



- (51) International Patent Classification:
H04L 12/717 (2013.01) H04L 12/911 (2013.01)
H04W 76/02 (2009.01) H04L 12/721 (2013.01)
- (21) International Application Number:
PCT/IB2015/052447
- (22) International Filing Date:
2 April 2015 (02.04.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
14/270,182 5 May 2014 (05.05.2014) US
- (71) Applicant: TELEFONAKTIEBOLAGET L M ERICSSON (PUBL) [SE/SE]; 164 83 Stockholm (SE).
- (72) Inventors: KEMPF, James; 369 Foxborough Drive, Mountain View, California 94041 (US). BEHESHTI-ZAVAREH, Neda; 3500 Palmilla Dr., Unit 1026, San Jose, California 95134 (US). ZHANG, Ying; 5474 Middyay Common, Fremont, California 94555 (US). NILSSON, Tord; Myrvägen 20, SE-430 94, SE, SE 430-94 Bohus-Björkö (SE). JOHANSSON, Bengt; Norra dragspels-gatan 4, 9 tr, SE 421-43 Västra Frölunda (SE). PETERSSON, Sten Rune; Nolviks Kile 61, SE 423-63 Torslanda (SE). LUNING, Harald; Kyrkäsgatan 12, SE 416-56 Göteborg (SE).

- (74) Agents: CAMERON, Michael G. et al.; 6300 Legacy Drive, MS EVR 1-C-11, Plano, Texas 75024 (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))

(54) Title: IMPLEMENTING A 3G PACKET CORE IN A CLOUD COMPUTER WITH OPENFLOW DATA AND CONTROL PLANES

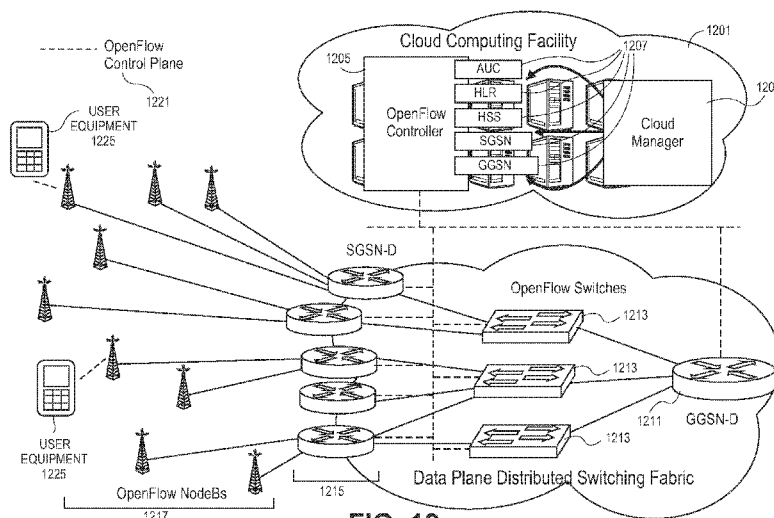


FIG. 12

(57) Abstract: A control plane device in a cloud computing system executes a plurality of virtual machines for implementing network function virtualization (NFV). The control plane device is operable to manage implementation of a general packet radio service (GPRS) tunnel protocol (GTP) in a packet core (PC) of a third generation (3G) network having a split architecture where a control plane of the PC of the 3G network is in the cloud computing system. The control plane communicates with a data plane of the PC through a control plane protocol. The data plane is implemented in a plurality of network devices of the 3G network. The control plane device and the plurality of virtual machines are operable to communicate with other control plane devices in the cloud computing system and with the plurality of network devices of the data plane.

WO 2015/170204 A1

**IMPLEMENTING A 3G PACKET CORE IN A CLOUD COMPUTER WITH
OPENFLOW DATA AND CONTROL PLANES**

CROSS-REFERENCE TO RELATED APPLICATION

5 The present patent application is a Continuation-In-Part of Application No. 13/220,471, filed August 29, 2011.

FIELD OF THE INVENTION

 The embodiments of the invention relate to a method and system for implementing a control plane of a third generation packet core in a cloud computer
10 system. Specifically, the embodiments of the invention relate to the use of the OpenFlow protocol to implement control of a data plane by the control plane being executed in a cloud computer system.

