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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2013 (17.01.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/009663 A1

(51) International Patent Classification:
H04L 29/06 (2006.01)

(21) International Application Number:
PCT/US2012/045880

(22) International Filing Date:
8 July 2012 (08.07.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/179,220 8 July 2011 (08.07.2011) US

(71) Applicant *(for all designated States except US)*: **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **SUNKARA, Ramu**; c/o Skype, International Patents, 70 Sir John Rogerson's Quay, Dublin, 2 (IE). **MARKOV, Alexander**; c/o Skype, International Patents, 70 Sir John Rogerson's Quay, Dublin, 2 (IE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

(54) Title: COMMUNICATION SYSTEM FOR ESTABLISHING A REAL-TIME COMMUNICATION SESSION

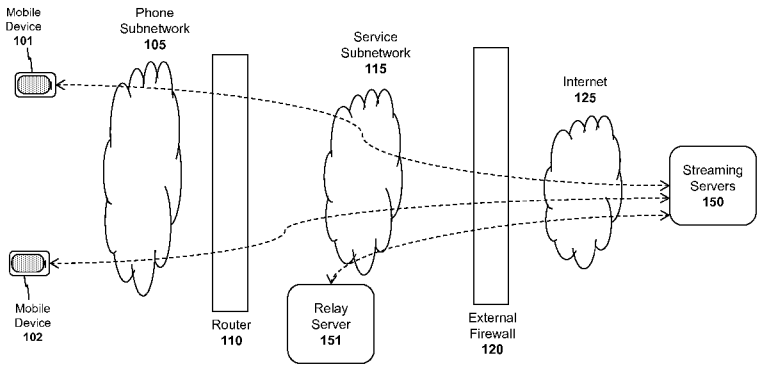

Fig. 2a

(57) Abstract: A computer-implemented system and method are described for improving the quality of experience (QoE) of real-time video sessions between mobile users. For example, a method according to one embodiment of the invention comprises: configuring one or more relay servers on the perimeter of a service provider network by a streaming server via the Internet; at the streaming server, receiving a request from a first mobile device to establish a real-time communication session with a second mobile device; the streaming server subsequently providing the first and second mobile devices with networking information for connecting via the relay server(s) which is(are) located close to at least one of the first and second mobile devices; and establishing the real-time communication session through the relay server(s).

WO 2013/009663 A1

BACKGROUND

Field of the Invention

[0001] This invention relates generally to the field of data networking. More particularly, the invention relates to an improved data communication system.

Description of the Related Art

[0002] Internet users today may communicate using a variety of different client applications including real-time messaging applications (e.g., instant messaging or "chat" applications) and real-time, two-way video applications. For example, certain client applications (such as, for example, the "Qik" client developed by Qik, Inc.), allow users to participate in two-way live video chat over 3G and 4G cellular phone networks.

[0003] **Figures 1a-b** illustrate an exemplary configuration in which two mobile devices 101-102 participate in a two-way video chat. **Figure 1a** illustrates control/signaling data routing used to establish a two-way video connection and **Figure 1b** illustrates the media data (i.e., the actual video streams) being transmitted between the two mobile devices 101-102. The control/signaling is typically carried using a transmission control protocol (TCP) or user datagram protocol (UDP) network transport and the media data is typically carried using a UDP network transport.

[0004] As illustrated, the path taken by both the control/signaling and media data includes a phone subnetwork 105 which may be a standard wireless cellular network (e.g., such as a 3G or 4G network). The phone subnetwork 105 is connected to a service sub-network 115 via a router 110. The service subnetwork is a data network used to support mobile data traffic and route the data traffic over the Internet 125 via an external firewall 120. The phone subnetwork 105 and service subnetwork 115 are maintained by a wireless service provider such as T-Mobile™, AT&T™, or Verizon™.

[0005] Currently, to enable real-time video conversations between the mobile devices 101-102, each of the mobile devices must be connected over the Internet 125 to a common set of streaming servers 150. Each mobile device 101-102 opens its own UDP datagram socket to the streaming servers 105 and the streaming servers then route the datagrams containing the video content over the appropriate datagram sockets. For example, the streaming servers 150 route video received from mobile device 101 over mobile device 102's datagram socket connection and vice versa.

[0006] One drawback of this configuration is that because each mobile device 101-102 must connect to a streaming server over the Internet 125, significant latency may be introduced. Consequently, a more efficient way to route video between the mobile devices 101-102 would be desirable.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] A better understanding of the present invention can be obtained from the following detailed description in conjunction with the following drawings, in which:

[0008] **FIGS. 1a-b** illustrate current techniques for establishing real time video sessions between two mobile devices.

[0009] **FIGS. 2a-b** illustrate techniques for improving latency by placing relay servers on the perimeter of a service provider's network in accordance with one embodiment of the invention.

[0010] **FIG. 3** illustrates a relay service positioned on the perimeter of a service provider's network routing video content between a mobile device connected via a WiFi connection and a mobile device connected via a cellular connection (e.g., a 3G or 4G connection).

[0011] **FIGS. 4a-b** illustrate two different embodiments in which multiple relay servers are used to establish communication between mobile devices.

DETAILED DESCRIPTION

[0012] In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the embodiments of the invention described below. It will be apparent, however, to one skilled in the art that the embodiments of the invention may be practiced without some of these specific details. For example, embodiments of the invention are described below within the context of real-time video applications, the underlying principles of the invention are not limited to any particular type of media communication. Additionally, while the discussion below focuses on specific service provider networks, the underlying principles of the invention are applicable to implementations on any form of radio networks. Finally, in some instances, well-known structures and devices are shown in block diagram form to avoid obscuring the underlying principles of the embodiments of the invention.

