



- (51) **International Patent Classification:**
G06F 15/173 (2006.01)
- (21) **International Application Number:**
PCT/US20 16/0 14442
- (22) **International Filing Date:**
22 January 2016 (22.01 .2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant: NOKIA SOLUTIONS AND NETWORKS OY** [FI/FI]; Karaportti 3, FIN-02610 Espoo (FI).
- (72) **Inventors; and**
- (71) **Applicants (for SC, TT only): CHANDRAMOULI, Devaki** [US/US]; 3532 Thorp Springs Drive, Piano, TX 75025 (US). **LIEBHART, Rainer** [DE/DE]; Leonistrasse 4, 81476 Munich (DE).
- (74) **Agents: GOLDHUSH, Douglas, H.** et al; Squire Patton Boggs (US) LLP, 8000 Towers Crescent Dr., 14th Floor, Vienna, VA 22182-6212 (US).
- (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, BN, 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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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, ST, 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, KM, ML, MR, NE, SN, TD, TG).

Published:
— with international search report (Art. 21(3))



WO 2017/127102 A1

(54) **Title:** APPLICATION RELOCATION BETWEEN CLOUDS

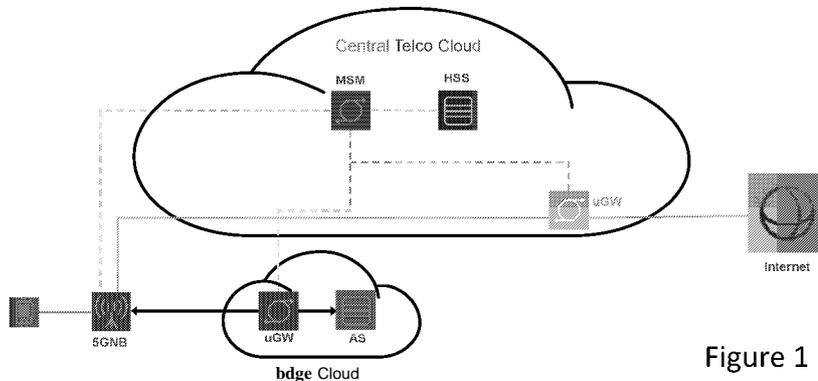


Figure 1

(57) **Abstract:** Various communication systems may benefit from improved handover procedures. For example, certain handover procedures involving edge clouds in a 5G or LTE network may benefit from a more seamless handover during communication. A method may include receiving at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, where the application server is running an application session for the user equipment. The method may also include transferring the application session to the second application server.

TITLE:

APPLICATION RELOCATION BETWEEN CLOUDS

BACKGROUND:

Field:

[0001] Various communication systems may benefit from improved handover procedures. For example, certain handover procedures involving edge clouds in a 5G or LTE network may benefit from seamless handover.

Description of the Related Art:

[0002] 5th generation (5G) is a new generation of radio systems and network architecture that can deliver extreme broadband and ultra-robust, low latency connectivity. 5G also allows for massive machine-to-machine connectivity for the Internet of Things (IoT). The 5G architecture can aid the programmable world, and help transform modern economies and societies.

[0003] 5G provides several areas of improvement. First, 5G can be used to provide massive broadband that delivers gigabytes of bandwidth per second on demand, in both uplink and downlink transmissions. Second, 5G can aid in machine-type communication that allows for immediate synchronous eye-hand feedback. For example, extreme low end-to-end (E2E) latency can aid with remote control of robots and cars. Third, 5G facilitates a massive machine-type communication that can connect billions of sensors and machines.

[0004] 5G is also designed to support a wide diversity of use cases. 5G may not only be a "new radio access technology family," but its architecture will expand to multiple dimensions by providing a common core for multiple radio technologies, such as cellular, fixed, and wireless local area network. The 5G core can also provide for multiple services, such as IoT, mobile broadband, and low-latency high reliability

service, as well as multiple network and service operators.

[0005] LTE (Long Term Evolution) or the E-UTRAN (Evolved Universal Terrestrial Access Network) is the access part of the Evolved Packet System (EPS). LTE can help to address user demands for higher data rate and quality of service. By providing high spectral efficiency, high peak data rates, short round trip time, as well as flexibility in frequency and bandwidth, LTE can help improve the user experience. The LTE may also provide various other benefits including packet switch optimization, continued demand for cost reduction, low complexity, and avoiding unnecessary fragmentation of technologies for paired and unpaired band operation.

SUMMARY:

[0006] A method, in certain embodiments, may include receiving at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, where the application server is running an application session for the user equipment. The method can also include transferring the application session to the second application server.

[0007] According to certain embodiments, an apparatus may include at least one memory including computer program code, and at least one processor. The at least one memory and the computer program code may be configured, with the at least one processor, to cause the apparatus at least to receive at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, wherein the application server is running an application session for the user equipment. The at least one memory and the computer program code may also be configured, with the at least one processor, to cause the apparatus at least to transfer

the application session to the second application server.

[0008] An apparatus, in certain embodiments, may include means for receiving at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, where the application server is running an application session for the user equipment. The apparatus may also include means for transferring the application session to the second application server.

[0009] According to certain embodiments, a non-transitory computer-readable medium encoding instructions that, when executed in hardware, perform a process. The process may include receiving at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, where the application server is running an application session for the user equipment. The process may also include transferring the application session to the second application server.

[0010] According to certain embodiments, a computer program product encoding instructions for performing a process according to a method including receiving at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, where the first application server is running an application session for the user equipment. The method may also include transferring the application session to the second application server.

[0011] A method, in certain embodiments, may include receiving a handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, where the first application server is running an application session for the user equipment. The method may also include sending the first application server a handover notification that the user equipment is moving to a

radio network being served by a second application server in a second cloud.

[0012] According to certain embodiments, an apparatus may include at least one memory including computer program code, and at least one processor. The at least one memory and the computer program code may be configured, with the at least one processor, to cause the apparatus at least to receive handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, where the first application server is running an application session for the user equipment. The at least one memory and the computer program code may also be configured, with the at least one processor, to send the first application server a handover notification that the user equipment is moving to a radio network being served by a second application server in a second cloud.

[0013] An apparatus, in certain embodiments, may include means for receiving a handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, where the first application server is running an application session for the user equipment. The apparatus may also include means for sending the first application server a handover notification that the user equipment is moving to a radio network being served by a second application server in a second cloud.

[0014] According to certain embodiments, a non-transitory computer-readable medium encoding instructions that, when executed in hardware, perform a process. The process may include receiving a handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, where the first application server is running an application session for the user equipment. The process may also include sending the first application server a handover notification that the user equipment is moving to a radio network being served by a second application server in a second cloud.

[0015] According to certain embodiments, a computer program product encoding

instructions for performing a process according to a method including receiving a handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, where the first application server is running an application session for the user equipment. The method may also sending the first application server a handover notification that the user equipment is moving to a radio network being served by a second application server in a second cloud.

BRIEF DESCRIPTION OF THE DRAWINGS:

[0016] For proper understanding of the invention, reference should be made to the accompanying drawings, wherein:

[0017] Figure 1 illustrates an exemplary network architecture illustrating cloud technology.

[0018] Figure 2 illustrates a signal flow diagram according to certain embodiments.

[0019] Figure 3 illustrates a signal flow diagram according to certain embodiments.

[0020] Figure 4 illustrates an exemplary network architecture illustrating cloud technology.

[0021] Figure 5 illustrates a signal flow diagram according to certain embodiments.

[0022] Figure 6 illustrates a signal flow diagram according to certain embodiments.

[0023] Figure 7 illustrates a flow diagram according to certain embodiments.

[0024] Figure 8 illustrates a flow diagram according to certain embodiments.

[0025] Figure 9 illustrates a system according to certain embodiments.

DETAILED DESCRIPTION:

[0026] In uses requiring extreme low latency applications, for example, below 10 milliseconds, a low E2E delay can be achieved with a new 5G radio interface design. The low latency E2E may also need the network application to reside close to the application client residing on the user equipment (UE). To do so, a gateway, which

terminates the UE's bearer, and the network application providing the service can both reside in an edge cloud. The edge cloud may either be provided close to the 5G base station (5GNB), or directly at or on the 5GNB platform serving the UE. In another embodiment, the edge cloud may be provided close to an LTE evolved NodeB (eNB), or directly at or on the eNB platform service the UE.

[0027] Edge clouds help to facilitate the proximity of the network applications and the gateway to the serving UE. Mobile Edge Computing (MEC), which can sometimes be synonymous to edge clouds, can also help to provide network applications close to the end user. Unlike a Telco cloud, which may be built from core network entities, the edge clouds may be built from locally distributed data centers.

