

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
1 April 2004 (01.04.2004)

PCT

(10) International Publication Number  
WO 2004/028110 A1

(51) International Patent Classification<sup>7</sup>: H04L 29/06

(21) International Application Number:  
PCT/IB2003/003962

(22) International Filing Date:  
16 September 2003 (16.09.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
10/245,476 17 September 2002 (17.09.2002) US

(71) Applicant: NORTEL NETWORKS LIMITED  
[CA/CA]; 2351 Boulevard Alfred-Nobel, St. Laurent, Quebec H4S 2A9 (CA).

(72) Inventors: SEN, Sanjoy; 6400 Independent Parkway, Apt. 4201, Plano, TX 75023 (US). MARCH, Sean; 8804 Jennifer Court, Plano, TX 75025 (US). AOUN, Cedric; 0032 Rue Des Peupliers, F-92100 Boulogne-Billancourt (FR).

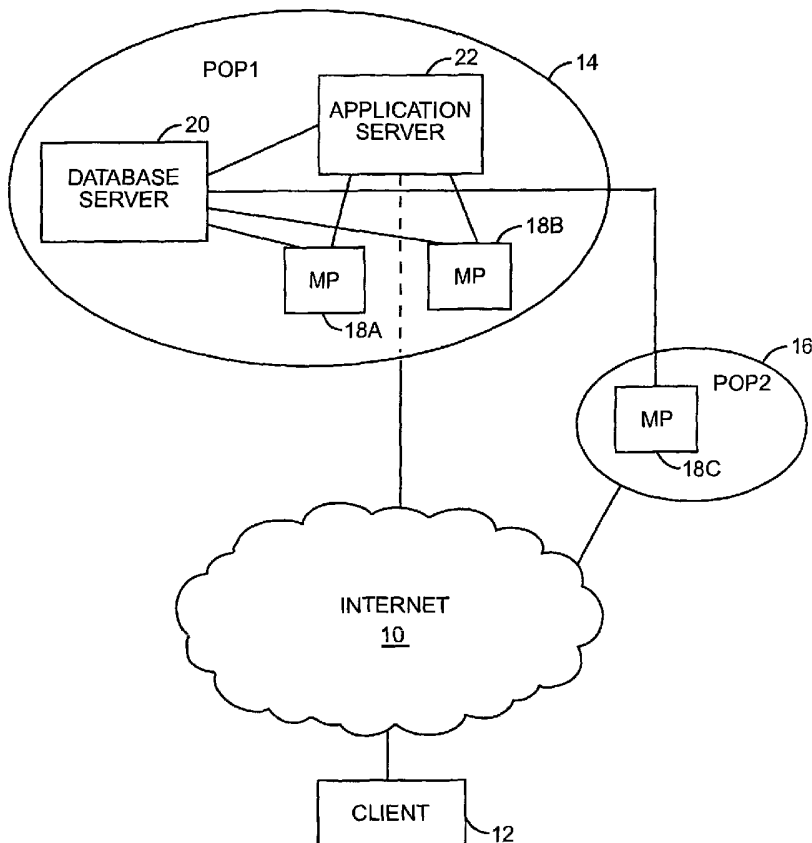
(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG,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

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

Published:  
— with international search report

[Continued on next page]

(54) Title: PROXIMITY DETECTION FOR MEDIA PROXIES



(57) Abstract: A method of detecting proximity between a media proxy and a client uses a proximity probe to query a plurality of media proxies, forcing the media proxies to respond to a proximity server. The proximity server uses an algorithm to determine which media proxy is closest to the client based on the responses. In an alternate embodiment, the same sorts of proximity probes may be used to determine if two media endpoints have a direct connection such that they may bypass a media proxy.

WO 2004/028110 A1



- 
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## ***PROXIMITY DETECTION FOR MEDIA PROXIES***

### Field of the Invention

**[0001]** The present invention relates to a technique to determine a best path over which a multimedia communication session may be held.

### Background of the Invention

**[0002]** Consumers of telecommunications services are faced with many options when it comes to services. One of the more dynamic fields currently is the provision of multimedia services. Typically, a consumer will request multimedia services by accessing an application server. The application server will direct the consumer to one of a plurality of media proxies in a round robin fashion. The media proxy may then set up and tear down the call as is well understood and provide other multimedia related services.

