** in the order specified by DNS server **105**. Accordingly, in the illustrated example, server P1 **204** may contact server I1 **214** first, as illustrated by message 1D.

When server P3 **208** receives a SIP request message for the same domain, server P3 **208** send a DNS request to DNS server **105**. DNS server
30 **105** returns servers I2, I3, and I1 in a different order of preference from that given in response to the previous DNS query by server P1 **204**. Server P3 **208** receives the server contact list and attempts to contact the server of the highest

order of preference in the list. In this example, the server of the highest order of preference is server 12 216.

For each successive DNS request, DNS server 105 may return a different order of priority to the requesting server. Since each requesting server
5 uses the order of priority specified by DNS server 105 to contact a terminating server and the order changes, messages will be load balanced equally over time among the destination servers. If one of the destination servers fails and is added to the exception list of the originating servers, messages will be load balanced among the remaining servers.

10 Figure 9 is a block diagram illustrating an exemplary architecture for P-CSCF 204. Referring to Figure 9, P-CSCF 204 includes a terminating SIP server database 900 for storing the list of available and unavailable SIP servers as described above. P-CSCF 204 may also include a SIP communications routing module 902 for communicating with terminating SIP servers to maintain
15 the availability and/or unavailability status information in database 900. SIP communications routing module 902 may also communicate with DNS server 105 for obtaining IP addresses corresponding to terminating SIP server domain names. SIP communications routing module 902 may also route communications based on information stored in terminating SIP server
20 database 900, as described above.

It will be understood that various details of the subject matter described herein may be changed without departing from the scope of the subject matter described herein. Furthermore, the foregoing description is for the purpose of illustration only, and not for the purpose of limitation, as the subject matter
25 described herein is defined by the claims as set forth hereinafter.

CLAIMS

What is claimed is:

1. A method for forwarding or inhibiting message traffic to a terminating session initiation protocol (SIP) server using a SIP server exception list,
5 the method comprising:
at an originating call session control function (CSCF):
 - (a) maintaining at least one list of terminating SIP servers accessible by the CSCF, the list including at least one of availability and unavailability information for the terminating SIP servers;
 - 10 (b) receiving a request to contact a destination accessible via one of the terminating SIP servers;
 - (c) indexing the list and identifying a first terminating SIP server;
 - (d) determining whether the destination SIP server is available using information stored in the list; and
 - 15 (e) in response to determining that the first terminating SIP server is not available, identifying a second terminating SIP server using the list.
2. The method of claim 1 wherein maintaining at least one list includes maintaining a first list including available and unavailable SIP servers
20 accessible by the CSCF and a second list including unavailable SIP servers accessible by the CSCF.
3. The method of claim 1 wherein maintaining at least one list includes maintaining a single list containing available and unavailable SIP servers accessible the CSCF.
- 25 4. The method of claim 1 wherein maintaining at least one list includes building a list including availability and unavailability status information for the SIP servers based on communications involving the SIP servers.
5. The method of claim 1 wherein maintaining at least one list includes marking a SIP server entry in the list as unavailable in response to
30 failing to receive a response from the corresponding server to a query message to the server within a timeout period, the timeout period being less than that for SIP call processing.

6. The method of claim 1 wherein maintaining at least one list includes sending a query to a SIP server, obtaining a response from the SIP server including status information, and storing the status from the query in the list.
- 5 7. The method of claim 1 wherein maintaining at least one list includes maintaining a retry timer for an entry identifying a server as unavailable and removing unavailability status from the entry in response to expiration of the retry timer.
8. The method of claim 1 wherein indexing a list includes randomly
10 indexing a list.
9. The method of claim 1 wherein indexing a list includes indexing a list according to a predetermined order of preference.
10. The method of claim 1 comprising, in response to the first terminating SIP server being available forwarding a SIP request message to the first
15 terminating SIP server.
11. A method for load balancing communications among terminating SIP servers, the method comprising:
- (a) receiving a first SIP request message including a domain name;
- (b) in response to the first SIP request message, resolving the
20 domain name into a list of terminating SIP servers having a first order of preference for contacting the terminating SIP servers;
- (c) receiving second SIP request message including the domain name;
- (d) in response to the second SIP request message, resolving the
25 domain into a list of the terminating SIP servers having a second order of preference for contacting the terminating SIP servers; and
- (e) forwarding the first and second SIP request messages to terminating SIP servers in accordance with the first and second
30 orders of preference.
12. A system for forwarding or inhibiting message traffic to a terminating session initiation protocol (SIP) server, the system comprising: a call session control function (CSCF), the CSCF including:

- (a) a terminating SIP server database for storing at least one list of terminating SIP servers and at least one of availability and unavailability status information for the terminating SIP servers; and
- 5 (b) a SIP communications routing module for receiving a SIP request message, for accessing the terminating SIP server database, for identifying a first terminating SIP server and status of the first terminating SIP server, and in response to determining that the first terminating SIP server is unavailable, for identifying a second
- 10 terminating SIP server to which the SIP request message should be forwarded.
13. The system of claim 12 wherein the terminating SIP server database includes a first list of available SIP servers and a second list of unavailable SIP servers.
- 15 14. The system of claim 12 wherein the terminating SIP server database includes a single list storing available and unavailable SIP server information.
15. The system of claim 12 wherein the SIP communications routing module is adapted to build the terminating SIP server database based on
- 20 information received from the terminating SIP servers in response to SIP request messages.
16. The system of claim 12 wherein the SIP communications routing module is adapted to mark a SIP server in the terminating SIP server database as unavailable in response to a timeout expiring for a SIP request
- 25 message transmitted by the SIP communications routing module, the timeout being less than a SIP call processing timeout.
17. The system of claim 12 wherein the SIP communications routing module is adapted to mark an entry in the terminating SIP server database as unavailable in response to receiving a message from a terminating SIP
- 30 server indicating unavailability of the terminating SIP server.
18. The system of claim 12 wherein the SIP communications routing module is adapted to maintain a retry interval for SIP servers marked as unavailable in the terminating SIP server database and to remove

unavailability status regarding the SIP servers marked as unavailable in response to expiration of the retry interval for each entry.

19. The system of claim 12 wherein the SIP communications routing module is adapted to randomly index a list of terminating servers in the terminating SIP server database to determine a server to which a message should be forwarded.
20. The system of claim 12 wherein the SIP communications routing module is adapted to access the terminating SIP server database in a predetermined order of preference specified by a domain name system (DNS) server.
21. The system of claim 12 wherein the SIP communications routing module is adapted to route the message to the first terminating SIP server in response to determining availability of the first terminating SIP server using the terminating SIP server database.
22. A computer program product comprising computer executable instructions embodied in a computer readable medium for performing steps comprising:
at an originating call session control function (CSCF):
(a) maintaining at least one list of terminating SIP servers accessible by the CSCF, the list including at least one of availability and unavailability information for the terminating SIP servers;
(b) receiving a request to contact a destination accessible via one of the terminating SIP servers;
(c) indexing the list and identifying a first terminating SIP server;
(d) determining whether the destination SIP server is available using information stored in the list; and
(e) in response to determining that the first terminating SIP server is not available, identifying a second terminating SIP server using the list.
23. The computer program product of claim 22 wherein maintaining at least one list includes maintaining a first list including available and unavailable SIP servers accessible by the CSCF and a second list including unavailable SIP servers accessible by the CSCF.

24. The computer program product of claim 22 wherein maintaining at least one list includes maintaining a single list containing available and unavailable SIP servers accessible the CSCF.
- 5 25. The computer program product of claim 22 wherein maintaining at least one list includes building a list including availability and unavailability status information for the SIP servers based on communications involving the SIP servers.
- 10 26. The computer program product of claim 22 wherein maintaining at least one list includes marking a SIP server entry in the list as unavailable in response to failing to receive a response from the corresponding server to a query message to the server within a timeout period, the timeout period being less than that for SIP call processing.
- 15 27. The computer program product of claim 22 wherein maintaining at least one list includes sending a query to a SIP server, obtaining a response from the SIP server including status information, and storing the status from the query in the list based on the query.
- 20 28. The computer program product of claim 22 wherein maintaining at least one list includes maintaining a retry timer for an entry identifying a server as unavailable and removing unavailability status from the entry in response to expiration of the retry timer.
- 25 29. The computer program product of claim 22 wherein indexing a list includes randomly indexing a list.
- 30 30. The computer program product of claim 22 wherein indexing a list includes indexing a list according to a predetermined order of preference.
31. The computer program product of claim 22 comprising, in response to the first terminating SIP server being available, forwarding a SIP request message to the first terminating SIP server.
32. A computer program product comprising computer executable instructions embodied in a computer readable medium for performing steps comprising:
- (a) receiving a first SIP request message including a domain name;

- (b) in response to the first SIP request message, resolving the domain name into a list of terminating SIP servers having a first order of preference for the terminating SIP servers;
- (c) receiving second SIP request message including the domain name;
- (d) in response to the second SIP request message, resolving the domain into a list of the terminating SIP servers having a second order of preference for the terminating SIP servers; and
- (e) forwarding the first and second SIP request messages to terminating SIP servers in accordance with the first and second orders of preference.