[0028] Figure 1 illustrates an exemplary network architecture of 5G cloud technology. Specifically, Figures 1 illustrates a central Telco cloud, including a 5G Mobility and Session Management entity (MSM) and a Home Subscriber Server (HSS). In Long Term Evolution (LTE) terms, the MSM can be a combination of a mobility management entity (MME), Serving Gateway (SGW) and Packet Data Network Gateway (PGW) control plane parts, and a Policy and Charging Rules Function (PCRF).

[0029] The edge cloud can also include a user plane gateway (uGW) that can be the distributed user plane gateway, which may be controlled by the MSM. In LTE terms, the uGW may be the Serving Gateway (SGW) and a Packet Data Network Gateway (PGW) user plane parts.

[0030] The uGW located in the edge cloud acts as an internet protocol (IP) anchor point. The UE's IP address can be assigned from the uGW's address pool or from the MSM or any other entity. As a result, if the UE moves to a new location closer to a different edge cloud, having a different gateway and a different application server, the UE may need to undergo handover. Some embodiments can ensure that the

application session is maintained without interruption, even when the application session may need to be transferred to a different application server.

[0031] Certain embodiments provide for a seamless continuation of an application session for a UE when servicing of the UE moves from a first edge cloud to a second edge cloud. In particular, the gateway in the second edge cloud may seamlessly begin to host the UE's IP address, while avoiding interruptions or breaks in the application session.

[0032] Some embodiments can provide a method for selecting the second edge cloud, including the second gateway and the second application server (AS). A procedure for efficiently transferring context data from the first gateway, in the first edge cloud, to a new gateway may also be provided. In addition, an efficient procedure for transferring context data from the first AS, in the first edge cloud, to the second AS may be provided.

[0033] In the case where a session application for a UE requires handover from the first edge cloud to the second edge cloud, the first AS, which is currently serving the UE, may be informed by the mobile network about the UE moving from the first edge cloud to the second edge cloud. The application session can then be transferred by different application techniques to the second or target AS, in the second edge cloud. Such a seamless handover can allow the application session for the UE to continue without any interruptions or breaks.

[0034] Figure 2 illustrates a signal flow diagram according to certain embodiments. Specifically, Figure 2 illustrates a signal flow diagram for relocating the uGW and AS between two edge clouds when the UE is moving from the serving area of one edge cloud into the serving area of another edge cloud. In addition, in the embodiment of Figure 2, the AS may initiate the session setup. In step 201, the UE can connect to a low latency application session. The first AS in the first edge cloud can subscribe to the MSM for mobility management, if the specific application

session in use has a need for seamless handover. As shown in Figure 1, the MSM may be located in the central Telco cloud. In other embodiments, the first AS may subscribe to any other network entity, located either in the Telco cloud or the edge cloud that are aware and/or are involved in the handover decisions. For example, the first AS may also subscribe to the 5GNB or to an LTE eNB.

[0035] In order for the first AS to be able to send a request to subscribe for mobility management events to the MSM, the first AS may need to be able to identify to which MSM the event notification request must be sent. The first AS may contact the correct MSM through various methods. For example, as the uGW in the edge cloud acts to terminate the bearer for the UE and allocate the UE IP address, it can become aware of the desired MSM identifier. The uGW in the edge cloud can therefore provide a MSM identifier, such as an IP address or a name, to the first AS. Alternatively, the first AS can also make a database lookup to receive the MSM address or send the request to a default MSM which further routes the request to the correct MSM. Because the first AS can be hosted in the same edge cloud as uGW, the Access Point Name (APN) of the first AS may be known to the uGW. In other embodiments, the first AS may be known to the uGW through other identification.

[0036] In certain embodiments, a tunnel between the first AS and the uGW can be established to ensure proper routing between the first AS and the uGW. In other embodiments, the routing path between the first AS and the uGW in the edge cloud can be created with the help of Software Defined Networking (SDN) mechanisms. For example, the MSM may instruct SDN to setup such a path.

[0037] Alternatively, the first AS may be configured with a logical MSM address that is resolved together with the UE identifier, such as MSISDN, IMSI, or optionally an application ID. The first AS can use the UE identifier to determine the actual MSM address serving the UE. Determining the MSM address may be included in the event notification request message sent via the Application

Programming Interface (API) from the first AS to the MSM. As such, an API gateway can translate and/or direct the event notification request message to the correct MSM. The HSS may be in the Telco cloud, where the MSM is also stored, can help in sending the message to the correct MSM using the API.

[0038] As described above, the first AS may subscribe to the MSM for HO event notification via an API, which can be provided by the MSM. The AS may then use the API to request to subscribe to the MSM for handover event notification. The handover event notification request from the first AS to the MSM may include a UE identifier. For example, the UE identifier may be a Mobile Subscriber Integrated Services Digital Network Number (MSISDN), an International Mobile Subscriber Identity (IMSI), or any other identifier that has previously been stored in the HSS, or another database, and may be known to both the MSM and the AS. The UE identifier may be used by the MSM to identify the UE.

[0039] In certain embodiments, the MSM may not need to identify the particular application. The UE identifier can be different from the application identifier (ID) used on the application layer. In some embodiments, however, the network may already have knowledge of the application ID. For example, some relationship can be assumed between the mobile network operator and the application service providers, in which application specific information, such as application ID, can be shared. The application ID may therefore be stored in a network entity in the mobile network. In certain embodiments the network entity may be an HSS.

[0040] Once the first AS has subscribed to the MSM, the MSM may inform the first AS of certain mobility events of the UE. One such event is an ongoing handover of the UE. When subscribing to the MSM, the first AS may subscribe to all handovers of a particular UE, or to all handovers of a particular set of UEs.

[0041] Alternatively, the AS may selectively subscribe to only some handovers. For example, the AS may only subscribe to a UE handover in a certain area. The area

may be defined as a set of cells, tracking area(s), presence area(s), or edge cloud service area(s). The edge cloud service area can consist of all 5GNB or LTE eNB that are connected to a particular edge cloud. Once the first AS specifies an area, the MSM may only inform the first AS about an ongoing handover if the UE moves from one such area to another area.

[0042] To accurately inform the first AS of UE movement between areas, the MSM may maintain the edge cloud service area, and a mapping of the edge cloud service area to corresponding 5GNBs and/or cell IDs, for example evolved cell global identifiers (ECGIs). In certain embodiments, the MSM may also be aware of area changes of a UE. The MSM may inform the UE and/or the first AS about the edge cloud service area mapping list. When the UE is informed of the edge cloud service area, it may notify the MSM and/or first AS when it moves from one edge cloud service area to another edge cloud service area. The UE can send such a notification when the UE is in connected or idle mode. This can help the MSM and/or the AS to determine accurately when such area changes occur due to both idle mode mobility and connected mode handover.

[0043] In step 201 of Figure 2, the UE is connected to a low latency application session, and the first AS has subscribed to the handover notifications at the MSM. An old 5GNB, which currently serves the UE, may detect that the UE needs to handover, in step 210. The old 5GNB may then initiate a request to a new target 5GNB, knowing that the new target 5GNB may be involved in the handover. In other embodiments, an old LTE eNB may initiate a request to a new target LTE eNB. In step 211, the handover prep request is sent from the old 5GNB, which currently serves the UE, to the new target 5GNB. The new target 5GNB can then prepare for the handover. The new target 5GNB can then send the request to the MSM, as shown in step 212. The handover request can include information about the old 5GNB and the new target 5GNB.

[0044] In step 213, if the first AS subscribed to the handover event, the MSM notifies the first AS about the handover request. The MSM can include in the handover request information regarding the new 5GNB, the UE identity, such as MSISDN, IMSI, or application ID, and the UE IP address. The MSM can also include information about the second AS. Receiving at least a part of this information can allow the first AS to adequately determine if it serves the new target 5GNB, and if the uGW connected to the first AS also serves the new target 5GNB. If not, meaning that the first AS and uGW do not serve the new target 5GNB, the first AS may provide the UE context information to the second AS in the new 5GNB serving area, as shown in step 214. In doing so, the first AS transfers application specific session data to the second AS.

[0045] In some embodiments, the first AS may use the MSM as an intermediary when providing the UE context to the second AS. In other words, the first AS may send the UE context to the MSM, which can then forward the context to the second AS. In other embodiments, the first AS may directly send the UE context information to the second AS.

[0046] Figure 3 illustrates a signal flow diagram according to certain embodiments. Once the second AS receives the application specific session data from the first AS, the second AS may inform the MSM about the transfer, and the second AS may initiate a session setup for the UE with the MSM, as shown in step 305. This session setup request message sent from the second AS to the MSM can include a UE ID, an application ID, and some information about the new gateway in the second edge cloud. The second AS can initiate the setup to ensure a make between the first AS and the second AS before a break occurs for the handover. This may allow the network to avoid a communication disruption caused by a change in the IP address from the first AS to the second AS.