**[0003]** The round robin approach suffers from some drawbacks. First, the round robin approach ignores the loads that are currently being experienced on the media proxies; for instance, there is no provision to evaluate loading on the media proxies before making the assignment. As such, a heavily loaded media proxy may still be assigned to provide services to a consumer just because it is next in the round robin cycle.

**[0004]** Second, the round robin approach ignores distances between the media proxy and the consumer. The distance may be a function of communication links, physical distance, and similar physical attributes. Distance introduces delay into the provision of services, which for multimedia applications can be extremely undesirable. The round robin approach assigns a media proxy to a consumer irrespective of whether another media proxy might be closer.

**[0005]** Thus, there is a need for a better technique to allocate media proxies to incoming requests.

### Summary of the Invention

**[0006]** The present invention addresses the shortcomings of the round robin approach by determining which media proxy is "closest" to the client and assigning that media proxy to the client. Initially, a proximity server, which in

a preferred embodiment is a database server, associated with an application server. The application server determines a round trip time required for a message to go from the database server to each media proxy and back. Subsequently, the application server may receive a request for multimedia services from a client. The application server returns a list of available media proxies to the client for which the database server has the round trip times. The client then sends out a probe query to each media proxy on the list. The media proxies, upon receipt of the probe, send a time-stamped message to the database server. The database server determines which media proxy is closest to the client based on the round trip times and the time required for the media proxy to receive the probe query from the client. After the determination, the application server assigns the client to a media proxy and the media services are provided as requested.

**[0007]** Those skilled in the art will appreciate the scope of the present invention and realize additional aspects thereof after reading the following detailed description of the preferred embodiments in association with the accompanying drawing figures.

#### Brief Description of the Drawings

**[0008]** The accompanying drawing figures incorporated in and forming a part of this specification illustrate several aspects of the invention, and together with the description serve to explain the principles of the invention.

**[0009]** Figure 1 illustrates a network according to one embodiment of the present invention;

**[0010]** Figure 2 illustrates, as a flow chart, the methodology of the present invention; and

**[0011]** Figure 3 illustrates a call flow diagram associated with the exemplary embodiment of Figure 2.

#### Detailed Description of the Preferred Embodiments

**[0012]** The embodiments set forth below represent the necessary information to enable those skilled in the art to practice the invention and illustrate the best mode of practicing the invention. Upon reading the following description in light of the accompanying drawing figures, those skilled in the art will understand the concepts of the invention and will

recognize applications of these concepts not particularly addressed herein. It should be understood that these concepts and applications fall within the scope of the disclosure and the accompanying claims.

**[0013]** Before discussing how the present invention works, an explanation of the network environment is presented so as to help understand the context of the present invention. A network 10 may be the public internet (shown), a private intranet, or other network through which various elements may interact, such as through any appropriate packet based protocol. A client 12 may be communicatively coupled to the network 10. The client 12 may be a personal computer, a multimedia enabled phone, or other multimedia enabled customer premises equipment. One or more points of presence (POPs) 14, 16 may be associated with one or more service providers. The POPs 14, 16 may have one or more media proxies (MPs) 18A, 18B, and 18C (hereinafter, generically a media proxy 18) associated therewith, as well as a proximity server 20, which in a preferred embodiment is a database server. In most installations, it is expected that a single application server 22 may control more than one media proxy 18, and further the application server 22 may control media proxies 18 distributed amongst multiple POPs 14, 16. The application server 22 may, in an exemplary embodiment, use Session Initiation Protocol (SIP) for communicating.

**[0014]** For the sake of providing a more complete explanation, a brief digression into the specifics of SIP is herein provided. A SIP endpoint is generally capable of running an application, which is generally referred to as a user agent (UA), and is capable of facilitating media sessions using SIP. User agents register their ability to establish sessions with a SIP proxy (the application server 22 in the present invention) by sending "REGISTER" messages to the SIP proxy. The REGISTER message informs the SIP proxy of one or more SIP universal resource locators (URL) that identify the user agent to the SIP network. The REGISTER message also contains information about how to reach specific user agents over the SIP network by providing the Internet Protocol (IP) address and port that the user agent will use for SIP sessions.

**[0015]** A "SUBSCRIBE" message may be used to subscribe to an application or service provided by a SIP endpoint. Further, "NOTIFY"

messages may be used to provide information between SIP endpoints in response to various actions or messages, including REGISTER and SUBSCRIBE messages.