[0013] One embodiment of the invention improves the quality of experience for end users when conducting real-time video sessions by deploying a set of in-network relays at

the perimeter of service provider networks. Specifically, As illustrated in **Figures 2a-b**, in this embodiment, a relay service is configured to communicate over the service subnetwork 115, inside the firewall 120 connecting the service subnetwork 115 to the Internet 125.

[0014] **Figure 2a** illustrates the control/signaling data routing used to establish a relay connection. Client applications installed on the mobile devices 101-102 discover these in-network relay server(s) 151 during this control/signaling phase. In one embodiment, each in-network relay server(s) 151 is registered with the streaming servers 150 using the outbound connection. In one embodiment, the mobile device 101 initiating the video session initially connects to the streaming servers 150 which then provides the network information needed to connect to the relay service (e.g., IP address and port) to both of the clients 101-102.

[0015] In one embodiment, a relay server 151 which is relatively "close" to one or both mobile devices 101, 102 on the network is selected by the streaming servers 150. Note that the proximity of the relay server 151 to the mobile devices 101, 102 may be determined based on network information provided to the streaming servers 150 (e.g., the TCP/IP addresses of the mobile devices and/or the relay server 151). The location of each relay server may be registered with the streaming servers 150 and the streaming servers 150 may use this information to determine the best relay server for servicing each request. Other information such as the current load on each of the relay servers 151 may also be factored in to the decision.

[0016] Once the two mobile devices 101-102 have the network information, they connect to the relay server(s) 151, which establishes a real-time video session between the devices, as illustrated in **Figure 2b**. In one embodiment, the media connections between the mobile devices 101-102 and the relay server 151 comprises UDP datagram sockets. The relay server(s) 151 receive video content packetized in datagrams from mobile device 101 over a first UDP datagram socket, and route the datagrams over a second UDP datagram socket to mobile device 102, which then extracts and decodes the video content from the UDP datagrams. Conversely, The relay server(s) 151 receive video packetized in datagrams from mobile device 102 over the second UDP datagram socket, and route the datagrams over the first UDP datagram socket to mobile device 101, which then extracts and decodes the video content from the UDP datagrams.

[0017] In one embodiment the router 110 is a well known device for routing network packets to an appropriate destination, as identified by the destination address stored in each packet header. For example, the mobile devices 101, 102 may use the TCP/IP address of the streaming servers 150 when sending connection requests to the streaming servers 150. The router 110 will then route the requests to the appropriate destinations.

[0018] When the embodiments of the invention shown in **Figures 2a-b** are compared against the implementation shown in **Figures 1a-b**, the benefits are clear. Because the media streams do not need to pass through the external firewall 120 and over the Internet 125, latency is significantly reduced, thereby improving the QoE for the end user. By way of example, and not limitation, current testing shows a latency reduction from 252 ms to 110 ms.

[0019] While the embodiment described above uses UDP datagrams, it should be noted that the underlying principles of the invention are not limited to any particular network communication protocols.

[0020] **Figure 3** illustrates how the relay server(s) 151 may also be used to establish a two-way video session between a mobile device 102 connected via a WiFi connection and a mobile device 101 connected via a cellular connection (e.g., a 3G or 4G connection). In this embodiment, one of the two devices may initially contact the streaming servers 150 to establish the video session. The streaming servers 150 may then notify the called party and provide the networking information needed for both mobile devices 101-102 to connect with the relay service 151. The relay service 151 then manages the real-time video session as described above (e.g., establishing UDP datagram sockets with each of the respective devices 101-102). To communicate with the mobile device 102 connected via WiFi, the relay server 151 opens a UDP hole 160 through the external firewall 120. Techniques for establishing a UDP hole through a firewall are well known and therefore will not be described in detail herein.

[0021] In the particular implementation shown in **Figure 3**, the mobile device 102 is connected via a WiFi link over an enterprise WLAN 127, which is connected to the Internet 125 via an enterprise NAT/firewall. While this may represent a common configuration, such a configuration is not required for complying with the underlying principles of the invention.

[0022] In one embodiment, multiple relay servers may be set up at convenient locations throughout the perimeter of the service provider's network. Additionally, multiple relay servers may be used to support communication across the service subnetworks of different service providers.

[0023] **Figure 4a** illustrates one embodiment in which relay server 151 and relay server 152 communicate with one another to route UDP datagrams between mobile devices 101 and 102. As in prior embodiments, the mobile device 101 initiating the connection may first contact the streaming servers 150. In response, the streaming servers 150 provide mobile device 101 networking information needed to connect with relay server 151 and mobile device 102 the networking information needed to connect with relay server 152. Upon connecting, the relay servers 151-152 then establish a UDP datagram socket between one another to route the UDP datagram traffic containing the video data.

[0024] **Figure 4b** illustrates an alternate implementation in which relay server 151 receives UDP datagram traffic containing video data from mobile device 101 and routes it to mobile device 102 and the second relay server 152 receives UDP datagram traffic containing video data from mobile device 102 and routes it to mobile device 101. As in the embodiments described above, an initial connection to the streaming servers 150 may be needed to retrieve the networking information needed to connect with the two relay servers 151-152.