1/9

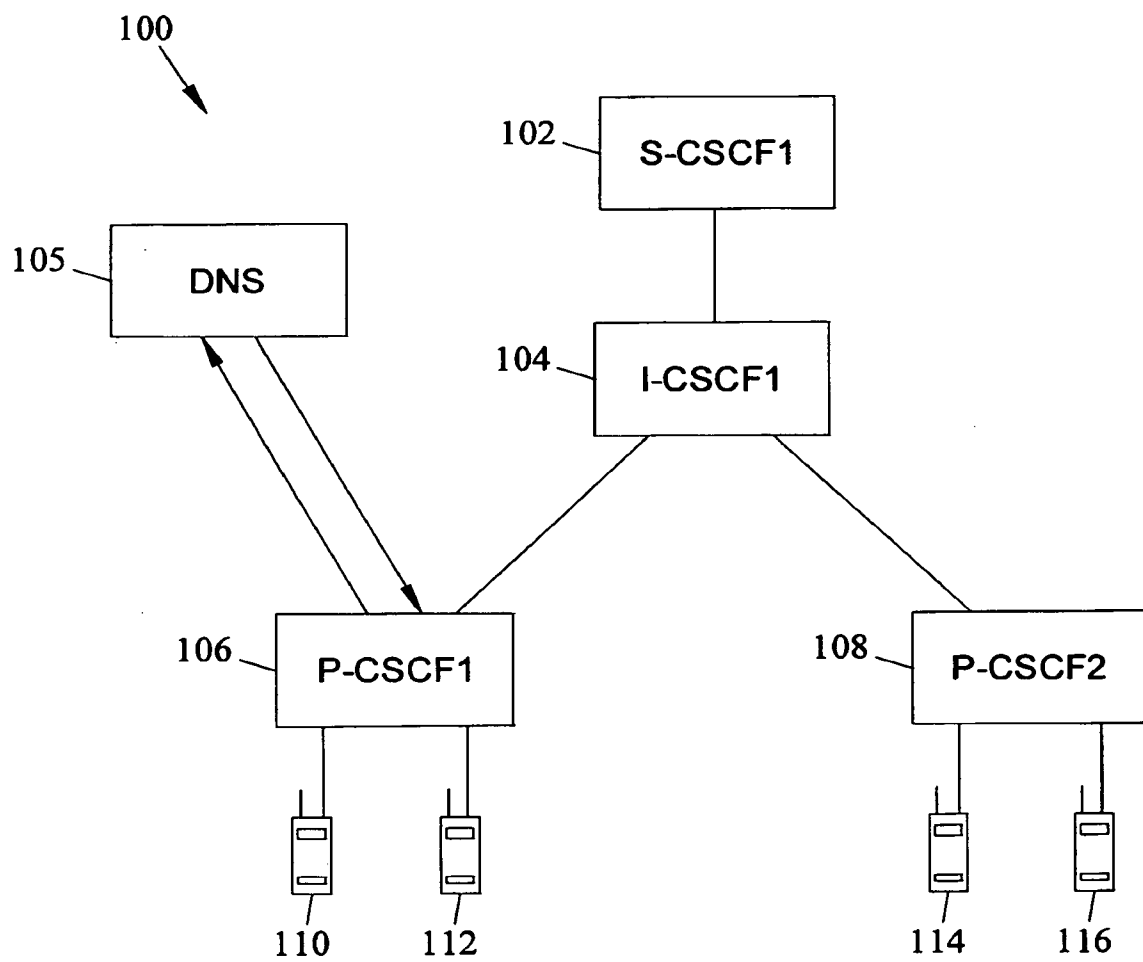


FIG. 1

2/9

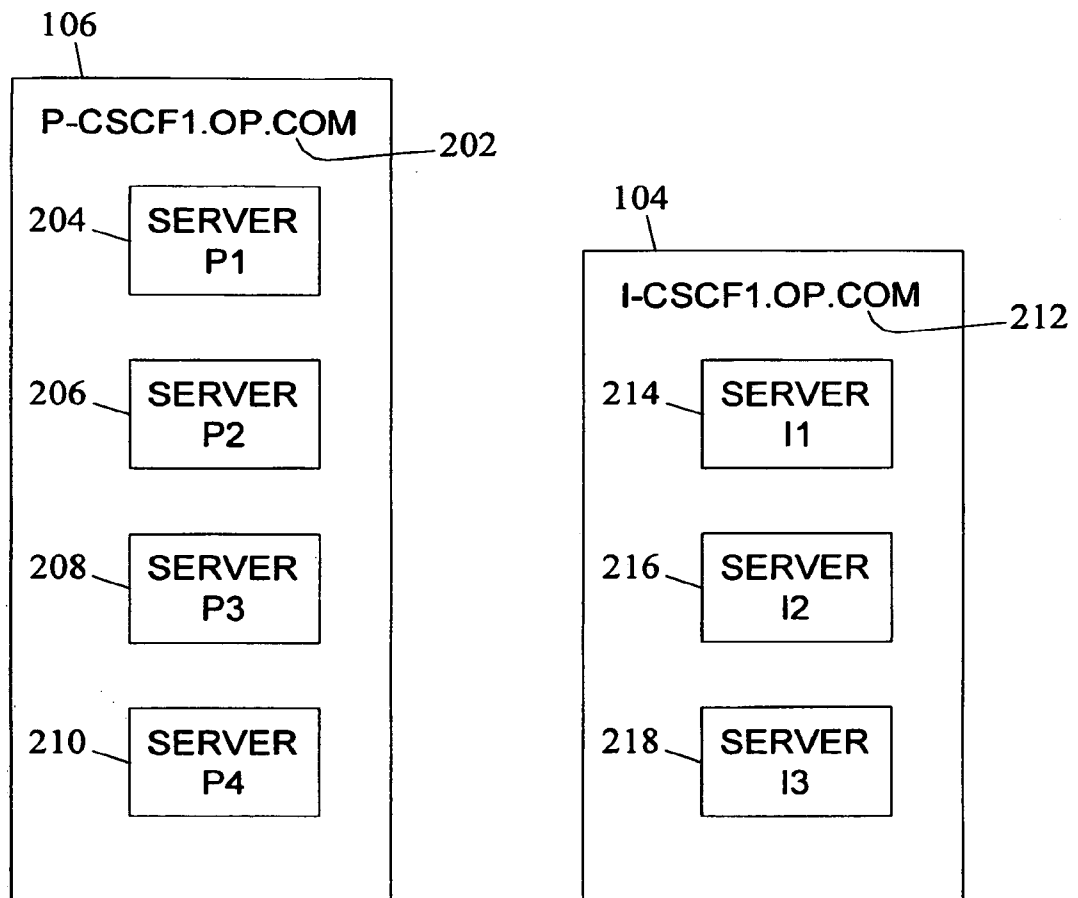


FIG. 2

3/9

300