[0047] However, in some embodiments, where the MSM may be used as an

intermediary in the transfer, step 305 may not be necessary, since the MSM would already know of the transfer of context information between be informed of the transfer.

[0048] In step 306, the MSM sets up the routing/context table in the new gateway in the second edge cloud. The routing/context may be UE specific, and may include the UE ID, UE IP address, information about the second AS, including the address of the second AS, and other relevant information. In step 307, the MSM similarly sets up routing/context table within the second AS, also located in the second edge cloud. The routing/context may be UE specific, and may include the UE ID, UE IP address, information about the new gateway in the second edge cloud, and other relevant information.

[0049] The MSM, in certain embodiments, can therefore inform the new GW of the second AS, and the second AS of the new GW, and help to facilitate the routing of packets between the two in the second edge cloud. UE context in new GW and second AS can be used to map uplink or downlink packets to the proper tunnel between the new GW and second AS. In some other embodiments, MSM can indicate to an SDN flow control device (SDN-C) that the UE IP is being served by a new GW, at which point the SDN-C may update switches and/or routers in the second edge cloud accordingly.

[0050] Figure 4 illustrates an exemplary network architecture illustrating cloud technology. In certain embodiments, a tunnel 401 may be established between the uGW 402 and the second AS 403 in the second edge cloud. Tunnel 401 may be established based on APN, as discussed above. Packets can then be mapped to tunnel 401 based on the IP address of UE 404 in the IP header of the data packets.

[0051] In certain embodiments, proper routing between uGW 402 and the second AS 403 may be ensured by at least one of L2 switching, L3 routing at uGW 402, L2/L3 tunnel between uGW and AS based on the APN, or uGW and AS collocated

on same platform in the second edge cloud using direct communication. The direct communication may be an inter-process or an inter-platform communication. In some embodiments, for L2 switching and L3 routing at uGW, the SDN may be used to update routers and/or switches in the second edge cloud.

[0052] In the embodiments of Figure 3, once the MSM sets up the routing tables in the new GW and the second AS, it may then send a handover response to the new 5GNB, in step 308. In step 309, the new 5GNB can then forward or send the handover response to the original or old 5GNB. The original or old 5GNB may then send a handover command to the UE. At this point, the new user plane path may already be established between the UE, the new 5GNB, and the second edge cloud, which includes both the new uGW and the second AS. This user plane path can help facilitate seamless handover of the application session to the second edge cloud without a service interruption to the UE.

[0053] In step 311, the UE can send a confirmation message to the new 5GNB confirming the handover, and the new 5GNB may then notify the MSM of the handover, in step 312. Once the MSM is notified, resources in the old 5GNB, old GW, and first AS may be released.

[0054] Figure 5 illustrates a signal flow diagram according to certain embodiments. Specifically, Figure 5 illustrates a signal flow diagram in which the MSM initiates the session setup. In step 510, the old 5GNB, which currently serves the UE, can detect that the UE may need a handover, and requests that a new target 5GNB to be prepared for a handover. In step 511 the old 5GNB may initiate and send the handover preparation request to the new target 5GNB. In step 512 the new 5GNB can send the handover request, including information about the old 5GNB and the new target 5GNB, to the MSM.

[0055] In certain embodiments, the MSM may need to know that the new 5GNB is served by the second edge cloud. Information pertaining to a relationship between

the second edge cloud and the new 5GNB may be configured in the MSM or fetched from a database in the Telco cloud, in which the MSM is located.

[0056] In some embodiments, the MSM can also select the new uGW and the second AS in the second edge cloud serving the new target 5GNB. The MSM may first determine if the first AS serves the new target 5GNB. If not, then the MSM may then select the second AS to initiate session setup. Information pertaining to the new uGW and the second AS can be configured in the MSM, or a database to which the MSM has access. In other embodiments, the new 5GNB may select the uGW and inform the MSM of its selection. For example, in step 512 the new 5GNB may include an indication of which uGW it has selected.

[0057] In step 513, the MSM can notify the first AS of the handover. In the notification message, the MSM may include information pertaining to the new 5GNB, UE ID, UE IP, and new gateway, and the second AS. In certain embodiments, the MSM may not be able to provide the first AS with the second AS address. In such embodiments, the first AS may be configured with data allowing it to select a second AS serving the new target 5GNB. Based on a logical second AS name, together with an ID of the second edge cloud provided by the MSM, the first AS may be able to determine the second AS address, without the MSM explicitly providing such information.

[0058] In the embodiment shown in Figure 5, the MSM may indicate to the SDN infrastructure that the IP address of the UE can be served by the new uGW, and can adapt the routing paths in the second edge cloud accordingly. In certain embodiments, the IP address of the UE may not be changed during handover. The routing paths adapted by the MSM, with help from the SDN, in the second edge cloud can be Layer 2 or Layer 3 connections, as described above in the discussion of Figure 4. In certain embodiments, the MSM, with help from the SDN, can update the second edge cloud switches or routers to ensure proper routing of packets between

the new gateway and the second AS. Yet in other embodiments, the MSM can provide the new gateway with information pertaining to the second AS and the UE IP.

[0059] Figure 6 illustrates a signal flow diagram according to certain embodiments. Specifically, Figure 6 illustrates a signal flow diagram in which the MSM initiates the session setup. In contrast to Figure 3, the MSM in the embodiment of Figure 6 may trigger the routing table/context setup. In some other embodiments, the SDN-C may trigger the routing table/context setup. In steps 614 and 615 the MSM sets up the routing tables in the new gateway and the second AS. The second AS can then initiate application session context retrieval with the first AS, as shown in step 616. In step 617, the MSM may send a handover response to the new target 5GNB. The new target 5GNB can then send the handover response to the old 5GNB, in step 618, which can then send a handover command to the UE, in step 619.

[0060] Since the new user plane between the UE, new target 5GNB, the new GW, and the second AS may already be established, routing of the application session can occur without disruption of service. In step 620, the UE can confirm the handover towards the new 5GNB, which may then notify the MSM, in step 621. Resources allocated by the old 5GNB, old uGW, and first application server can then be released.

[0061] Figure 7 illustrates a flow diagram according to certain embodiments. In step 701, the first AS in a first edge cloud may receive from a user plane gateway an identifier of the network entity for mobility management, such as MSM. In step 702, the first AS may send the network entity for mobility management a subscribe request. In the request, the first AS may specify which handover notifications it would like to receive.

[0062] In step 704, the first AS may receive from the MSM, or any other network entity, a handover notification that the UE is moving to a radio network provided by,

for example, a 5GNB or LTE eNB, being served by a second AS located in a second edge cloud. In step 705, the first AS may receive address information relating to the second application server. The address information may also be included in the handover notification sent to the first AS from the MSM. The first AS can then transfer the application session to the second AS, as shown in step 706.

[0063] Figure 8 illustrates a flow diagram according to certain embodiments. In step 801, an MSM may receive a handover event notification request from a first AS. The request may specify to which handover notifications the first AS would like to subscribe. In step 802, the MSM may receive a request for a handover from a new 5GNB. In step 803, the MSM may select a new uGW and/or a second AS belonging to a new 5GNB. In other embodiments, in step 804, the MSM may receive an address for the second AS. The MSM can also send the first AS a handover notification, including information relating to the second AS and the new uGW, as shown in step 805. The MSM can then trigger initiations of the session setup with at least the new uGW or the second AS, as shown in step 806. In step 807, the uGW may instruct a SDN to setup a routing path between the new uGW and the second AS.

[0064] Figure 9 illustrates a system according to certain embodiments. It should be understood that each block of the signal flow charts in Figures 2, 3, 5, 6, and the flow charts in Figures 7 and 8, or any combination thereof, may be implemented by various means or their combinations, such as hardware, software, firmware, one or more processors and/or circuitry. In one embodiment, a system may include several devices, such as, for example, a network entity 920 or UE or user device 910. The system may include more than one UE 910 and more one network entities 920, although only one access node shown for the purposes of illustration. A network entity can be a network node, a base station, an eNB, a GW, a MSM, an AS, a 5GNB, a server, a host or any of the other access or network node

discussed herein.

[0065] Each of these devices may include at least one processor or control unit or module, respectively indicated as 911 and 921. At least one memory may be provided in each device, and indicated as 912 and 922, respectively. The memory may include computer program instructions or computer code contained therein. One or more transceiver 913 and 923 may be provided, and each device may also include an antenna, respectively illustrated as 914 and 924. Although only one antenna each is shown, many antennas and multiple antenna elements may be provided to each of the devices. Other configurations of these devices, for example, may be provided. For example, a network entity 920 and UE 910 may be additionally configured for wired communication, in addition to wireless communication, and in such a case antennas 914 and 924 may illustrate any form of communication hardware, without being limited to merely an antenna.