[0025] While the relay servers 151-152 are configured within the service subnetwork in the discussion above, in other embodiments, some relay servers may be configured within the phone subnetwork 1056-106. In such a case, if both mobile devices are connected to the same subnetwork, a relay server on that subnetwork may be used to connect the devices. If the mobile devices are on different phone subnetworks, however, then a relay server on the service subnetwork 115 may be selected.

[0026] In the foregoing specification, the invention has been described with reference to specific exemplary embodiments thereof. It will be evident that various modifications may be made thereto without departing from the broader spirit and scope of the invention as set forth in the following claims. The specification and drawings are, accordingly, to be regarded in an illustrative sense rather than a restrictive sense.

[0027] Embodiments of the invention may include various steps as set forth above. The steps may be embodied in machine-executable instructions which cause a general-purpose

or special-purpose processor to perform certain steps. Alternatively, these steps may be performed by specific hardware components that contain hardwired logic for performing the steps, or by any combination of programmed computer components and custom hardware components. Elements of the present invention may also be provided as a machine-readable medium for storing the machine-executable program code. The machine-readable medium may include, but is not limited to, floppy diskettes, optical disks, CD-ROMs, and magneto-optical disks, ROMs, RAMs, EPROMs, EEPROMs, magnetic or optical cards, or other type of media/machine-readable medium suitable for storing electronic program code.

[0028] Throughout the foregoing description, for the purposes of explanation, numerous specific details were set forth in order to provide a thorough understanding of the invention. It will be apparent, however, to one skilled in the art that the invention may be practiced without some of these specific details. For example, it will be readily apparent to those of skill in the art that the functional modules and methods described herein may be implemented as software, hardware or any combination thereof. Moreover, although some embodiments of the invention are described herein within the context of a client P2P application, the underlying principles of the invention may be implemented in the form of a server application or any other form of client application. Accordingly, the scope and spirit of the invention should be judged in terms of the claims which follow.

CLAIMS

1. A computer-implemented method comprising:

configuring one or more servers within the perimeter of a service provider network;

receiving a request from a first mobile device to establish a real-time communication session with a second mobile device;

providing the first and second mobile devices with networking information for connecting to the servers; and

establishing the real-time communication session through the server.

2. The method as in claim 1 wherein the perimeter of a service provider network comprises a service subnetwork of the service provider.

3. The method as in claim 1 wherein receiving further comprises:

receiving the request at a server located outside of the service provider network, and

registering the one or more servers with the server located outside of the service provider network.

4. The method as in claim 1 wherein the real-time communication session comprises a real-time, two-way video communication session.

5. The method as in claim 1 wherein the networking information includes an IP address of the one or more servers.

6. The method as in claim 1 wherein the service provider comprises a cellular service provider.

7. The method as in claim 1 wherein two or more servers are used to establish the real-time communication session, wherein a first one of the servers (a) communicates directly with a first one of the mobile devices and a second one of the servers communicates directly with a second one of the mobile devices and wherein the first and second servers establish socket connections between one another, or (b) receives an input stream from a first one of the mobile devices and routes the input stream to a second one

of the mobile devices and wherein a second one of the servers receives an input stream from the second one of the mobile devices and routes the input stream to the first one of the mobile devices.

8. The method as in claim 1 wherein the server comprises a relay server.

9. The method as in claim 1 wherein the server is configured on a phone subnetwork to which the first and second mobile devices are connected.

10. A machine-readable medium having program code stored thereon which, when executed by a machine, causes the machine to perform the operations of the method of any of claims 1 to 9.