DOMAIN IDENTIFER	GLOBAL SERVER LIST			
P-CSCF1.OP.COM	P1	P2	P3	P4
I-CSCF1.OP.COM	I1	I2	I3	-

FIG. 3A

302

ACCESSIBLE SERVER	IP ADDRESS
I-CSCF1 I1	22.94.128.1
I-CSCF1 I2	22.94.128.2
I-CSCF1 I3	22.94.128.3

FIG. 3B

304

ACCESSIBLE SERVER	STATUS	RETRY INTERVAL
I-CSCF1 I2	CONGESTED	2 SECONDS
I-CSCF1 I3	FAILED	30 SECONDS

FIG. 3C

4/9

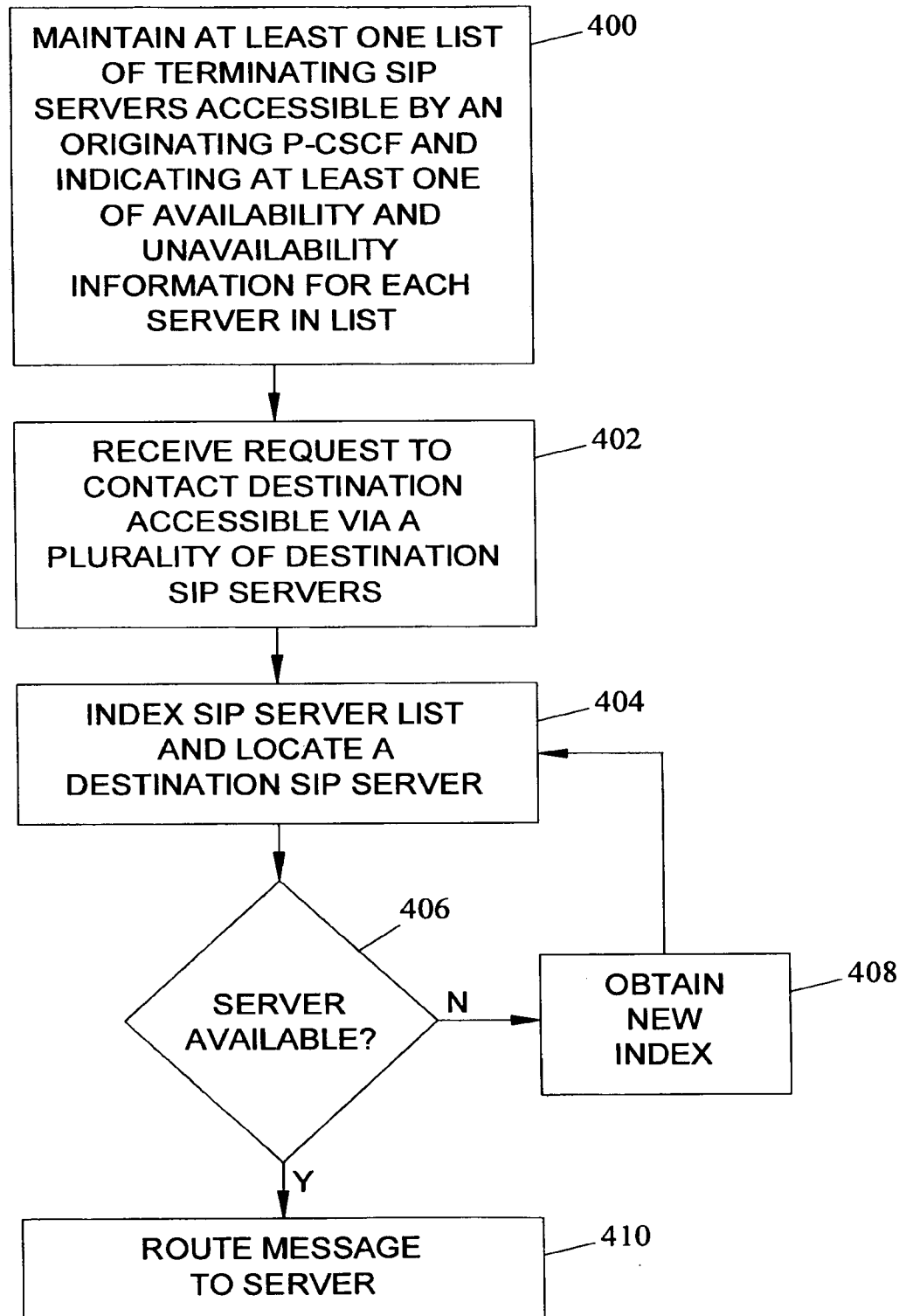


FIG. 4