**[0016]** When a user agent wants to establish a session with another user agent, the user agent initiating the session will send an INVITE message to the SIP proxy and specify the targeted user agent in the "TO:" header of the INVITE message. Identification of the user agent takes the form of a SIP URL. In its simplest form, the URL is represented by a number or "<username>@<domain>," such as "janedoe@nortelnetworks.com." Generally, the user name is unique within the name space of the specified domain.

**[0017]** If the targeted user agent has registered with the SIP proxy, the SIP proxy will forward any received INVITE message directly to the targeted user agent. The targeted user agent will respond with a 200 OK message, and a session between the respective user agents will be established as per the message exchange required in the SIP specification. Media capabilities are passed between the two user agents of the respective endpoints as parameters embedded within the session setup messages, such as the INVITE, 200 OK, and acknowledgement (ACK) messages. The media capabilities are typically described using the Session Description Protocol (SDP). Once respective endpoints are in an active session with each other and have determined each other's capabilities, the specified media content may be exchanged during an appropriate media session.

**[0018]** Returning now to the present invention, the proximity server 20 has a pre-configured list of media proxies 18, and more specifically, a list of addresses and ports on the media proxies 18 that serve the domains served by the application server 22. Further, at least one port on each media proxy 18 is reserved for running the proximity detection method of the present invention.

**[0019]** Note that, in general, the client 12, the media proxy 18, the proximity server 20, and the application server 22 all may contain data processing devices, such as a microprocessor. Further, the microprocessor may be controlled by software or hardware that performs the functions of the present invention. The software may be stored on any appropriate computer

readable medium such as a hard drive, a disc, or other memory device. Note further that it is possible that the proximity server 20 and the application server 22 be collocated or integrated into a single entity, but such is not required. The proximity server is one possible host for the proximity detection algorithm. The algorithm is independent of any platform.

**[0020]** The methodology of the present invention will now be presented with reference to Figures 2 and 3. Figure 2 is a flowchart of the methodology and Figure 3 shows the call flows. Steps in the flow chart begin with the numeral one, such as block 100, while steps in the call flow begin with the numeral two, such as message 200. While particular protocols and messaging examples are provided, they are intended to be exemplary and assist in comprehension of the steps of the present invention. Other messaging protocols could be used as needed or desired.

**[0021]** In an exemplary embodiment, the proximity server 20 periodically estimates a round trip time (RTT) between the proximity server 20 and each of the media proxies 18 under its jurisdiction (block 100). This RTT estimation may be done using an Internet Control Message Protocol (ICMP) ping as is well understood. At some time subsequent to the estimation of the RTT, a client 12, which may be a SIP client, sends a SIP REGISTER message (200) requesting registration with the application server 22 (block 102) as part of a request for multimedia services. The application server 22 sends the pre-configured list of addresses (including port addresses, if appropriate) to the client 12 for media proxies 18 that are available to serve the domain from which the client's request originates (block 104). This list of media proxies 18 can be sent in the body of the 200 OK response to the REGISTER message (202).

**[0022]** The client 12 sends a Simple Traversal of UDP Through Network Address Translators (STUN), ICMP ping, or an application layer probe message (204A, 204B, 204C) to all of the media proxies 18 in the list received from the application server 22 (block 106). This message may include the user identification (USER ID) of the client 12. The probe may be sent out essentially simultaneously.

**[0023]** Each media proxy 18, upon receiving the probe message, sends a notification message (206A, 206B, 206C) reporting the probe event to the

proximity server 20 (block 108). This notification message may contain the user identification of the client 12 that triggered the event, as well as a timestamp (TIME). The media proxy 18 also sends back an acknowledgement (208A, 208B, 208C) to the client 12 (block 110) probe. The acknowledgement informs the client 12 which media proxies 18 are reachable. Optionally, the media proxy 18 may include in the notification message (206A, 206B, 206C) a timestamp delta representing the time between the receipt of the client probe and the time of sending the notification message to the proximity server 20. In an exemplary embodiment, the timestamp delta value has a granularity of milliseconds, although finer gradations could be used.