BACKGROUND

 The general packet radios system (GPRS) is a system that is used for
15 transmitting Internet Protocol packets between user devices such as cellular phones and the Internet. The GPRS system includes the GPRS core network, which is an integrated part of the global system for mobile communication (GSM). These systems are widely utilized by cellular phone network providers to enable cellular phone services over large areas.

20 The GPRS tunneling protocol (GTP) is an important communication protocol utilized within the GPRS core network. GTP enables end user devices (e.g., cellular phones) in a GSM network to move from place to place while continuing to connect to the Internet. The end user devices are connected to the Internet through a gateway GPRS support node (GGSN). The GGSN tracks the end user device's data from the end

user device's serving GPRS support node (SGSN) that is handling the session originating from the end user device.

Three forms of GTP are used by the GPRS core network. GTP-U is used for transfer of user data in separated tunnels for each packet data protocol (PDP) context.

5 GTP-C is used within the GPRS core network for signaling between GGSN and SGSN.

GTP' is used for carrying charging data from Charging Data Function (CDF) of the GSM or UMTS network to the charging gateway function (CGF), which provides requisite end user device usage information to a billing system. GTP' uses the same message structure as GTP-C and GTP-U.

10 SUMMARY

A control plane device in a cloud computing system executes a plurality of virtual machines for implementing network function virtualization (NFV). The control plane device is operable to manage implementation of a general packet radio service (GPRS) tunnel protocol (GTP) in a packet core (PC) of a third generation (3G) network
15 having a split architecture where a control plane of the PC of the 3G network is in the cloud computing system. The control plane communicates with a data plane of the PC through a control plane protocol. The data plane is implemented in a plurality of network devices of the 3G network. The control plane device and the plurality of virtual machines are operable to communicate with other control plane devices in the
20 cloud computing system and with the plurality of network devices of the data plane.

The control plane device includes a storage medium to store a centralized control plane software including control plane modules for implementing the control plane of the PC and a processor communicatively coupled to the storage medium. The processor is operable to execute the plurality of virtual machines, where at least one of

the plurality of virtual machines is operable to execute the centralized control plane (CCP) software including at least one of the control plane modules. Each control plane module provides a set of control plane functions for managing the data plane. The CCP software is operable to receive a request to create a GTP tunnel in the PC of the 3G network between a serving GPRS support node (SGSN) and a gateway GPRS support node (GGSN) for a subscriber session. The CCP software is operable to configure a switch implementing a data plane of the SGSN, via the control plane protocol, to encapsulate and decapsulate packets of the subscriber session and to establish a first GTP tunnel endpoint. The CCP software is operable to configure at least one switch in a route of the GTP tunnel, via the control plane protocol, to forward packets of the subscriber session according to the GTP tunnel, and the CCP software is operable to configure a switch implementing a data plane of the GGSN, via the control plane protocol, to encapsulate and decapsulate the packets of the subscriber session and to establish a second GTP tunnel endpoint.

15 BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings in which like references indicate similar elements. It should be noted that different references to "an" or "one" embodiment in this disclosure are not necessarily to the same embodiment, and such references mean at least one. Further, when a particular feature, structure, or characteristic is described in connection with an embodiment, it is submitted that it is within the knowledge of one skilled in the art to effect such feature, structure, or characteristic in connection with other embodiments whether or not explicitly described.

Figure 1 is a diagram of one embodiment of an example network with an OpenFlow switch.

Figure 2 is a diagram illustrating one embodiment of the contents of a flow table entry.

5 **Figure 3** illustrates another example architecture implementing OpenFlow.

Figure 4 illustrates one embodiment of the processing of packets through an OpenFlow switch packet processing pipeline.

Figure 5 is a flowchart of one embodiment of the OpenFlow rule matching process.

10 **Figure 6** is a diagram of the fields, which a matching process can utilize for identifying rules to apply to a packet.

Figures 7A and 7B are a flowchart of one embodiment of a process for OpenFlow header processing.

15 **Figure 8** is a diagram of one embodiment of a third generation (3G) packet core (PC).