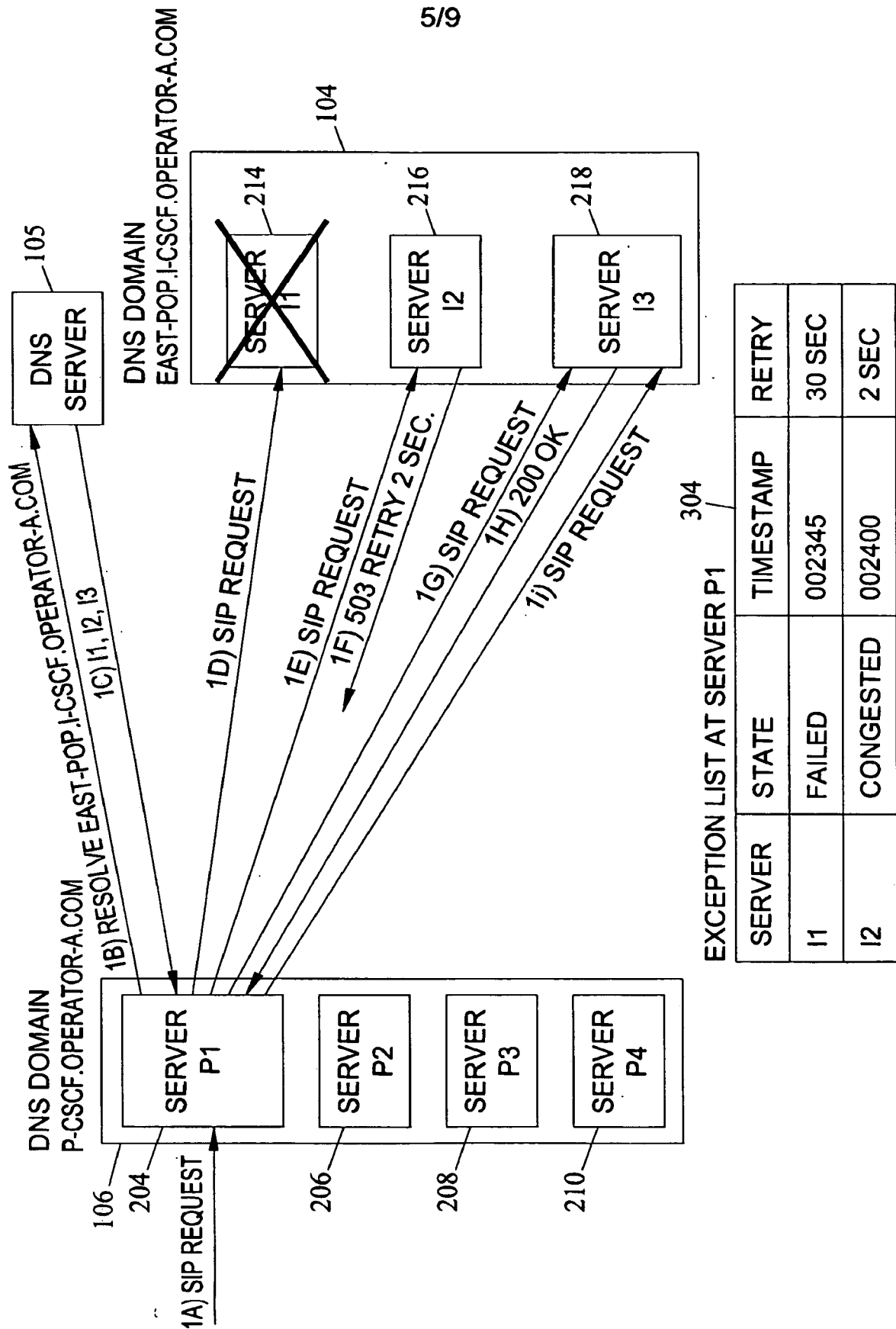


FIG. 5

6/9

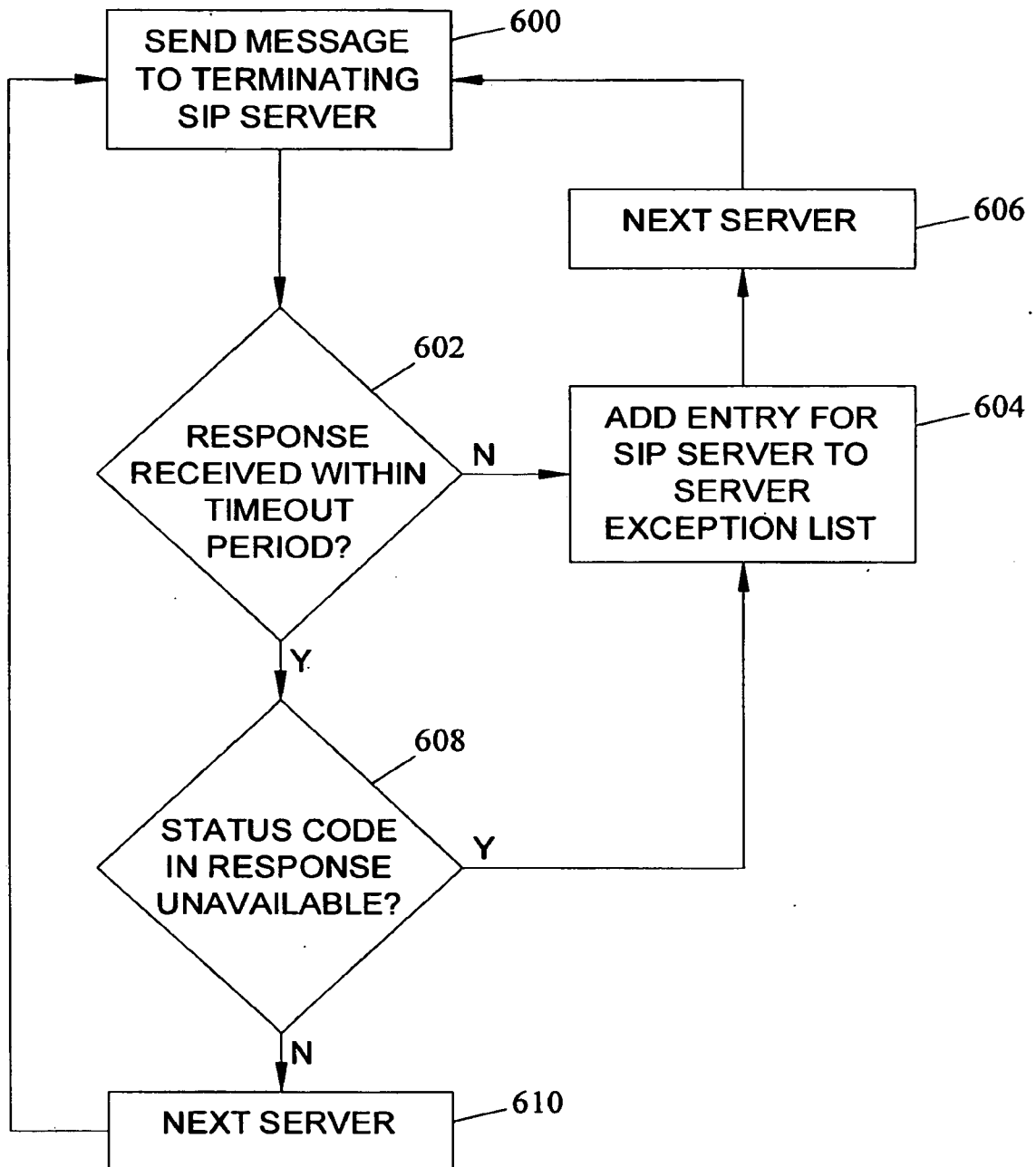


FIG. 6

7/9

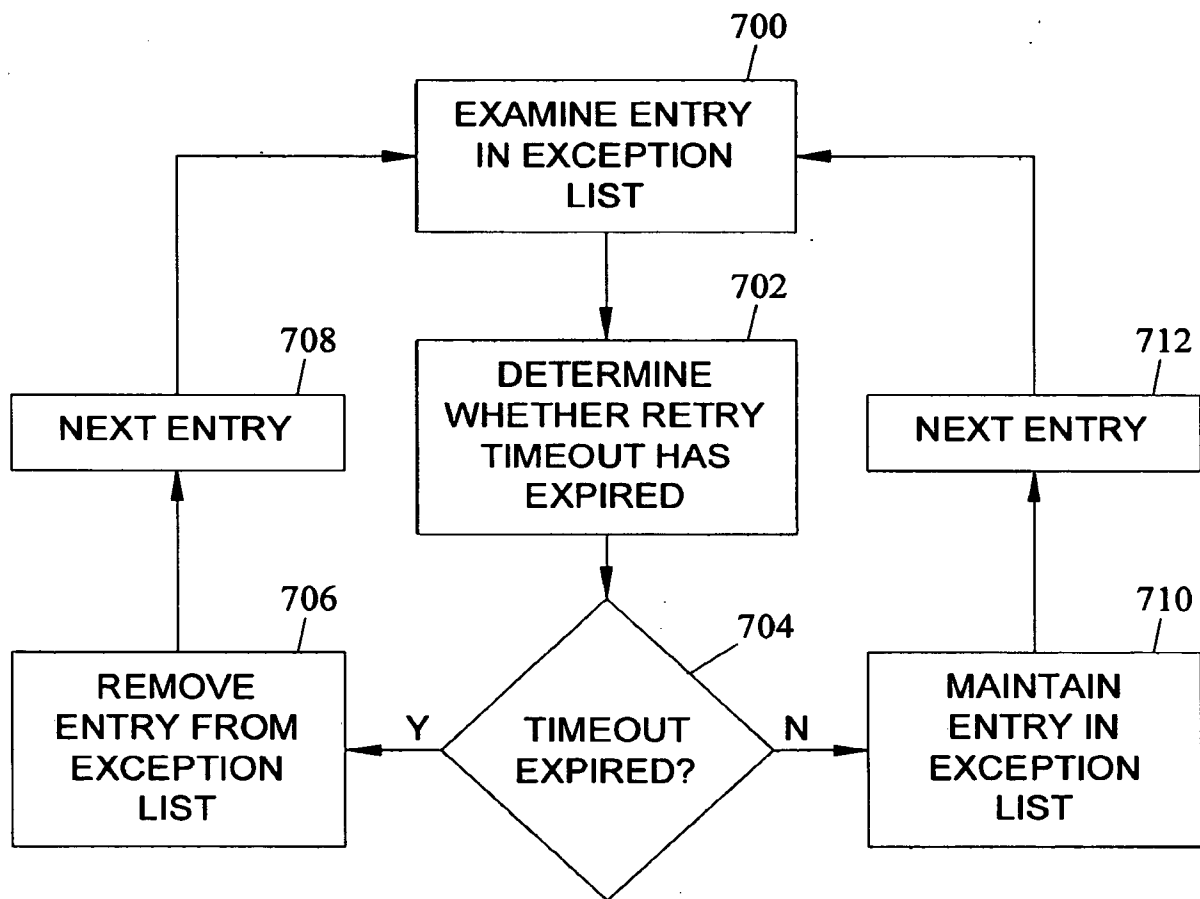