**[0024]** The proximity server 20 notes the time of receipt of the notification messages (block 112). The proximity server 20 then uses these values and the round trip time (RTT) previously gathered for each of the media proxies 18 to estimate the relative proximities of the client 12 to each of the media proxies 18 (block 114). While there may be several ways to estimate the proximity, an exemplary method to estimate the proximity is as follows.

**[0025]** RTT<sub>A</sub>, RTT<sub>B</sub>, and RTT<sub>C</sub> are the RTTs for the media proxies 18A, 18B, and 18C, respectively. TA, TB, and TC are the recorded times at the proximity server 20 representing the receipt of the notification messages. The relative proximity values may thus be computed as:

- i.  $Prox(A, B) = (TA - TB) - (RTT_A - RTT_B)$
- ii.  $Prox(A, C) = (TA - TC) - (RTT_A - RTT_C)$
- iii.  $Prox(B, C) = (TB - TC) - (RTT_B - RTT_C)$

**[0026]** Note that if the timestamp deltas are present, they may be subtracted out of the calculation for greater accuracy. For example, if  $\Delta_A$  and  $\Delta_B$  are present, then  $Prox(A, B) = (TA - TB) - (RTT_A - RTT_B) - (\Delta_A - \Delta_B)$ . Following the rule that if  $Prox(A, B) < 0$  (or negative), then media proxy 18A is relatively closer to the client 12 than media proxy 18B, the following table may be constructed.

Prox (A,B)	Prox (A,C)	Prox (B,C)	Relative Proximities from client
-	-	-	A<B<C



-	-	+	A<C<B
-	+	-	ERROR
-	+	+	C<A<B
+	-	-	B<A<C
+	-	+	ERROR
+	+	-	B<C<A
+	+	+	C<B<A

**[0027]** In the event of the two error cases, the relative proximities cannot be properly determined, and the RTTs of the media proxies 18 may be used to represent client 12 proximity. Using the RTTs is a good approximation so long as the proximity server 20 is located close to the client 12. Note that this proximity detection may, in an exemplary embodiment, only be performed once per registration.

**[0028]** Armed with the relative proximities, the proximity server 20 may then evaluate other factors (block 116), which include such relatively static factors as: 1) the number of hops between the media proxy 18 and the client 12 and 2) media proxy 18 capacity. Additional dynamic parameters may include the number of User Data Protocol (UDP) ports available, the processing power available, the bandwidth available, and statistics relating to jitter, delay, and packet loss at each port. These additional factors may be normalized and weighted as needed or desired to effectuate an appropriate analysis of the suitability of assigning a client 12 to a given media proxy 18.

**[0029]** The proximity server 20 then sends a best fit message (210) to the application server 22 (block 118). The best fit message (210) indicates which media proxy 18 is most appropriate for the client 12 to use and is based on the proximity and the additional factors outlined above.

**[0030]** The application server 22 then instructs with message 212 the client 12 to use the selected media proxy 18 (block 120). Note that the proximity server 20 could provide the instruction to the client 12. Likewise, the processing of the proximity calculations could be performed in the application server 22, although such is not preferred.

**[0031]** Appropriate authentication or encryption steps may be taken to prevent duplication of the user identification that may be sent in some of the various messages. Likewise, denial of service attacks may be prevented in standard ways, such as monitoring traffic levels.

**[0032]** Those skilled in the art will recognize improvements and modifications to the preferred embodiments of the present invention. All such improvements and modifications are considered within the scope of the concepts disclosed herein and the claims that follow.

## Claims

What is claimed is:

1. A method of determining proximity between a media proxy and a client comprising:
  - receiving a request from a client for a media session;
  - passing a list of media proxies to the client;
  - receiving responses from the media proxies of the list of media proxies;and
  - calculating a proximity between the client and each of the media proxies on the list of media proxies based at least in part on the responses.
2. The method of claim 1 further comprising assigning a media proxy to the client based, at least in part, on the calculating.
3. The method of claim 1 further comprising determining a round trip time (RTT) between a proximity server and each of the media proxies on the list of media proxies.
4. The method of claim 3 wherein calculating a proximity comprises calculating a proximity based at least in part on the RTT.
5. The method of claim 1 wherein calculating a proximity comprises calculating a proximity in a proximity server.
6. The method of claim 1 wherein said receiving and passing comprise receiving and passing in an application server.
7. A method of determining proximity between a media proxy and a client comprising:
  - receiving a list of media proxies at the client;
  - generating a ping probe to each of the media proxies on the list of media proxies such that each of the media proxies responds to a proximity server upon receiving the respective ping probe; and

receiving an assignment of a media proxy from the list based, at least in part, on the responses the media proxies sent upon receiving the respective ping probes.