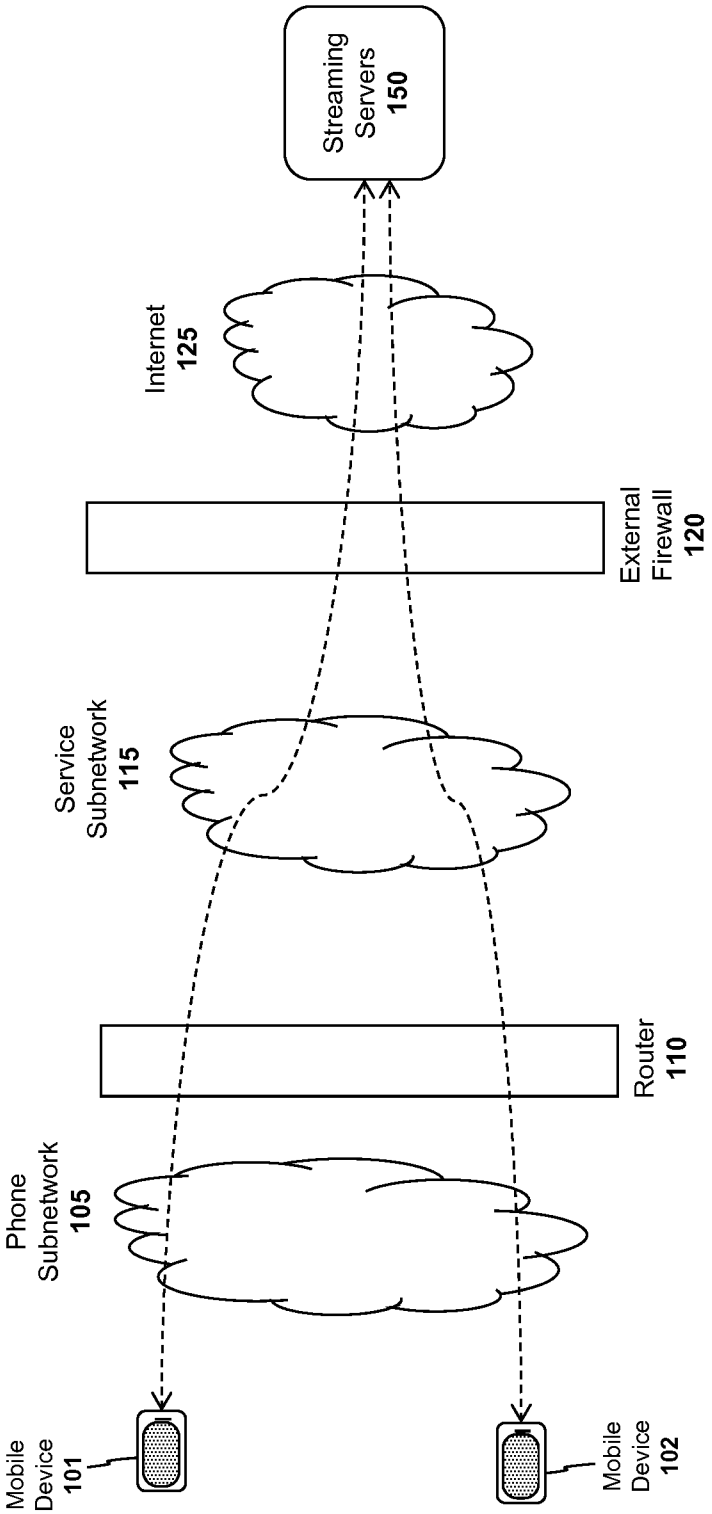

Fig. 1a
(prior art)

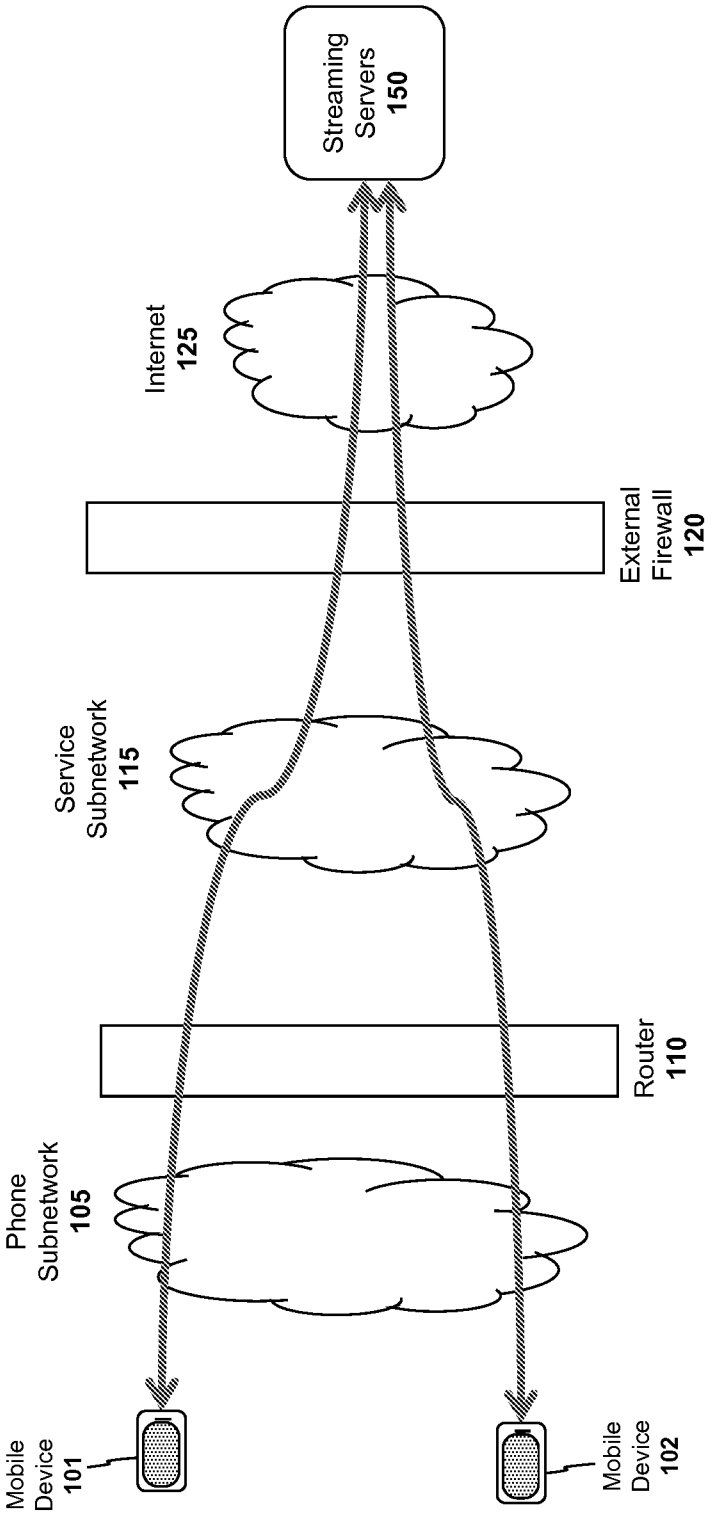

Fig. 1b
(prior art)