Figure 9 is a diagram of one embodiment of the header fields in the primary GTP-U encapsulation header.

Figure 10 is a diagram of one embodiment of a cloud computing system servicing a set of clients.

20 **Figure 11** is a diagram of another embodiment of a cloud computing system showing the process of adding a new service instance to a client's VPC.

Figure 12 is a diagram of one embodiment of the 3G PC implemented in the cloud computing system.

Figure 13 is a flowchart of one embodiment of a process for the basic operation of the 3G PC network.

Figure 14 is a diagram of one embodiment of how the 3G PC in the cloud computing system enables a managed services company to manage multiple operator
5 networks out of a single data center.

Figure 15 is a diagram of one embodiment of a process for 3G PC peering and differential routing for specialized service treatment.

Figure 16 is a diagram of one embodiment of the OpenFlow flow table modification for GTP TEID routing.

10 **Figure 17** is a diagram of the structure of a flow table row.

Figures 18A-C are flowcharts of session creation, modification and deletion in the 3G PC.

Figure 19 is a diagram of one embodiment of an OpenFlow message flow for the create session request procedure.

15 **Figure 20** is a diagram of one embodiment of the OpenFlow message sequence for the modify session request procedure.

Figure 21 is a diagram of one embodiment of the OpenFlow message sequence for the delete session request procedure

20 **Figure 22A** illustrates connectivity between network devices (NDs) within an exemplary network, as well as three exemplary implementations of the NDs, according to some embodiments of the invention.

Figure 22B illustrates an exemplary way to implement the special-purpose network device 602 according to some embodiments of the invention.

Figure 22C illustrates various exemplary ways in which virtual network elements (VNEs) may be coupled according to some embodiments of the invention.

Figure 22D illustrates a network with a single network element (NE) on each of the NDs of Figure 22A, and within this straight forward approach contrasts a traditional distributed approach (commonly used by traditional routers) with a centralized approach for maintaining reachability and forwarding information (also called network control), according to some embodiments of the invention.

Figure 22E illustrates the simple case of where each of the NDs 2200A-H implements a single NE 2270A-H (see Figure 22D), but the centralized control plane 2276 has abstracted multiple of the NEs in different NDs (the NEs 2270A-C and G-H) into (to represent) a single NE 2270I in one of the virtual network(s) 2292 of Figure 22D, according to some embodiments of the invention.

Figure 22F illustrates a case where multiple VNEs (VNE 2270A.1 and VNE 2270H.1) are implemented on different NDs (ND 2200A and ND 2200H) and are coupled to each other, and where the centralized control plane 2276 has abstracted these multiple VNEs such that they appear as a single VNE 670T within one of the virtual networks 2292 of Figure 22D, according to some embodiments of the invention.

Figure 23 illustrates a general purpose control plane device 2304 including hardware 2340 comprising a set of one or more processor(s) 2342 (which are often Commercial off-the-shelf (COTS) processors) and network interface controller(s) 2344 (NICs; also known as network interface cards) (which include physical NIs 2346), as well as non-transitory machine readable storage media 748 having stored therein centralized control plane (CCP) software 2350), according to some embodiments of the invention.

DESCRIPTION OF EMBODIMENTS

The following description describes methods and apparatus for computation of
*. In the following description, numerous specific details such as logic
implementations, opcodes, means to specify operands, resource
5 partitioning/sharing/duplication implementations, types and interrelationships of system
components, and logic partitioning/integration choices are set forth in order to provide
a more thorough understanding of the present invention. It will be appreciated,
however, by one skilled in the art that the invention may be practiced without such
specific details. In other instances, control structures, gate level circuits and full
10 software instruction sequences have not been shown in detail in order not to obscure the
invention. Those of ordinary skill in the art, with the included descriptions, will be able
to implement appropriate functionality without undue experimentation.

References in the specification to “one embodiment,” “an embodiment,” “an
example embodiment,” etc., indicate that the embodiment described may include a
15 particular feature, structure, or characteristic, but every embodiment may not
necessarily include the particular feature, structure, or characteristic. Moreover, such
phrases are not necessarily referring to the same embodiment. Further, when a
particular feature, structure, or characteristic is described in connection with an
embodiment, it is submitted that it is within the knowledge of one skilled in the art to
20 affect such feature, structure, or characteristic in connection with other embodiments
whether or not explicitly described.