8. The method of claim 7 wherein generating a ping probe includes generating a ping probe with a user identification therein.
9. The method of claim 7 further comprising initiating a media session using the media proxy assigned to the client.
10. A method of determining proximity between a media proxy and a client comprising:
  - receiving a ping probe at the media proxy from the client;
  - generating a response at the media proxy to the ping probe directed to a proximity server including a timestamp corresponding to when the ping probe was received; and
  - being assigned to the client.
11. The method of claim 10 wherein generating a response further comprises including a timestamp delta corresponding to time elapsed from receipt of the ping probe and generation of the response at the media proxy.
12. The method of claim 11 wherein including a timestamp delta comprises including a timestamp delta with millisecond granularity.
13. The method of claim 11 wherein receiving a ping probe from the client comprises receiving a ping probe with a user identification of the client.
14. The method of claim 11 further comprising responding to a round trip time inquiry from a proximity server such that the database may determine a round trip time elapsed for messages sent between the proximity server and the media proxy.

15. An apparatus for determining proximity between a media proxy and a client comprising a data processing device adapted to:
- receive a request from a client for a media session;
  - pass a list of media proxies to the client;
  - receive responses from the media proxies of the list of media proxies;
- and
- calculate a proximity between the client and each of the media proxies on the list of media proxies based at least in part on the responses.
16. The apparatus of claim 15 wherein said data processing device is further adapted to assign a media proxy to the client based, at least in part, on the calculating.
17. The apparatus of claim 15 wherein said data processing device is further adapted to determine a round trip time (RTT) between a proximity server and each of the media proxies on the list of media proxies.
18. The apparatus of claim 17 wherein said data processing device is further adapted to calculate a proximity based, at least in part, on the RTT.
19. A client seeking media services comprising a processor adapted to:
- receive a list of media proxies at the client;
  - generate a ping probe to each of the media proxies on the list of media proxies such that each of the media proxies responds to a proximity server upon receiving the respective ping probe; and
  - receive an assignment of a media proxy from the list based, at least in part on the responses the media proxies sent upon receiving the respective ping probes.
20. The client of claim 19 wherein said processor adapted to generate a ping probe generates a ping probe with a user identification therein.
21. The client of claim 19 wherein said processor is further adapted to initiate a media session using the media proxy assigned to the client.

22. A media proxy for providing media services to a client, said media proxy comprising a data processing device adapted to:

receive a ping probe from the client;

generate a response to the ping probe directed to a proximity server including a timestamp corresponding to when the ping probe was received;

and

be assigned to the client.

23. The media proxy of claim 22 wherein the data processing device is adapted to include a timestamp delta corresponding to time elapsed from receipt of the ping probe and generation of the response at the media proxy.

24. The media proxy of claim 22 wherein the data processing device is adapted to include a timestamp delta with a msec granularity.

25. The media proxy of claim 22 wherein the data processing device is adapted to receive a ping probe with a user identification of the client.

26. The media proxy of claim 22 wherein the data processing device is further adapted to respond to a round trip time inquiry from a proximity server such that the database may determine a round trip time elapsed for messages sent between the proximity server and the media proxy.

27. A computer readable medium comprising software for determining proximity between a media proxy and a client, said software programmed to:

receive a request from a client for a media session;

pass a list of media proxies to the client;

receive responses from the media proxies of the list of media proxies;

and

calculate a proximity between the client and each of the media proxies on the list of media proxies based at least in part on the responses.