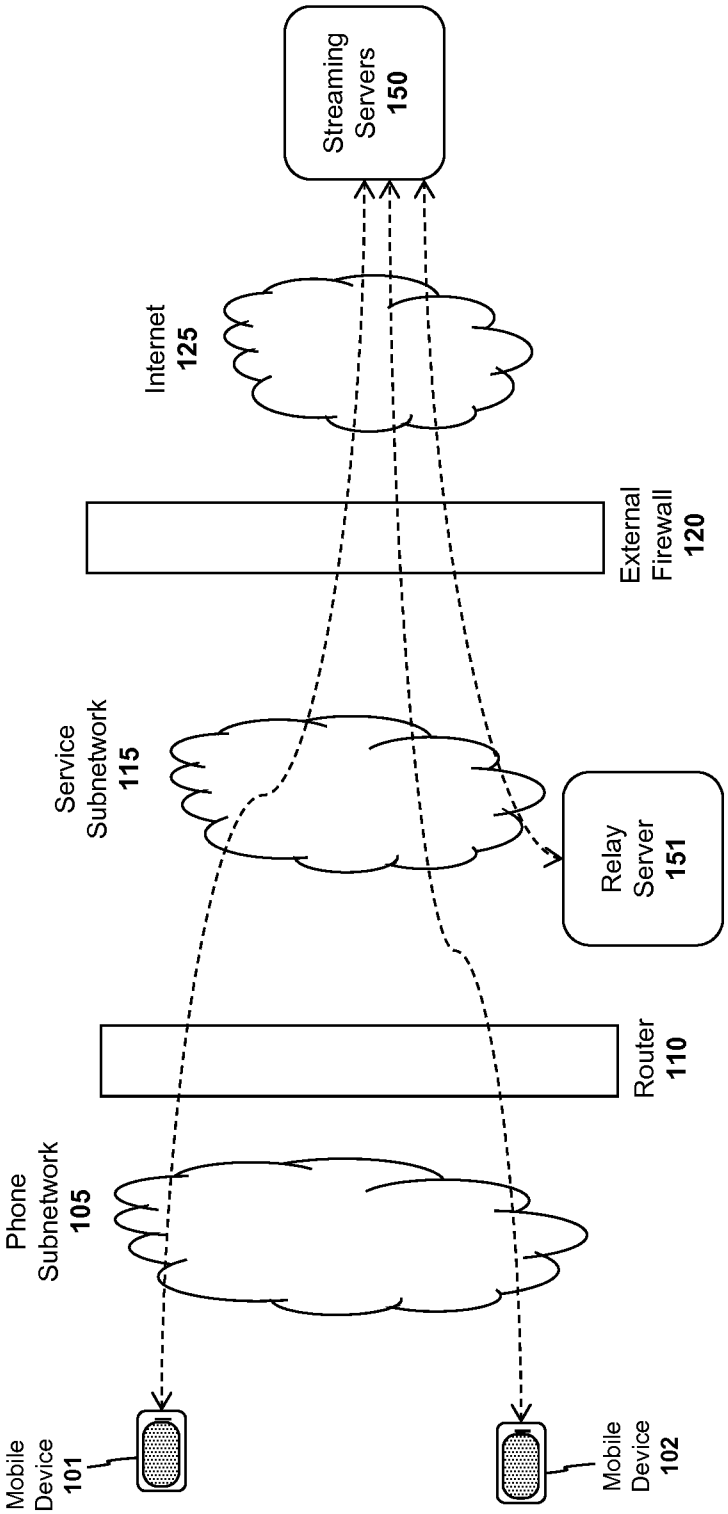
Streaming Servers 150
Internet 125
External Firewall 120
Service Subnetwork 115
Relay Server 151
Router 110
Phone Subnetwork 105
Mobile Device 101
Mobile Device 102