Bracketed text and blocks with dashed borders (e.g., large dashes, small dashes,
dot-dash, and dots) may be used herein to illustrate optional operations that add
additional features to embodiments of the invention. However, such notation should not

be taken to mean that these are the only options or optional operations, and/or that blocks with solid borders are not optional in certain embodiments of the invention.

In the following description and claims, the terms “coupled” and “connected,” along with their derivatives, may be used. It should be understood that these terms are not intended as synonyms for each other. “Coupled” is used to indicate that two or
5 more elements, which may or may not be in direct physical or electrical contact with each other, co-operate or interact with each other. “Connected” is used to indicate the establishment of communication between two or more elements that are coupled with each other.

10 An electronic device stores and transmits (internally and/or with other electronic devices over a network) code (which is composed of software instructions and which is sometimes referred to as computer program code or a computer program) and/or data using machine-readable media (also called computer-readable media), such as machine-
readable storage media (e.g., magnetic disks, optical disks, read only memory (ROM),
15 flash memory devices, phase change memory) and machine-readable transmission media (also called a carrier) (e.g., electrical, optical, radio, acoustical or other form of propagated signals – such as carrier waves, infrared signals). Thus, an electronic device (e.g., a computer) includes hardware and software, such as a set of one or more
processors coupled to one or more machine-readable storage media to store code for
20 execution on the set of processors and/or to store data. For instance, an electronic device may include non-volatile memory containing the code since the non-volatile memory can persist code/data even when the electronic device is turned off (when power is removed), and while the electronic device is turned on that part of the code that is to be executed by the processor(s) of that electronic device is typically copied

from the slower non-volatile memory into volatile memory (e.g., dynamic random access memory (DRAM), static random access memory (SRAM)) of that electronic device. Typical electronic devices also include a set or one or more physical network interface(s) to establish network connections (to transmit and/or receive code and/or data using propagating signals) with other electronic devices. One or more parts of an embodiment of the invention may be implemented using different combinations of software, firmware, and/or hardware.

The operations in the flow diagrams will be described with reference to the exemplary embodiments of the other figures. However, it should be understood that the operations of the flow diagrams can be performed by embodiments of the invention other than those discussed with reference to the other figures, and the embodiments of the invention discussed with reference to these other figures can perform operations different than those discussed with reference to the flow diagrams.

The embodiments of the present invention provide a method and system for avoiding the disadvantages of the prior art. The disadvantages of the prior art are that prior implementations of the 3G packet core use a pool of servers that are dedicated to a specific network entity, such as a server pool that is dedicated to hosting a SGSN. When additional signaling demands require that extra capacity, then a new SGSN instance is instantiated in the server pool. However, when demand is high for the services of the home subscriber server (HSS), the server pool dedicated to the HSS servers will be heavily utilized, but the server pool for the SGSN is underutilized. These underutilized server pools continue to require maintenance and incur operating expenses, but are not providing optimum performance for the network operator.

In some situations, managed services companies build and run mobile operator networks, while the mobile operator itself handles marketing, billing, and customer relations. The signaling and data traffic for each mobile operator network is kept private and isolated from the traffic of their competitors, even though their network and their competitors' networks may be managed by the same managed services company. The managed services company must maintain a completely separate server pool and physical signaling network for each mobile operator it supports. As a result, there is a large duplication of resources and an underutilization of server capacity. This increases operating expenses for the managed services companies and the mobile operator network due to the additional equipment, power and cooling requirements.

The 3G packet core architecture as it is currently defined allows only one point of presence (PoP) between the mobile operator's fixed core/Internet and the mobile aggregation network, that is, there is a single gateway GPRS support node (GGSN). Mobile network operators cannot set up multiple peering points and PoPs between neighboring operators within the aggregation network. This would reduce the amount of traffic that flows into the mobile operator's core network, thereby reducing the need for expensive and time consuming core network upgrades. In addition, peering points are usually without cost to operators as long as the service level agreements (SLAs) are observed. However, this flexibility of deployment is unavailable to mobile operators due to the need to anchor their PoP with the core/Internet at a single mobile gateway.