28. A computer readable medium adapted for use at a client and comprising software programmed to:

receive a list of media proxies at the client;

generate a ping probe to each of the media proxies on the list of media proxies such that each of the media proxies responds to a proximity server upon receiving the respective ping probe; and

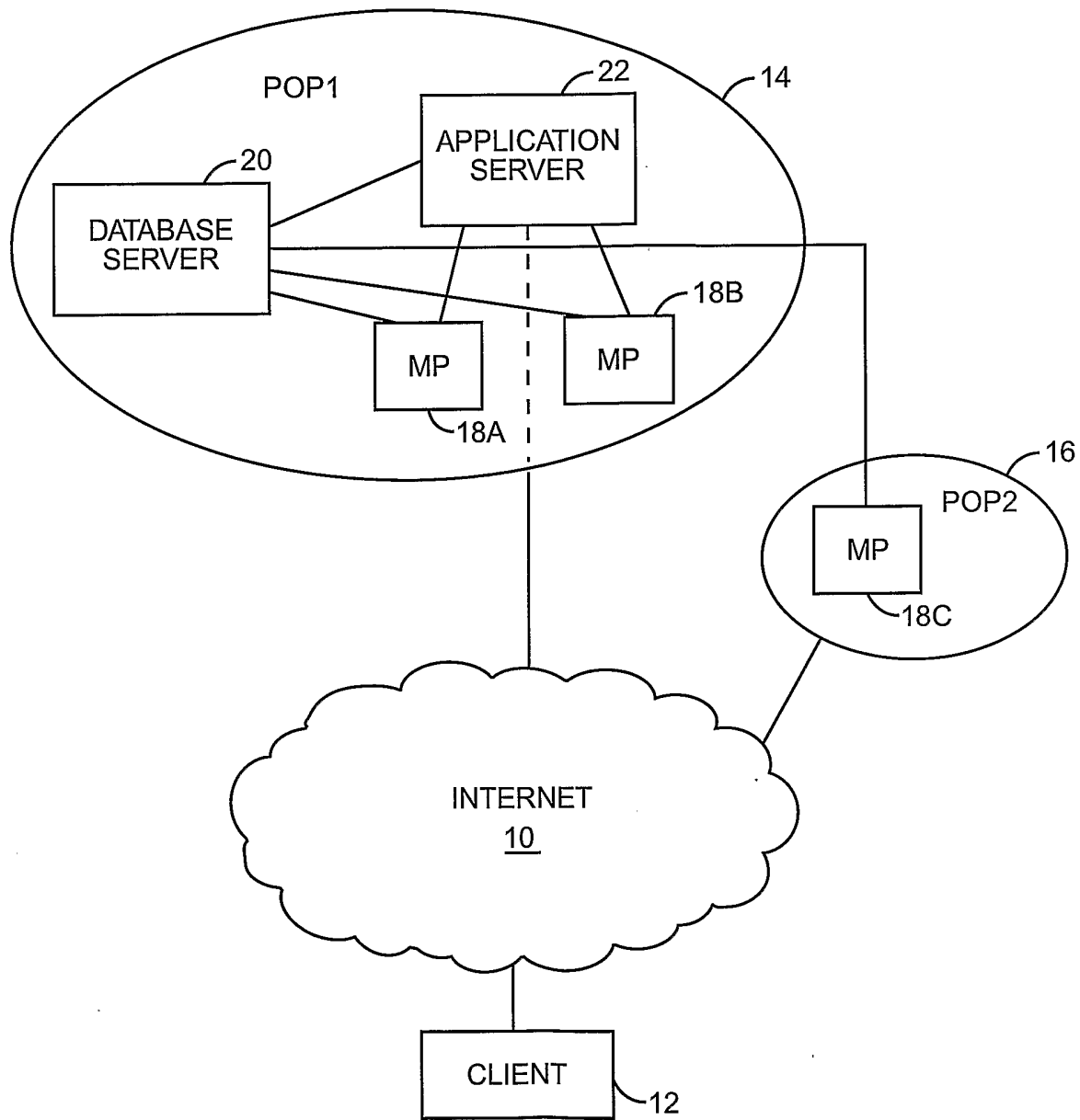
receive an assignment of a media proxy from the list based, at least in part on the responses the media proxies sent upon receiving the respective ping probes.

29. A computer readable medium adapted for use at a media proxy comprising for providing media services to a client and programmed to:

receive a ping probe from the client;

generate a response to the ping probe directed to a proximity server including a timestamp corresponding to when the ping probe was received; and

be assigned to the client.