Fig. 2a

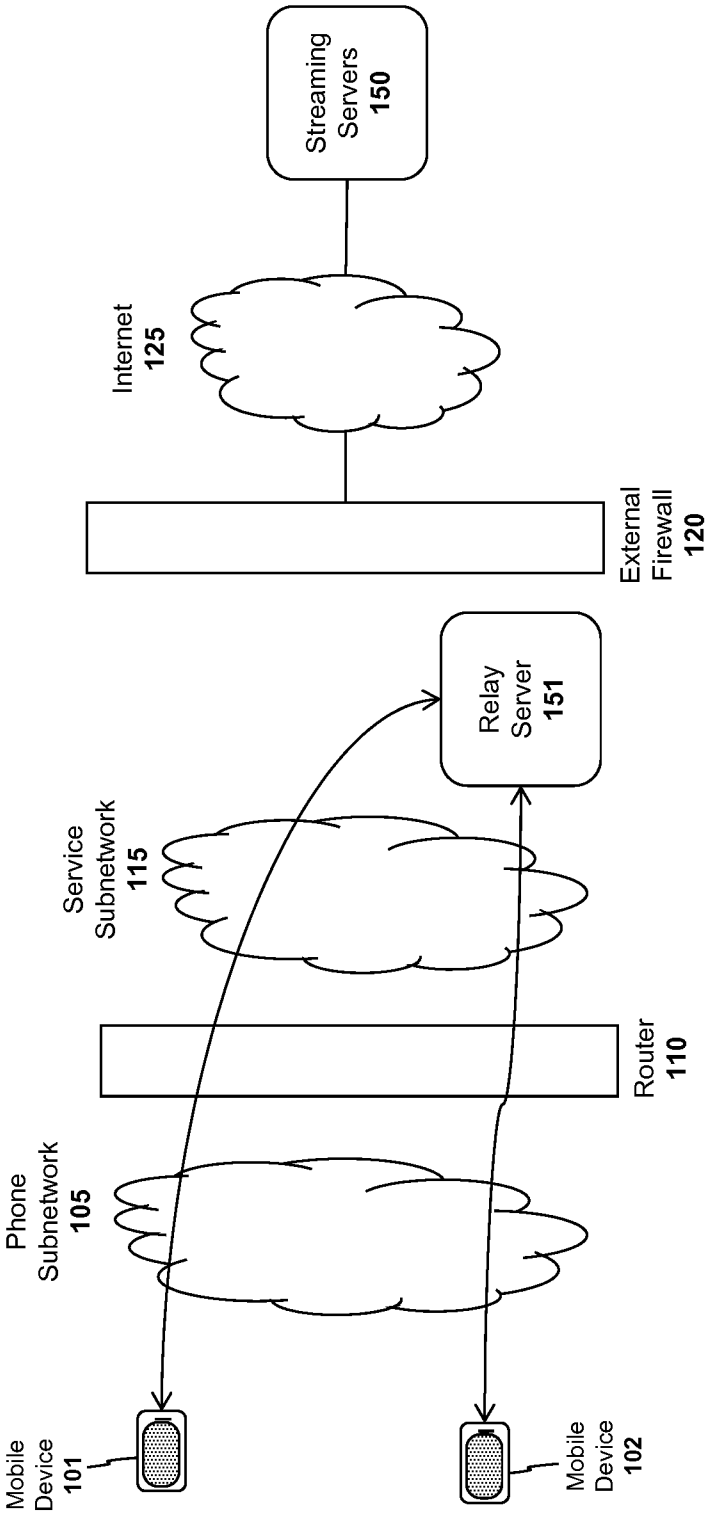
Streaming Servers 150
Internet 125
External Firewall 120
Relay Server 151
Service Subnetwork 115
Router 110
Phone Subnetwork 105
Mobile Device 101
Mobile Device 102

Fig. 2b

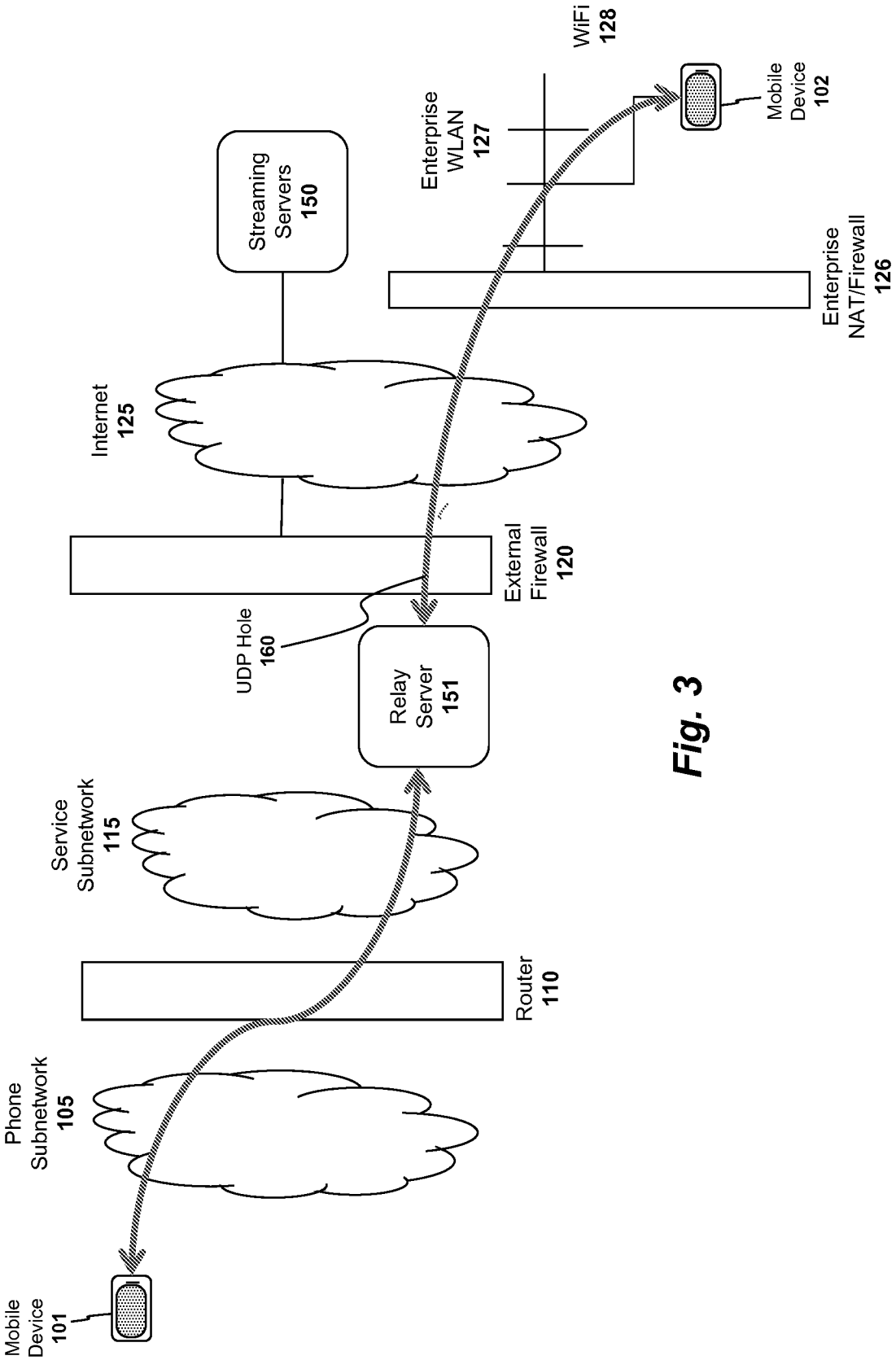
Mobile Device 101
Phone Subnetwork 105
Service Subnetwork 115
Router 110
Relay Server 151
UDP Hole 160
External Firewall 120
Internet 125
Streaming Servers 150
Enterprise WLAN 127
WiFi 128
Mobile Device 102
Enterprise NAT/Firewall 126