The 3G PC architecture also contains little flexibility for specialized treatment of user flows. Though the architecture does provide support for establishing quality of service (QoS), other sorts of data management are not available. For example, services involving middleboxes, such as specialized deep packet inspection or interaction with

local data caching and processing resources that might be utilized for transcoding or augmented reality applications is difficult to support with the current 3G PC architecture. Almost all such applications require the packet flows to exit through the GGSN, thereby being de-tunneled from GTP, and to be processed within the wired
5 network.

Implementing the control plane of an 3G PC in a cloud computing facility and the data plane of the 3G PC using a set of OpenFlow switches, as well as managing communication between the control plane and the data plane using the OpenFlow protocol (e.g., OpenFlow 1.1), creates a problem that the OpenFlow protocol does not
10 support GTP or GTP tunnel endpoint identifier (TEID) routing, which is necessary for implementing the data plane of the 3G PC.

The embodiments of the invention overcome these disadvantages of the prior art. The disadvantages of the prior art are avoided by splitting the control plane and the data plane for the 3G PC architecture and to implement the control plane by deploying
15 the 3G PC control plane entities in a cloud computing facility, while the data plane is implemented by a distributed collection of OpenFlow switches. The OpenFlow protocol is used to connect the two, with enhancements to support GTP routing. While the 3G PC architecture already has a split between the control plane and the data plane, in the sense that the SSGN and the GGSN are primarily data plane entities while the
20 home location register (HLR), home subscriber server (HSS) and authentication center (AUC) are primarily control plane entities, this split was made at the level of the mobility management protocol, GTP.

The standard 3G PC architecture assumes a standard routed IP network for transport on top of which the mobile network entities and protocols are implemented.

The enhanced 3G PC architecture described herein is instead at the level of IP routing and media access control (MAC) switching. Instead of using L2 routing and L3 internal gateway protocols to distribute IP routing and managing Ethernet and IP routing as a collection of distributed control entities, L2 and L3 routing management is centralized in a cloud facility and the routing is controlled from the cloud facility using the OpenFlow protocol. As used herein, the “OpenFlow protocol” refers to the OpenFlow network protocol and switching specification defined in the OpenFlow Switch Specification at www.openflowswitch.org a web site hosted by Stanford University. As used herein, an “OpenFlow switch” refers to a network device implementing the OpenFlow protocol.

The standard 3G PC control plane entities HSS, HLR, AUC, visitor location register (VLR), equipment identity register (EIR), short message service interworking message center (SMS-IWMSC), SMS gateway message center (SMS-GMSC) and subscriber location function (SLF) and control plane aspects of the SGSN and GGSN are deployed in the cloud. The data plane consists of standard OpenFlow switches with enhancements as needed for routing GTP packets, rather than IP routers and Ethernet switches. At a minimum, the data plane parts of the SGSN and GGSN and the packet routing part of the NodeB in the E-UTRAN require OpenFlow enhancements for GTP routing. Additional enhancements for GTP routing may be needed on other switches within the 3G PC architecture depending on how much fine grained control over the routing an operator requires. The enhancements for GTP routing include processes for establishing GTP tunnels, modifying GTP tunnels and tearing down GTP tunnels within the 3G PC architecture.

The split between the control and data plane parts of the 3G PC can be used together with virtual private cloud (VPC) technology to implement multiple PoPs within a single 3G PC, provide GTP flow-specific routing for specialized applications, and run multiple operator networks from a single cloud computing facility.

5 In one embodiment, the cloud based 3G PC system can be implemented as a set of hardware devices. In another embodiment, the system components are implemented in software (for example microcode, assembly language or higher level languages). These software implementations can be stored on a non-transitory computer-readable medium. A non-transitory “computer-readable” medium can
10 include any medium that can store information. Examples of the non-transitory computer-readable medium include a read only memory (ROM), a floppy diskette, a CD Rom, a DVD, a flash memory, a hard drive, an optical disc or similar medium.