[0066] Transceivers 913 and 923 may each, independently, be a transmitter, a receiver, or both a transmitter and a receiver, or a unit or device that may be configured both for transmission and reception. The transmitter and/or receiver (as far as radio parts are concerned) may also be implemented as a remote radio head which is not located in the device itself, but in a mast, for example. The operations and functionalities may be performed in different entities, such as nodes, hosts or servers, in a flexible manner. In other words, division of labor may vary case by case. One possible use is to make a network node deliver local content. One or more functionalities may also be implemented as virtual application(s) in software that can run on a server.

[0067] A user device or user equipment 910 may be a mobile station (MS) such as a mobile phone or smart phone or multimedia device, a computer, such as a tablet, provided with wireless communication capabilities, personal data or digital assistant (PDA) provided with wireless communication capabilities, portable

media player, digital camera, pocket video camera, navigation unit provided with wireless communication capabilities or any combinations thereof.

[0068] In some embodiment, an apparatus, such as an access node, may include means for carrying out embodiments described above in relation to Figures 2, 3, 5, 6, 7, and 8. In certain embodiments, at least one memory including computer program code can be configured to, with the at least one processor, cause the apparatus at least to perform any of the processes described herein.

[0069] According to certain embodiments, an apparatus 920 may include at least one memory 922 including computer program code, and at least one processor 921. The at least one memory 922 and the computer program code are configured, with the at least one processor 921, to cause the apparatus 920 at least to receive at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, where the application server is running an application session for the user equipment. The at least one memory 922 and the computer program code are configured, with the at least one processor 921, to also cause the apparatus 920 at least to transfer the application session to the second application server.

[0070] According to certain embodiments, an apparatus 920 may include means for receiving information about a streaming service, and means for receiving at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, where the application server is running an application session for the user equipment. The apparatus 920 may also include means for transferring the application session to the second application server.

[0071] According to certain embodiments, an apparatus 920 may include at least one memory 922 including computer program code, and at least one processor

921. The at least one memory 922 and the computer program code are configured, with the at least one processor 921, to cause the apparatus 920 at least to receive a handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, where the first application server is running an application session for the user equipment. The at least one memory 922 and the computer program code are configured, with the at least one processor 921, to also cause the apparatus 920 at least to send the first application server a handover notification that the user equipment is moving to a radio network being served by a second application server in a second cloud.

[0072] According to certain embodiments, an apparatus 920 may include means for receiving information about a streaming service, and means for receiving a handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, where the first application server is running an application session for the user equipment. The apparatus 920 may also include means for sending the first application server a handover notification that the user equipment is moving to a radio network being served by a second application server in a second cloud.

[0073] Processors 911 and 921 may be embodied by any computational or data processing device, such as a central processing unit (CPU), digital signal processor (DSP), application specific integrated circuit (ASIC), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), digitally enhanced circuits, or comparable device or a combination thereof. The processors may be implemented as a single controller, or a plurality of controllers or processors.

[0074] For firmware or software, the implementation may include modules or unit of at least one chip set (for example, procedures, functions, and so on). Memories 912 and 922 may independently be any suitable storage device, such as a non-transitory computer-readable medium. A hard disk drive (HDD), random access

memory (RAM), flash memory, or other suitable memory may be used. The memories may be combined on a single integrated circuit as the processor, or may be separate therefrom. Furthermore, the computer program instructions may be stored in the memory and which may be processed by the processors can be any suitable form of computer program code, for example, a compiled or interpreted computer program written in any suitable programming language. The memory or data storage entity is typically internal but may also be external or a combination thereof, such as in the case when additional memory capacity is obtained from a service provider. The memory may be fixed or removable.

[0075] The memory and the computer program instructions may be configured, with the processor for the particular device, to cause a hardware apparatus such as a network entity 920 or UE 910, to perform any of the processes described above (see, for example, Figures 2, 3, 5, 6, 7, and 8). Therefore, in certain embodiments, a non-transitory computer-readable medium may be encoded with computer instructions or one or more computer program (such as added or updated software routine, applet or macro) that, when executed in hardware, may perform a process such as one of the processes described herein. Computer programs may be coded by a programming language, which may be a high-level programming language, such as objective-C, C, C++, C#, Java, etc., or a low-level programming language, such as a machine language, or assembler. Alternatively, certain embodiments may be performed entirely in hardware.

[0076] Furthermore, although Figure 9 illustrates a system including a network entity 920 and UE 910, certain embodiments may be applicable to other configurations, and configurations involving additional elements, as illustrated and discussed herein. For example, multiple user equipment devices and multiple network entities may be present, or other nodes providing similar functionality, such as nodes that combine the functionality of a user equipment and a network

entity, such as a relay node. For example, the UE 910 may be configured for device-to-device communication.

[0077] Certain embodiments described above can allow for seamless handover of an application session between a first edge cloud and a second edge cloud. Some embodiments may allow for the preservation of the UE's IP address in low latency applications. Some embodiments may also be applicable to Mobile Edge Computing (MEC) scenarios.

[0078] The features, structures, or characteristics of certain embodiments described throughout this specification may be combined in any suitable manner in one or more embodiments. For example, the usage of the phrases "certain embodiments," "some embodiments," "other embodiments," or other similar language, throughout this specification refers to the fact that a particular feature, structure, or characteristic described in connection with the embodiment may be included in at least one embodiment of the present invention. Thus, appearance of the phrases "in certain embodiments," "in some embodiments," "in other embodiments," or other similar language, throughout this specification does not necessarily refer to the same group of embodiments, and the described features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

[0079] One having ordinary skill in the art will readily understand that the invention as discussed above may be practiced with steps in a different order, and/or with hardware elements in configurations which are different than those which are disclosed. Therefore, although the invention has been described based upon these preferred embodiments, it would be apparent to those of skill in the art that certain modifications, variations, and alternative constructions would be apparent, while remaining within the spirit and scope of the invention. While some embodiments can be directed to a 5G environment, other embodiments can be directed to an LTE environment.

[0080] Partial Glossary

[0081] API	Application Programming Interface
[0082] APN	Access Point Name
[0083] AS	Application Server
[0084] E2E	End-to-End
[0085] HO	Handover
[0086] HSS	Home Subscriber Server
[0087] IoT	Internet of Things
[0088] IP	Internet Protocol
[0089] M2M	Machine-to-Machine
[0090] MSM	Mobility and Session Management
[0091] MME	Mobility Management Entity
[0092] GW	Gateway
[0093] uGW	user plane Gateway
[0094] MEC	Mobile Edge Computing
[0095] SGW	Serving Gateway
[0096] PCRF	Policy and Charging Rules Function
[0097] MSISDN	Mobile Subscriber Integrated Services Digital Network Number
[0100] IMSI	International Mobile Subscriber Identity
[0101] UE	User Equipment
[0102] SDN	Software Defined Networking

WE CLAIM:

1. A method comprising:

receiving at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, wherein the application server is running an application session for the user equipment; and

transferring the application session to the second application server.

2. The method according to claim 1, further comprising:

subscribing to a network entity for mobility management, wherein the network entity is configured to participate in the handover signaling of the user equipment.

3. The method according to claim 2, further comprising:

sending the network entity for mobility management a request to subscribe for mobility management events including an identity for the user equipment.

4. The method according to claim 2, further comprising:

receiving from a user plane gateway an identifier of the network entity for mobility management.

5. The method according to claim 1, further comprising:
receiving an address for the second application server.

6. The method according to claims 1, wherein the second application server triggers a session setup request for initiating relocation of the application session to the second application server.

7. The method according to claim 1, further comprising:
determining by the first application server that the user equipment is moving to the radio network being served by the second application server.

8. A method comprising:
receiving a handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, wherein the first application server is running an application session for the user equipment;
and
sending the first application server a handover notification that the user equipment is moving to a radio network being served by a second application server in a second cloud.

9. The method according to claim 8, further comprising:
receiving a request for the handover of the application session from the first application server to the second application server.
10. The method according to claim 8, further comprising:
triggering the handover of the application session from the first application server in the first cloud to the second application server.
11. The method according to claim 8, further comprising:
maintaining a service area of the first cloud and a service area of the second cloud; and
maintaining a mapping of the service area of the first cloud and a mapping of the service area in the second cloud.
12. The method according to claim 8, further comprising:
instructing a software defined networking mechanism to setup a routing path between a user plane gateway and the first application server.
13. The method according to claim 8, further comprising:
selecting a new user plane gateway or the second application server based

on information received from a network entity.

14. The method according to claim 8, further comprising:
triggering a session setup request for initiating relocation of the application session to the second application server.

15. The method according to claim 14, further comprising:
sending the session setup request to a new user plane gateway and the second application server.