**FIG. 1**



2/3

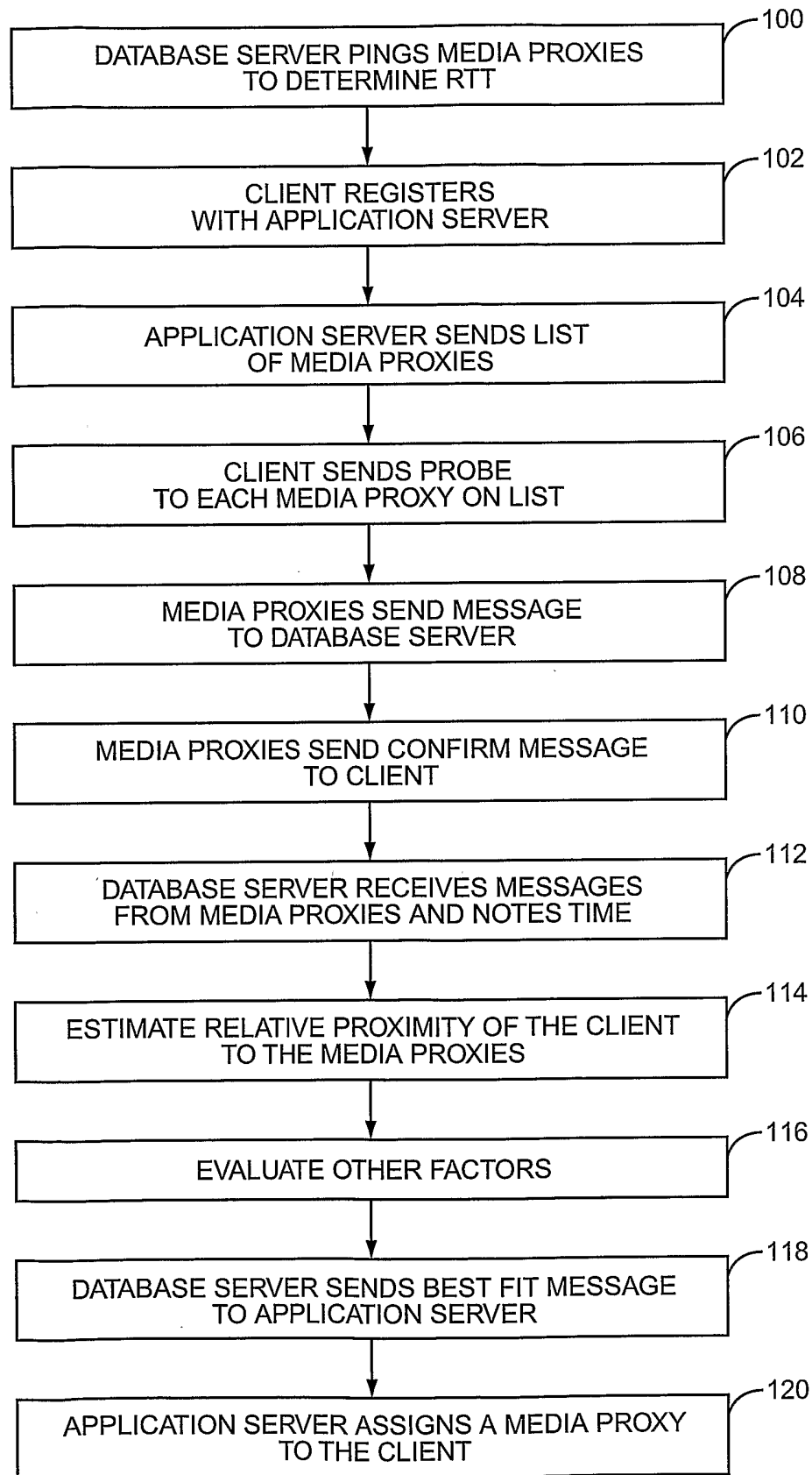


FIG. 2

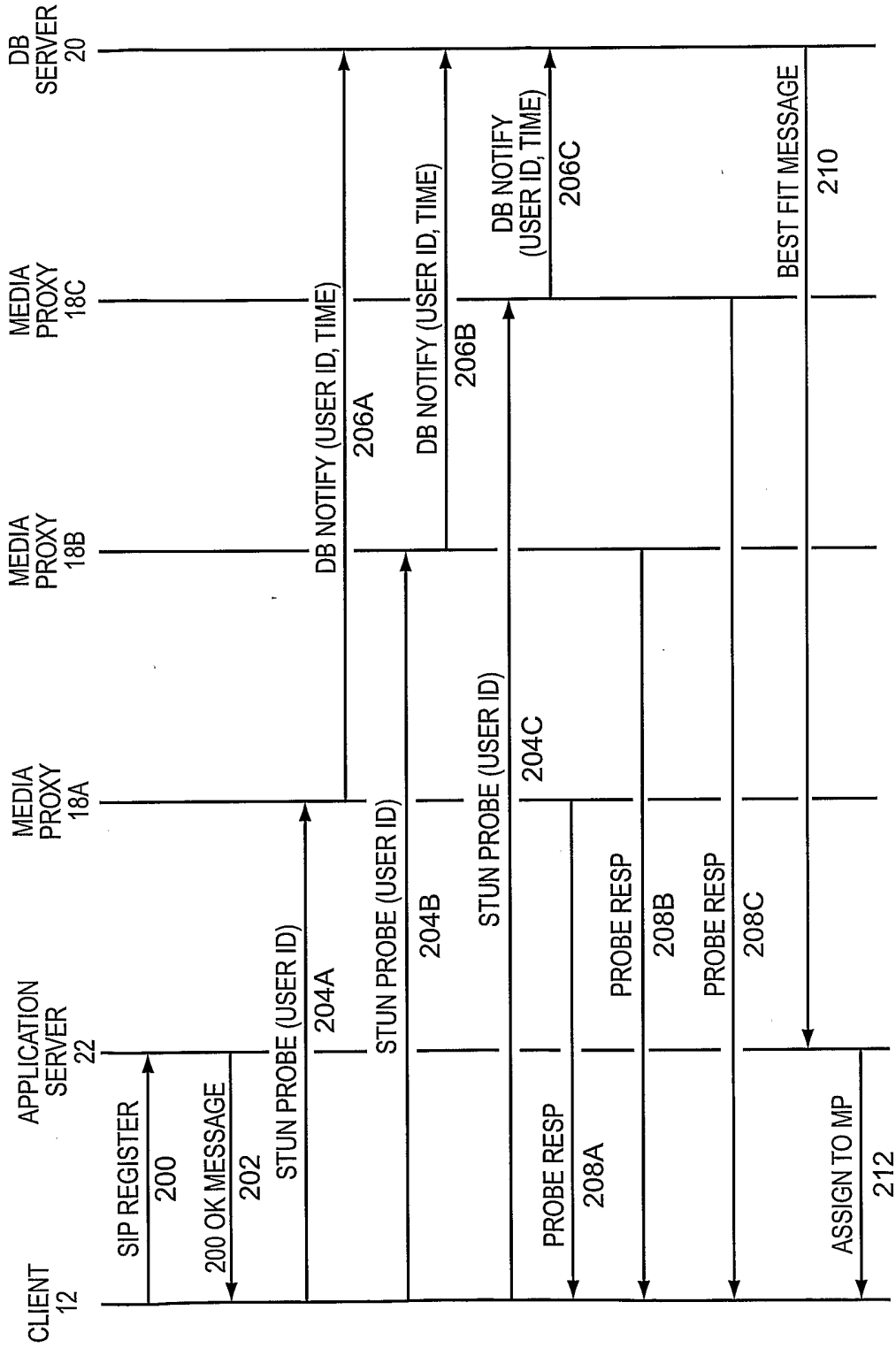


FIG. 3

## INTERNATIONAL SEARCH REPORT

PCT/IB 03/03962

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 H04L29/06

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, COMPENDEX, IBM-TDB

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2002/059622 A1 (VELIKANOV CYRILLE ET AL) 16 May 2002 (2002-05-16) abstract page 2, paragraph 20 column 3, paragraph 25 page 5, paragraph 61 - paragraph 71 page 6, line 76 - line 79 -----	1-29
A	US 2002/038360 A1 (WINKLER PETER ET AL) 28 March 2002 (2002-03-28) page 1, paragraph 10 - paragraph 11 page 2, paragraph 16 page 4, line 41 - line 57 -----	1-29

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

° Special categories of cited documents :

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\* & \* document member of the same patent family

Date of the actual completion of the international search

21 January 2004

Date of mailing of the international search report

28/01/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Karavassilis, N

INTERNATIONAL SEARCH REPORT

PCT/IB 03/03962

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2002059622	A1	16-05-2002	NONE
US 2002038360	A1	28-03-2002	NONE