OpenFlow 1.0 Networks

Figure 1 is a diagram of one embodiment of an example network with an
15 OpenFlow switch, conforming to the OpenFlow 1.0 specification. The OpenFlow 1.0 protocol enables a controller 101 to connect to an OpenFlow 1.0 enabled switch 109 using a secure channel 103 and control a single forwarding table 107 in the switch 109. The controller 101 is an external software component executed by a remote computing device that enables a user to configure the OpenFlow 1.0 switch 109. The secure
20 channel 103 can be provided by any type of network including a local area network (LAN) or a wide area network (WAN), such as the Internet.

Figure 2 is a diagram illustrating one embodiment of the contents of a flow table entry. The forwarding table 107 is populated with entries consisting of a rule 201 defining matches for fields in packet headers; an action 203 associated to the flow

match; and a collection of statistics 205 on the flow. When an incoming packet is received a lookup for a matching rule is made in the flow table 107. If the incoming packet matches a particular rule, the associated action defined in that flow table entry is performed on the packet.

5 A rule 201 contains key fields from several headers in the protocol stack, for example source and destination Ethernet MAC addresses, source and destination IP addresses, IP protocol type number, incoming and outgoing TCP or UDP port numbers. To define a flow, all the available matching fields may be used. But it is also possible to restrict the matching rule to a subset of the available fields by using wildcards for the
10 unwanted fields.

The actions that are defined by the specification of OpenFlow 1.0 are Drop, which drops the matching packets; Forward, which forwards the packet to one or all outgoing ports, the incoming physical port itself, the controller via the secure channel, or the local networking stack (if it exists). OpenFlow 1.0 protocol data units (PDUs) are
15 defined with a set of structures specified using the C programming language. Some of the more commonly used messages are: report switch configuration message; modify state messages (including a modify flow entry message and port modification message); read state messages, where while the system is running, the datapath may be queried about its current state using this message; and send packet message, which is
20 used when the controller wishes to send a packet out through the datapath.

OpenFlow 1.0 supports “vendor extensions” that allow certain protocol elements to be extended. Protocol messages and table actions can be extended, but flow-matching rules cannot. The use of these extensions in connection with the cloud based EPC architecture is discussed further herein below.

OpenFlow 1.1 Networks

Figure 3 illustrates another example architecture implementing OpenFlow, conforming to the OpenFlow 1.1 specification. In this embodiment, there is an explicit provision for multiple flow tables 301. This allows the packet processing pipeline to
5 mix and match particular rules and actions without causing a combinatorial explosion in the table size. For example, one flow table can perform QoS processing while a second flow table does routing.

Figure 4 illustrates one embodiment of the processing of packets through an OpenFlow 1.1 switched packet processing pipeline. A received packet is compared
10 against each of the flow tables 401. After each flow table match, the actions are accumulated into an action set. If processing requires matching against another flow table, the actions in the matched rule include an action directing processing to the next table in the pipeline. Absent the inclusion of an action in the set to execute all
accumulated actions immediately, the actions are executed at the end 403 of the packet
15 processing pipeline. An action allows the writing of data to a metadata register, which is carried along in the packet processing pipeline like the packet header.

Figure 5 is a flowchart of one embodiment of the OpenFlow 1.1 rule matching process. OpenFlow 1.1 contains support for packet tagging. OpenFlow 1.1 allows matching based on header fields and multi-protocol label switching (MPLS) labels.
20 One virtual LAN (VLAN) label and one MPLS label can be matched per table. The rule matching process is initiated with the arrival of a packet to be processed (Block 501). Starting at the first table 0 a lookup is performed to determine a match with the received packet (Block 503). If there is no match in this table, then one of a set of default actions is taken (i.e., send packet to controller, drop the packet or continue to

next table) (Block 509). If there is a match, then an update to the action set is made along with counters, packet or match set fields and meta data (Block 505). A check is made to determine the next table to process, which can be the next table sequentially or one specified by an action of a matching rule (Block 507). Once all of the tables have
5 been processed, then the resulting action set is executed (Block 511).