16. An apparatus comprising:
at least one memory comprising computer program code;
at least one processor;
wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:
receive at a first application server located in a first cloud a handover notification that a user equipment is moving to a radio network being served by a second application server located in a second cloud, wherein the application server is running an application session for the user equipment; and
transfer the application session to the second application server.

17. The apparatus according to claim 16, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

subscribe to a network entity for mobility management events, wherein the network entity is configured to participate in the handover signaling of the user equipment.

18. The apparatus according to claim 16, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

send the network entity for mobility management a request to subscribe for mobility management events including an identity for the user equipment.

19. The apparatus according to claim 16, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

receive from a user plane gateway an identifier of the network entity for mobility management.

20. The apparatus according to claim 16, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

receive an address for the target application server.

21. The apparatus according to claim 16, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

trigger relocation of the application session to the second application server.

22. The apparatus according to claim 16, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

determine by the first application server that the user equipment is moving to the radio network being served by the second application server.

23. An apparatus comprising:

at least one memory comprising computer program code;

at least one processor;

wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

receive a handover event notification request from a first application server in a first cloud, including an identifier for a user equipment, wherein the first application server is running an application session for the user equipment; and

send the first application server a handover notification that the user equipment is moving to a radio network being served by a second application server in a second cloud.

24. The apparatus according to claim 23, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

receive a request for the handover of the application session from the first application server to the second application server.

25. The apparatus according to claim 23, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

trigger the handover of the application session from the first application server in the first cloud to the second application server in the second cloud.

26. The apparatus according to claim 23, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

maintain a service area of the first cloud and a service area of the second cloud; and

maintain a mapping of the service area of the first cloud and a mapping of the service area in the second cloud.

27. The apparatus according to claim 23, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

instruct a software defined networking mechanism to setup a routing path between a user plane gateway and the first application server.

28. The apparatus according to claim 23, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

select a new user plane gateway or the second application server based on information received from a network entity.

29. The apparatus according to claim 23, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

trigger relocation of the application session to the second application server.

30. The apparatus according to claim 29, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus at least to:

sending the session setup request to a new user plane gateway and the second application server.

31. A non-transitory computer-readable medium encoding instructions that, when executed in hardware, perform a process according to any of claims 1-15.