Fig. 3

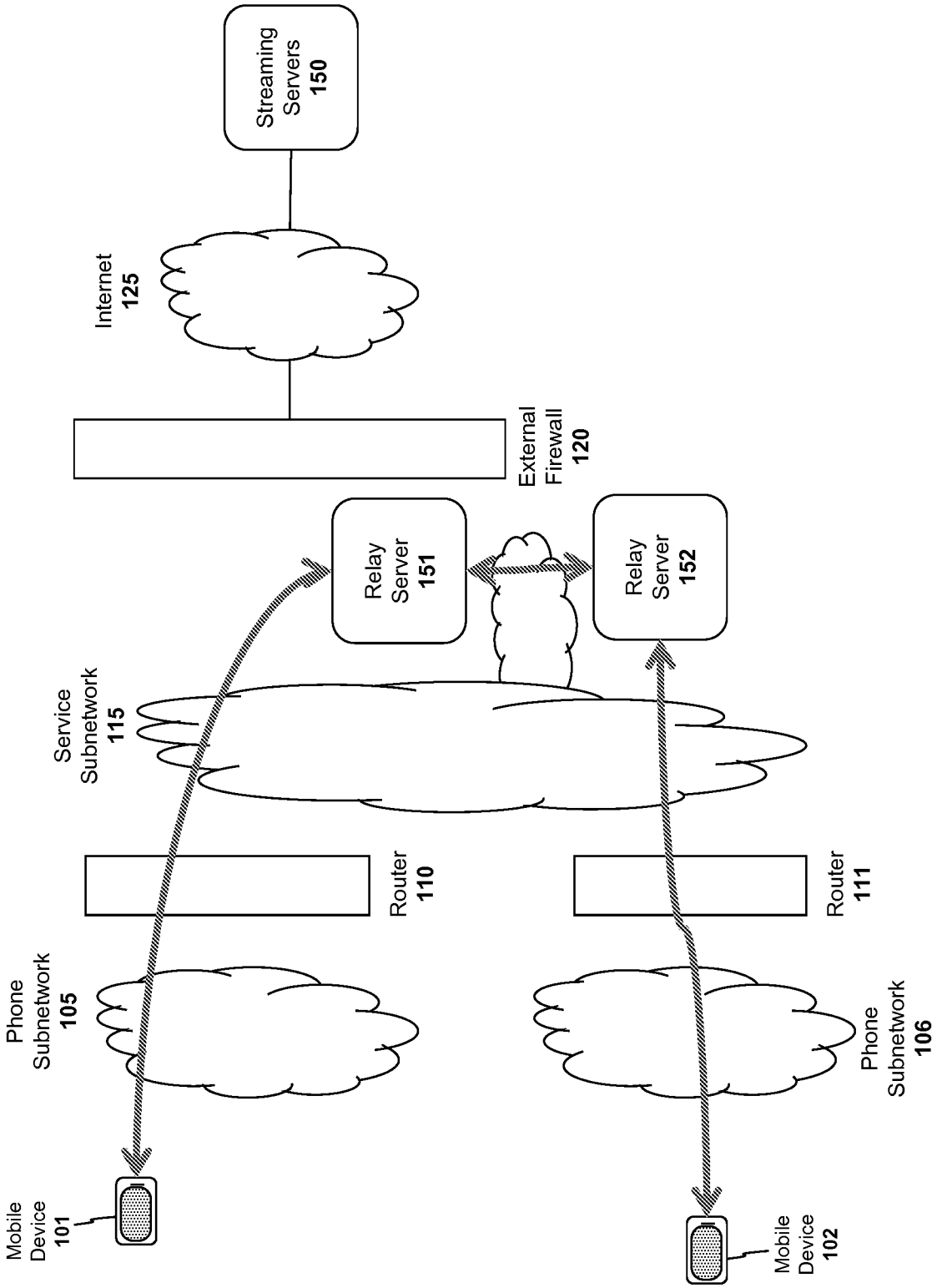
Streaming Servers 150
Internet 125
External Firewall 120
Relay Server 151
Relay Server 152
Service Subnetwork 115
Router 110
Router 111
Phone Subnetwork 105
Phone Subnetwork 106
Mobile Device 101
Mobile Device 102