Figure 6 is a diagram of the fields, which a matching process can utilize for identifying rules to apply to a packet. Actions allow manipulating of tag stacks by pushing and popping labels. Combined with multiple tables, VLAN or MPLS label stacks can be processed by matching one label per table.

10 **Figure 7** is a flow chart of one embodiment of a header parsing process. The parsing process matches a packet header by initializing a set of match fields (Block 701) and checking for the presence of a set of different header types. The process checks for a VLAN tag (Block 703). If the VLAN tag is present, then there are a series of processing steps for the VLAN tag (Blocks 705-707). If the switch supports MPLS
15 (Block 709), then there are a series of steps for detecting and processing the MPLS header information (Blocks 711-715). If the switch supports address resolution protocol (ARP), then there are a series of steps for processing the ARP header (Blocks 719 and 721). If the packet has an IP header (Block 723), then there are a series of steps for processing the IP header (Blocks 725-733). This process is performed for
20 each received packet.

In one embodiment, a group table can be supported in conjunction with the OpenFlow 1.1 protocol. Group tables enable a method for allowing a single flow match to trigger forwarding on multiple ports. Group table entries consist of four fields: a group identifier, which is a 32 bit unsigned integer identifying the group; a group type

that determines the group's semantics; counters that maintain statistics on the group; and an action bucket list, which is an ordered list of action buckets, where each bucket contains a set of actions to execute together with their parameters.

There are four different types of groups: All, which execute all actions in the bucket list, this is used for broadcast or multicast forwarding; Select, which execute one bucket per packet, based on an algorithm determined by the switch which is outside the OpenFlow protocol, this is used to implement multipath forwarding; Indirect, which execute the single bucket on all packets, this allows multiple flows or groups to point to a single collection of actions rather than having the actions defined in multiple forwarding table entries; Fast Failover, which execute the first live bucket, where each bucket is associated with a port that controls its liveness, this enables the switch to failover to another port without involving the controller.

OpenFlow 1.1 can be utilized to support virtual ports. A virtual port, as used herein, is an "action block" that performs some kind of processing action other than simply forwarding the packet out to a network connection like physical ports do. Examples of a few built-in virtual ports include: ALL, which forwards the port out all ports except for the ingress port and any ports that are marked "Do Not Forward;" CONTROLLER, which encapsulates the packet and sends it to the controller; TABLE, which inserts the packet into the packet processing pipeline by submitting it to the first flow table, this action is only valid in the action set of a packet-out message; and IN_PORT, which sends the packet out the ingress port. In other embodiments, there can also be switched-defined virtual ports.

3G PC Architecture

Figure 8 is a diagram of one embodiment of a 3G packet core (3G PC) network. A 3G PC network consist of three interacting domains: a core network (CN) 801, a radio access network (RAN) 803 and User Equipment (UE) 805. The user
5 equipment 805 can be computing devices, cellular phones and similar devices. The radio access networks 803 can be any number or combination of networks including universal mobile telecommunications system (UMTS) 841 or global systems for mobile communications (GSM) networks 843. These networks can interface with the core
10 network 801 through a radio network center (RNC) 831 or a packet control unit (PCU) 833.

The main function of the core network 801 is to provide switching, routing and transit for user traffic from the user equipment 805. The core network 801 also contains databases and network management functions. It is the common packet core network for GSM/GPRS, wideband code division multiple access (WCDMA)/ high speed
15 packet access (HSPA) and non-3GPP mobile networks. The core network 801 is used for transmitting Internet Protocol (IP) packets. The core network 801 interfaces with the Internet 851 and other networks 853 through the GGSN 819.

The core network 801 is divided into circuit switched and packet switched domains. The circuit switched elements include a mobile services switching center
20 (MSC) 811, visitor location register (VLR) 813 and Gateway MSC 815. Packet switched elements are SGSNs 817 and GGSN 819. Other network devices, like EIR 821, HLR 823, VLR 813 and AUC 825 are shared by both domains.

The architecture of the core network 801 can change when new services and features are introduced. In other embodiments, a number portability database (NPDB)

can be used to enable a user to change networks while keeping an old phone number. A gateway location register (GLR) can be used to optimize the subscriber handling between network boundaries.