32. An apparatus comprising means for performing a process according to any of claims 1-15.

33. A computer program product encoding instructions for performing a process according to any of claims 1-15.

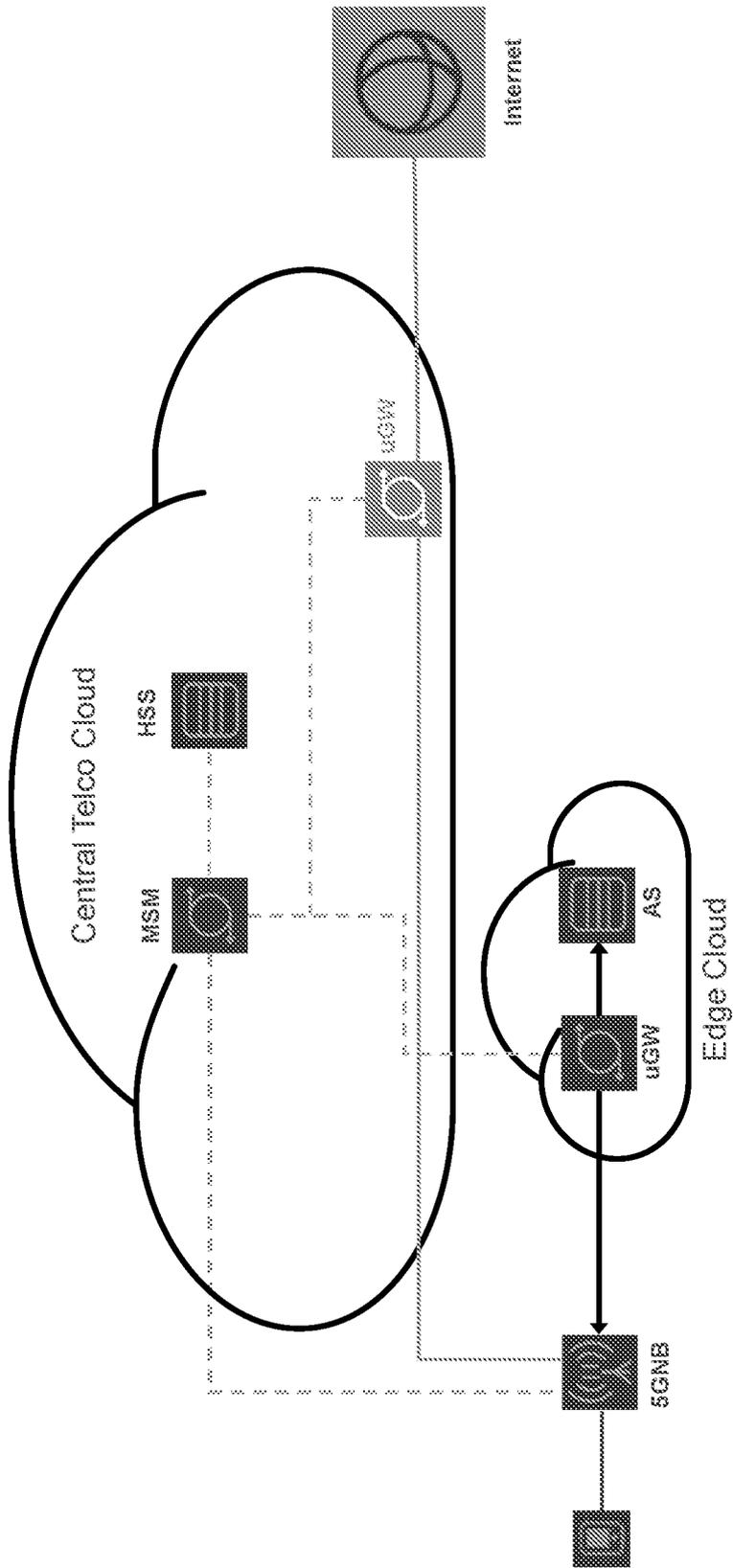


Figure 1

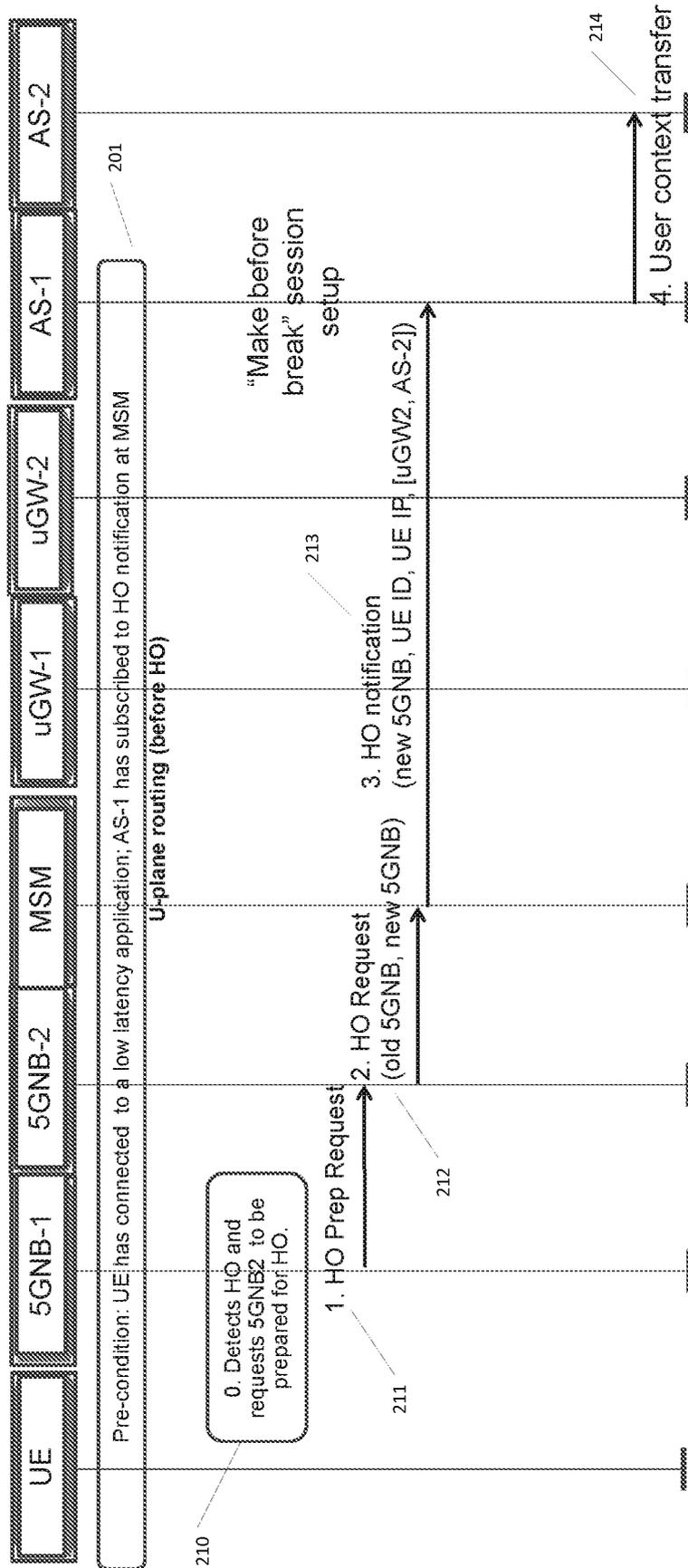


Figure 2

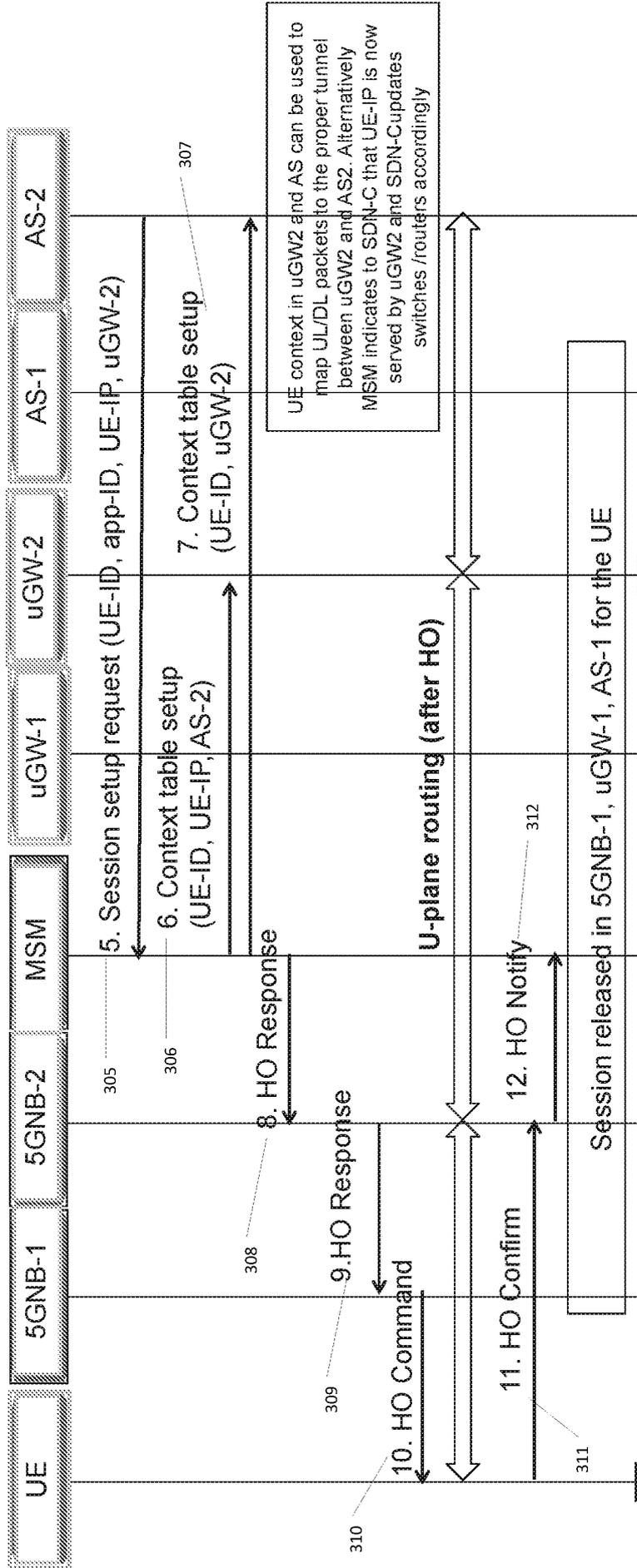


Figure 3

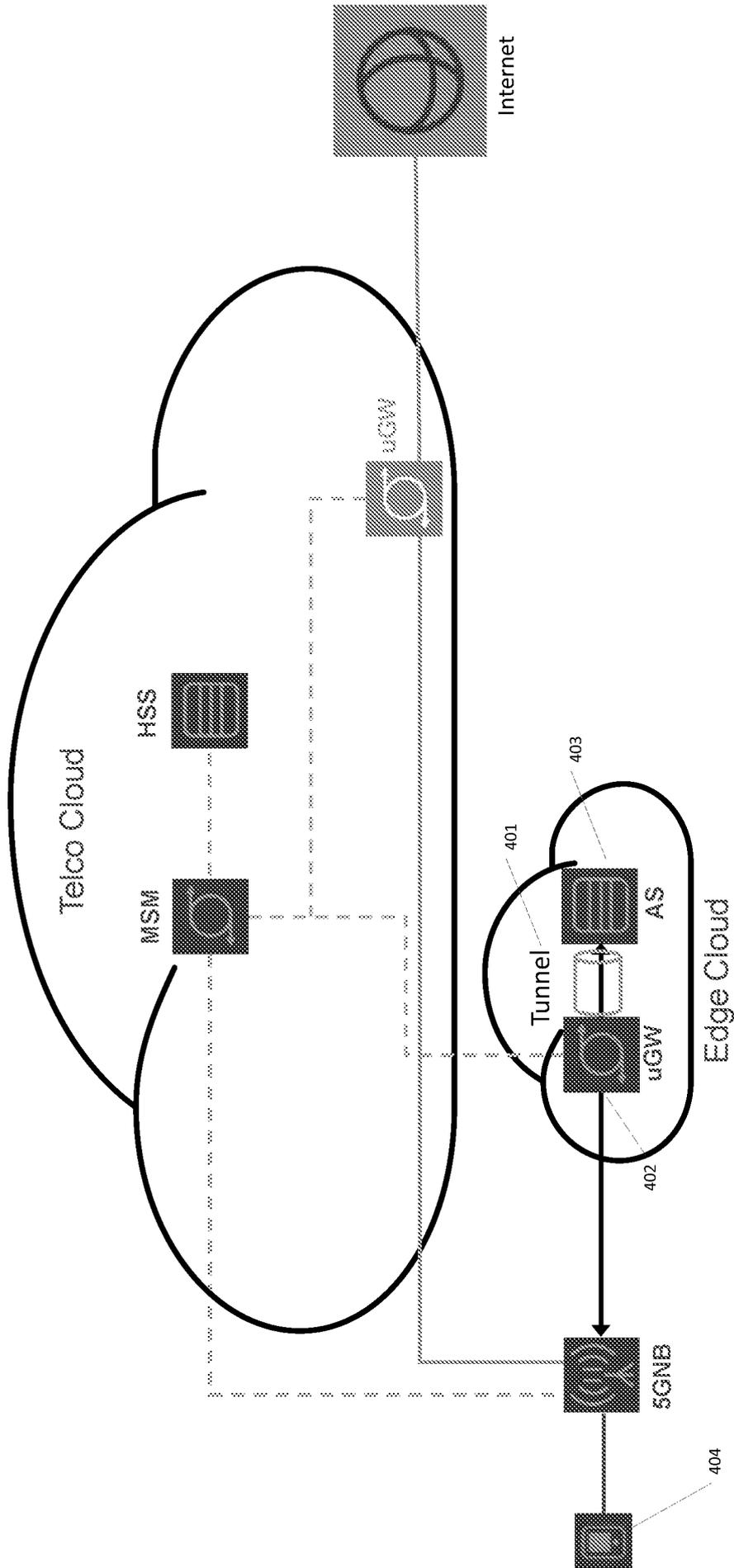


Figure 4

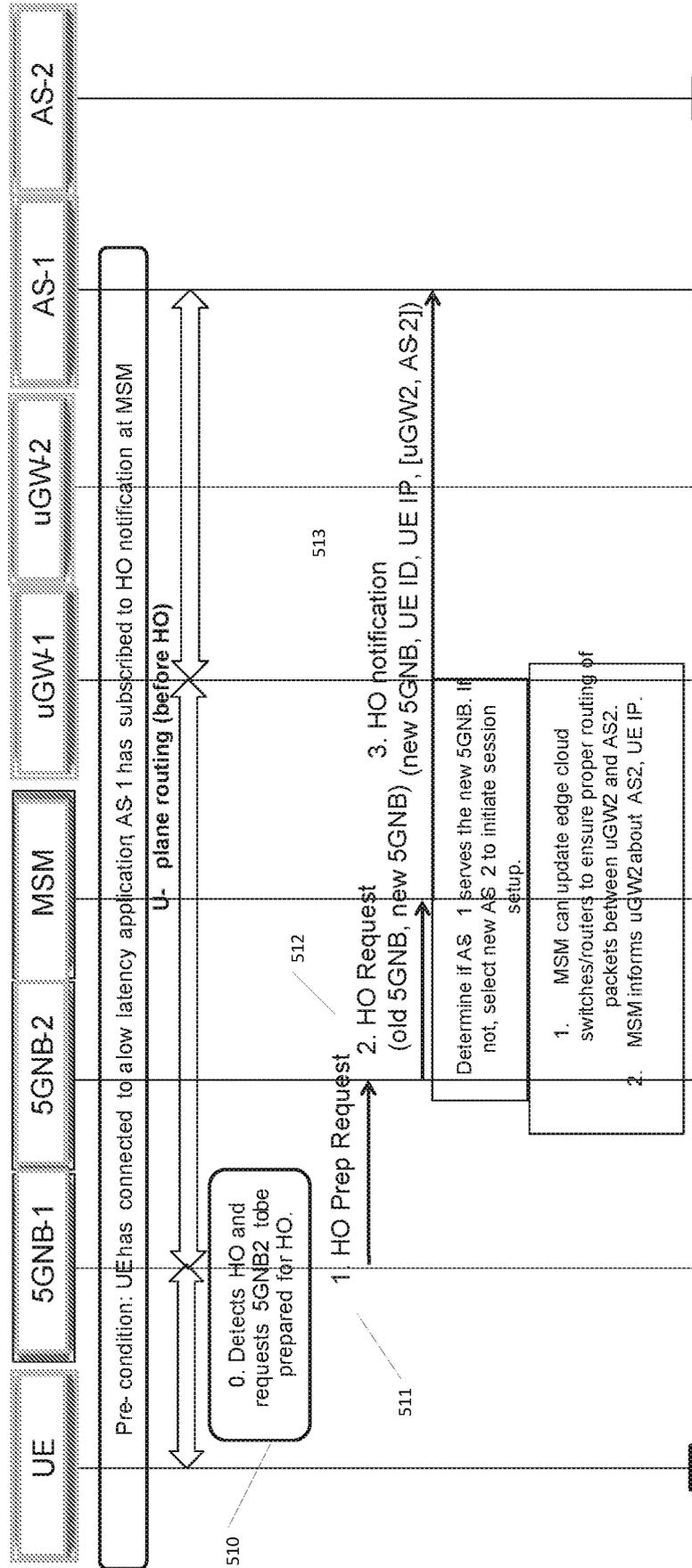


Figure 5

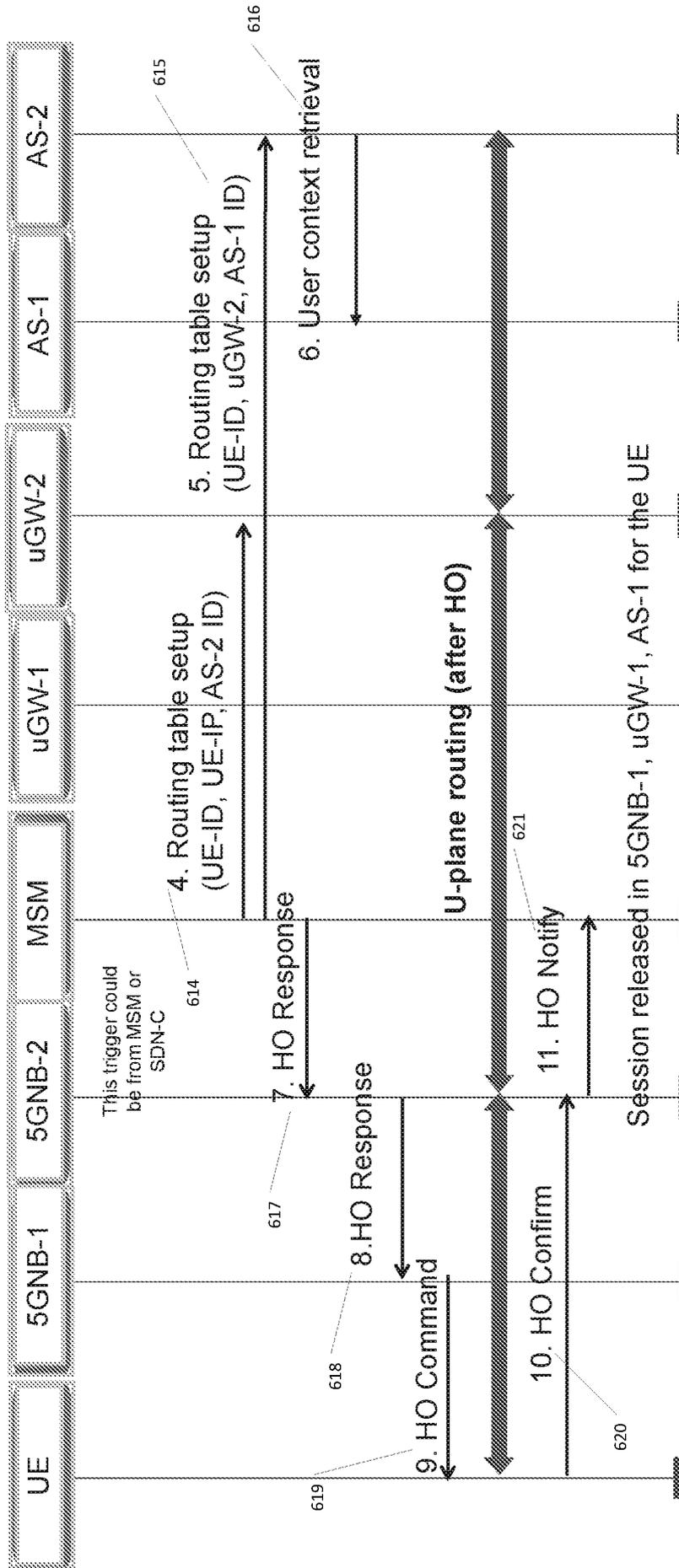


Figure 6

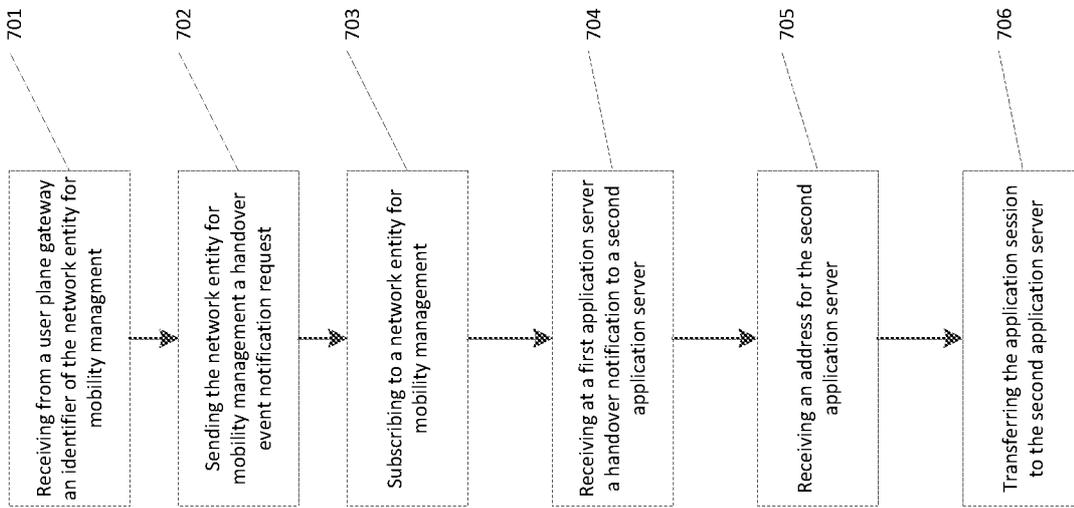


Figure 7

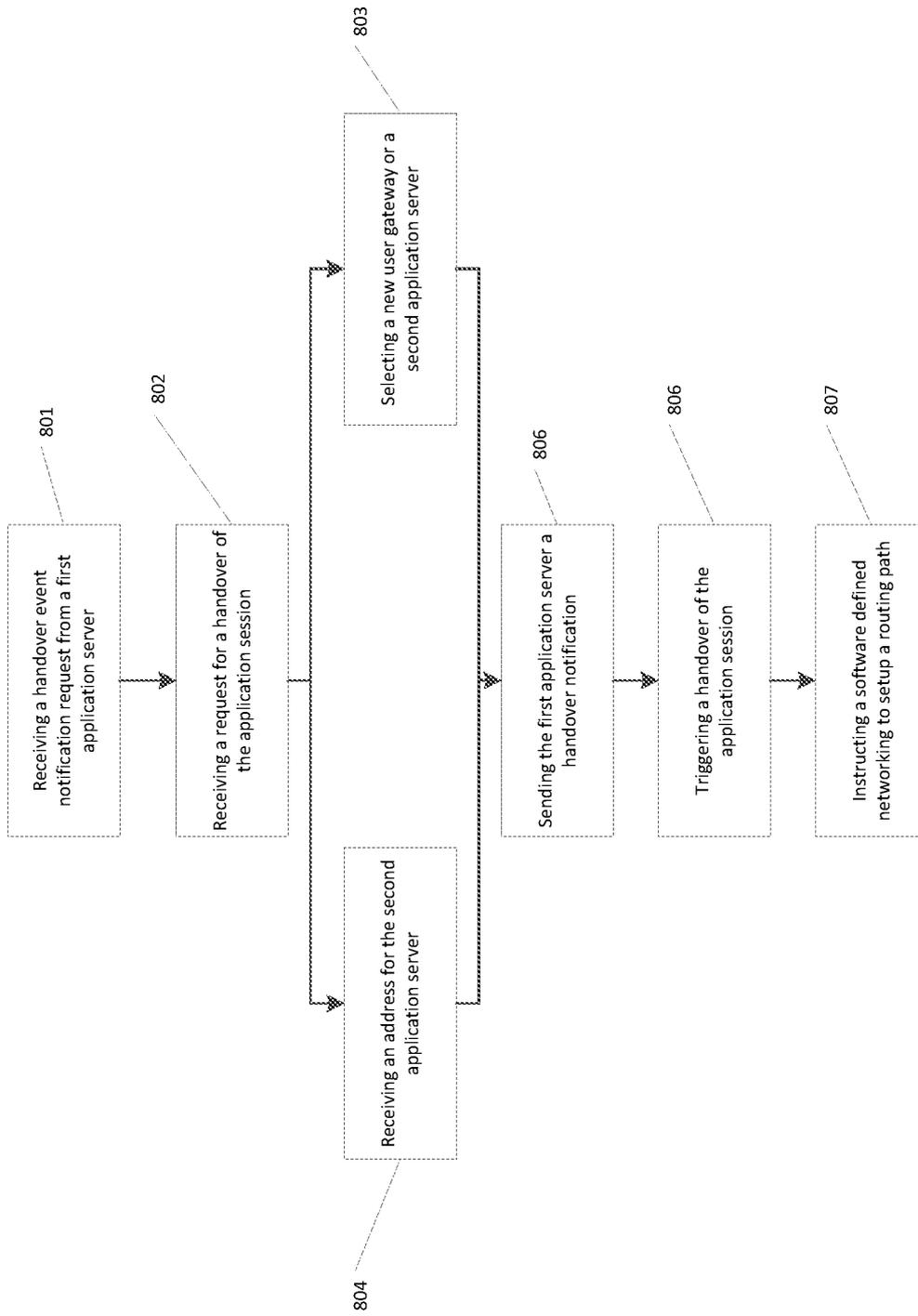


Figure 8

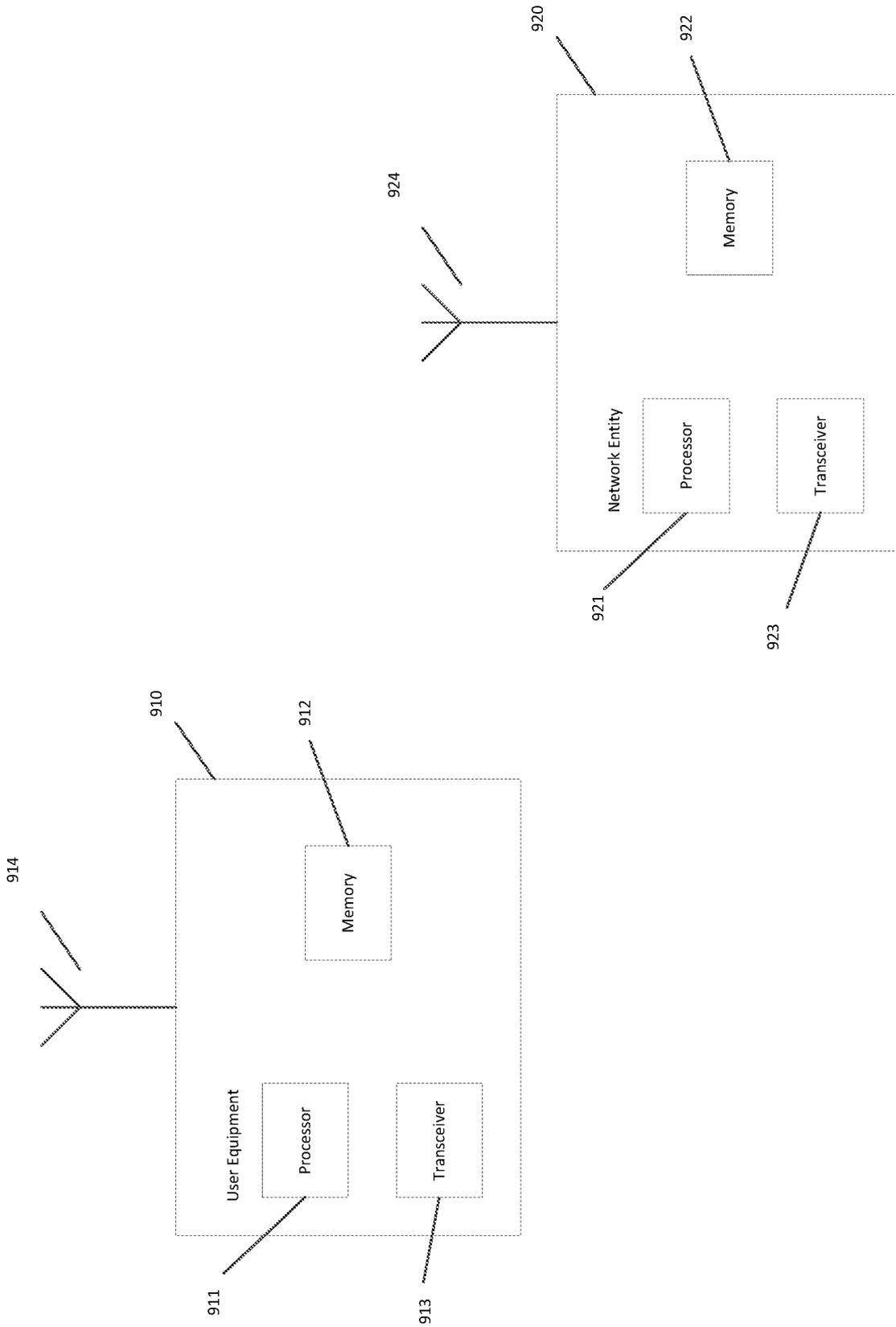


Figure 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 16/14442

A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - G06F 15/173 (2016.01) CPC - G06F 15/173; G06F 15/17306; H04L 61/106; H04L 61/103 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(8) - G06F 