FIG. 7

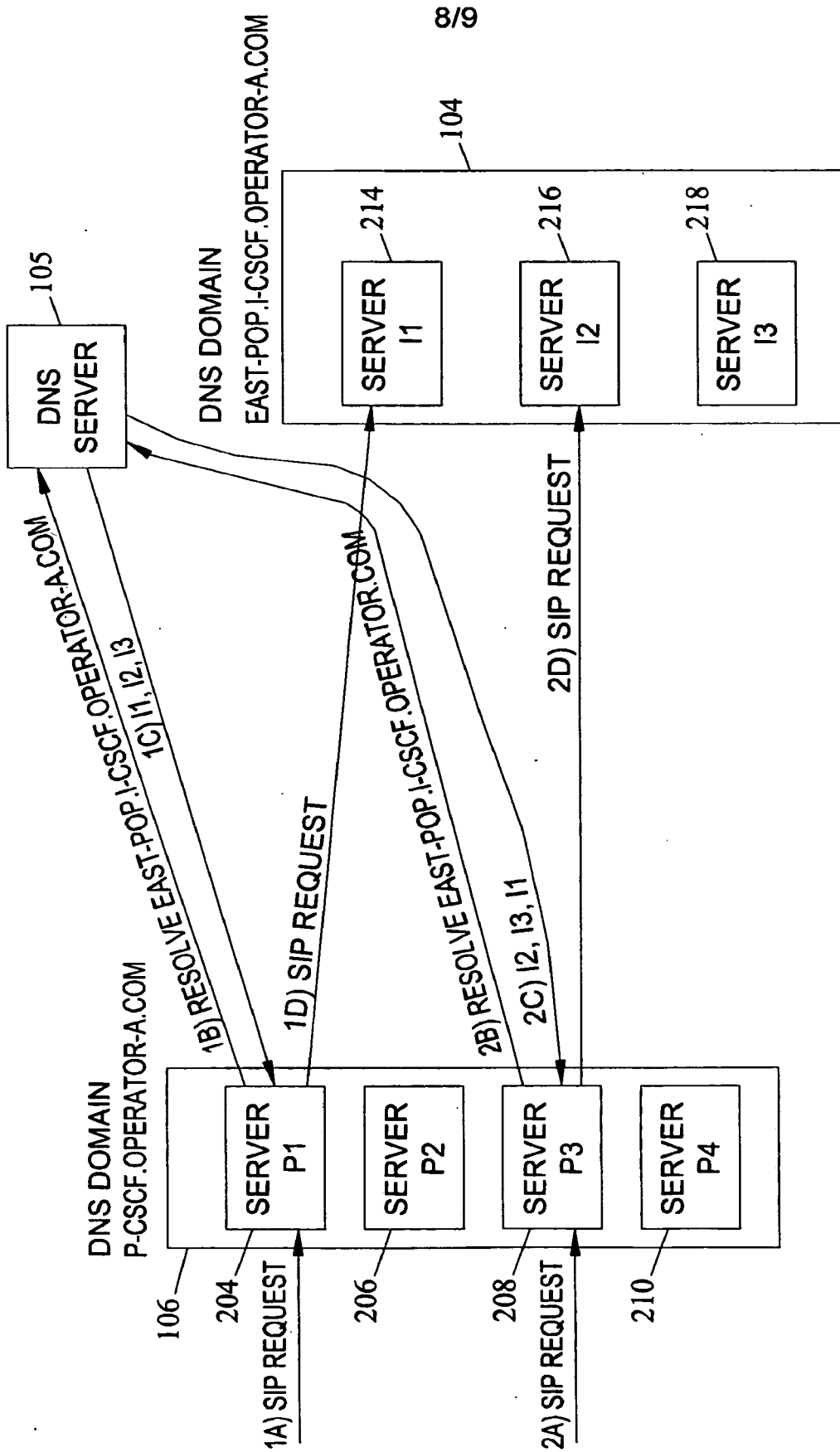


FIG. 8

9/9

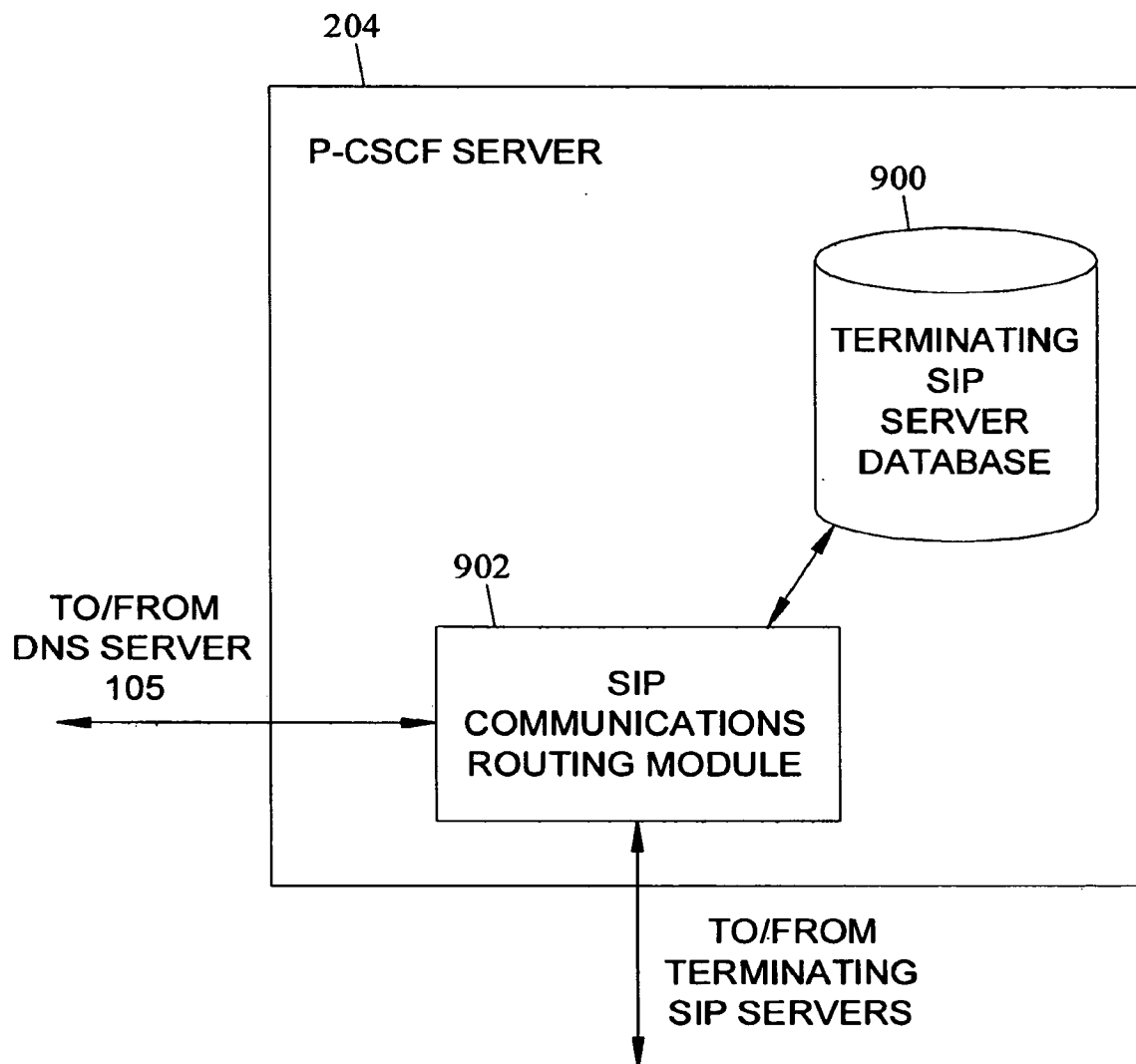


FIG. 9