Fig. 4a

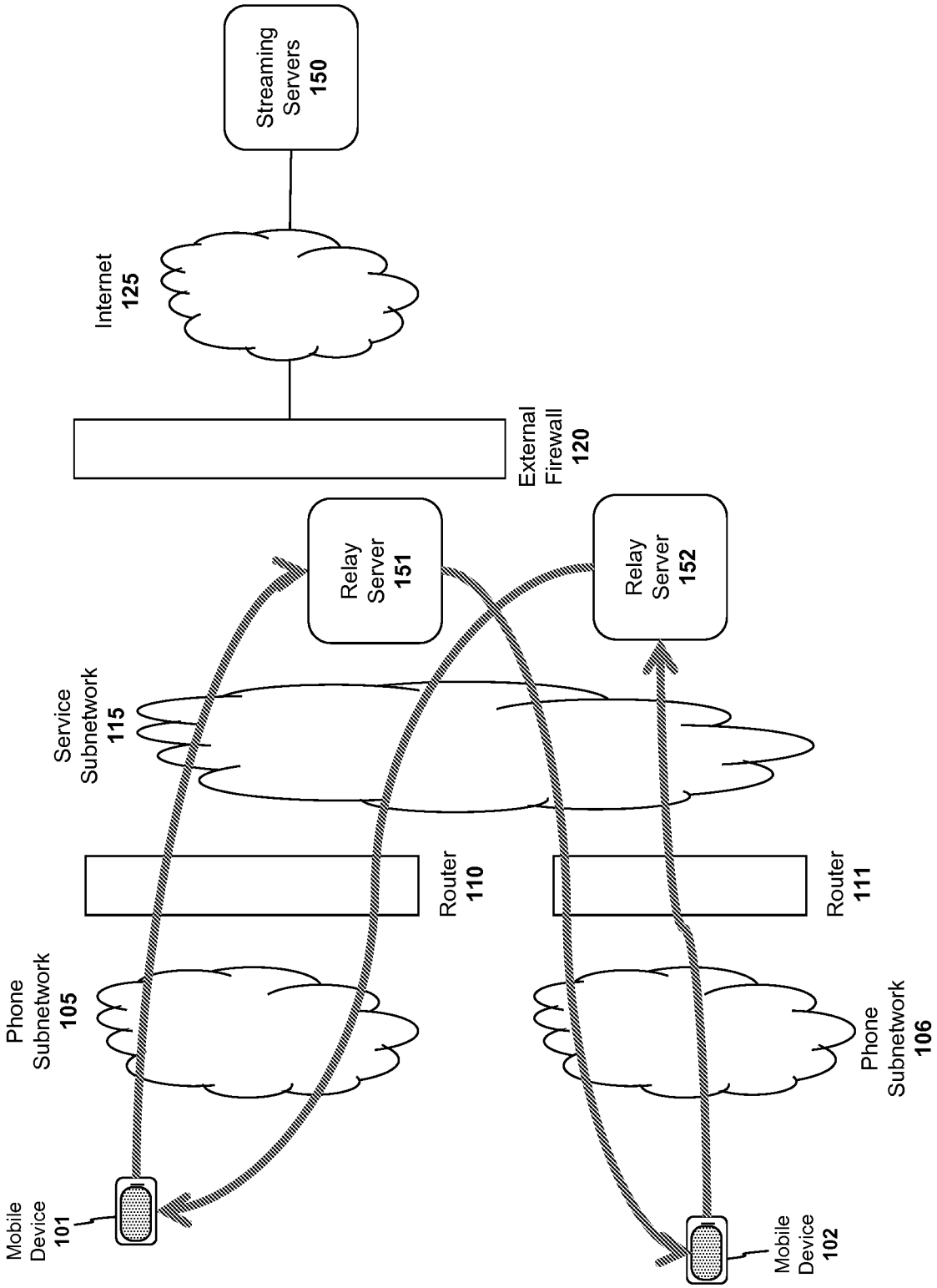
Streaming Servers 150
Internet 125
External Firewall 120
Relay Server 151
Relay Server 152
Service Subnetwork 115
Router 110
Router 111
Phone Subnetwork 105
Phone Subnetwork 106
Mobile Device 101
Mobile Device 102

Fig. 4b

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/045880

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04L29/06
ADD.

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)
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 practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Poikselkä, M. and Mayer, G.: "The IMS: IP Multimedia Concepts and Services", 1 January 2009 (2009-01-01), John Wiley and Sons, XP002685037, ISBN: 978-0-470-72196-4 page 3 - page 4 Section 2.1.1; page 15 Section 2.1.10; pages 21-22; figure 2.6 page 48 - page 51; figures 3.1, 3.3 page 58 page 249 - page 250 page 380; figure 12.13 ----- -/--	1-10

[X] Further documents are listed in the continuation of Box C.

[X] See patent family 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 application or patent 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	Date of mailing of the international search report
11 October 2012	29/10/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Konrad, Markus

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/045880

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 770 948 A1 (AT & T CORP [US]) 4 April 2007 (2007-04-04) paragraph [0016] - paragraph [0018]; figure 1 -----	1-10
A	US 7 437 465 B2 (ROUSSEAU OLIVIER [FR]) 14 October 2008 (2008-10-14) column 2 - column 5; figures 1-6 -----	1-10

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2012/045880

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1770948 A1	04-04-2007	CA 2561213 A1	29-03-2007
		EP 1770948 A1	04-04-2007
		US 2007070982 A1	29-03-2007
US 7437465 B2	14-10-2008	AT 313904 T	15-01-2006
		DE 60302840 T2	31-08-2006
		EP 1389863 A1	18-02-2004
		FR 2843666 A1	20-02-2004
		US 2004081160 A1	29-04-2004