15/173 (2016.01) CPC - G06F 15/173; G06F 15/17306; H04L 61/106; H04L 61/103		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched CPC - G06F 15/173; G06F 15/17306; H04L 61/106; H04L 61/103; USPC - 709/226,245, 249		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Electronic Database Searched: Patbase, Google Patents, Google Search Search Terms Used: application server, cloud, handover, notification, ue, radio network, second, transferring, network entity, mobility management, handover, signaling, subscribe, gateway, radio network, mapping, routing path, etc.,		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	US 2013/0138814 A1 (Kotecha) 30 May 2013 (30.05.2013); entire document, especially, Abstract, FIG. 1, 3,5, para [0016], [0017], [0021], [0024], [0026]-[0029], [0033], [0043], [0044], [0033],[0054],	1-10, 12-14, 16-25, 27-29, 31/1-10, 31/12-14, 32/1-10, 32/12-14, 33/1-10, 33/12-14 ----- 11,15, 26,30, 31/1 1, 31/ 15, 32/1 1, 32/15, 33/1 1, 33/15
Y	US 2012/0064908 A1 (Fox et al.) 15 March 2012 (15.03.2012); especially, Abstract, FIG. 14, para [0022], [0030], [0073], [0074], [0329]	11, 26, 31/1 1, 32/1 1, 33/1 1
Y	US 201 1/0170517 A1 (Bakker et al.) 14 July 201 1 (14.07.201 1); entire document, especially, Abstract, FIG. 2, para [0089], [01 12]-[01 14], [0137])	15, 30, 31/15 32/15, 33/15
A	US 2012/0016977 A1 (Robertson et al.) 19 January 2012 (19.01 .2012); entire document,	1-33
A	US 2006/0075073 A1 (Bichot) 27 February 2003 (27.02.2003); entire document,	1-33
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/>		
* 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 23 March 2016 (23.03.2016)		Date of mailing of the international search report 19 APR 2016
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300		Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774