The primary functions of the core network 801 with respect to mobile wireless networking are mobility management and QoS. These functions are not typically provided in a fixed broadband network, but they are crucial for wireless networks. Mobility management is necessary to ensure packet network connectivity when a wireless terminal moves from one base station to another. QoS is necessary because, unlike fixed networks, the wireless link is severely constrained in how much bandwidth it can provide to the terminal, so the bandwidth needs to be managed more tightly than in fixed networks in order to provide the user with acceptable quality of service.

The signaling for implementing the mobility management and QoS functions is provided by the GPRS Tunneling Protocol (GTP). GTP has two components: GTP-C – a control plane protocol that supports establishment of tunnels for mobility management and bearers for QoS management that matches wired backhaul and packet core QoS to radio link QoS; and GTP-U – a data plane protocol used for implementing tunnels between network devices that act as routers. There are two versions of GTP-C protocol, i.e. GTP version 1 (GTPv1-C and GTPv1-U) and GTP version 2-C (designed for LTE). GTPv1 is primarily utilized in conjunction with the 3G PC based system.

Network Services are considered to be end-to-end, this means from a Terminal Equipment (TE) to another TE. An End-to-End Service may have a certain Quality of Service (QoS) which is provided for the user of a network service. It is the user that decides whether he is satisfied with the provided QoS or not. To realize a certain

network QoS Service with clearly defined characteristics and functionality is to be set up from the source to the destination of a service.

In addition to the QoS parameters, each bearer has an associated GTP tunnel. A GTP tunnel consists of the IP address of the tunnel endpoint nodes (radio base station, SGSN, and GGSN), a source and destination UDP port, and a Tunnel Endpoint Identifier (TEID). GTP tunnels are unidirectional, so each bearer is associated with two TEIDs, one for the uplink and one for the downlink tunnel. One set of GTP tunnels (uplink and downlink) extends between the radio base station and the SGSN and one set extends between the SGSN and the GGSN. The UDP destination port number for GTP-U is 2152 while the destination port number for GTP-C is 2123. The source port number is dynamically allocated by the sending node. **Figure 9** is a diagram of one embodiment of the header fields in the primary GTP-U encapsulation header.

Cloud Computing

Data centers offer computing, storage, and network communication resources to outside customers. The offered services can consist of elastic, on demand processing, storage that for most practical purposes is limited only by the customer's ability to pay, and network bandwidth into the Internet. This set of services provided by a data center is referred to herein as cloud computing.

Server virtualization technology allows a pool of servers to be managed as essentially one large compute resource. A layer of software called a hypervisor sits between the operating system and the hardware. The hypervisor schedules the execution of virtual machines (VMs). A VM is an operating system image packaged with some applications. The hypervisor allows a VM to be suspended and moved between servers to load balance. Load balancing and monitoring of VM execution to

catch crashes provides the same kind of fault tolerance and scalability services for enterprise applications that are achieved at much higher cost with specialized solutions. A cloud manager system oversees the execution of VMs, the scheduling of execution to meet the demand of the VMs and the optimization of server utilization and
5 minimization of power consumption. The cloud manager or cloud operating system is a software program that can schedule execution to allow an in-service upgrade of hardware and software without impacting ongoing service provisioning to the VMs and their applications in the cloud computing system.

To support the arbitrary movement of VMs between machines, the networking
10 within the data center must also be virtualized. Cloud computing systems can virtualize the network by incorporating a virtual switch into the hypervisor. The virtual switch provides virtual network ports to the VMs executing under the control of the hypervisor. The virtual switch software also allows the network resources to be virtualized in a manner similar to how the server resources are virtualized by the
15 hypervisor. The hypervisor and the virtual switch can thereby co-operate to allow VMs to be moved between servers. When the hypervisor moves a VM, it communicates with the virtual switch about the new location, and the virtual switch ensures that the network routing tables for the VM's addresses (L2 MAC address, potentially also the IP address) are updated so packets are routed to the new location.

20 A cloud computing system can be composed of any number of computing devices having any range of capabilities (e.g., processing power or storage capacity). The cloud computing system can be a private or public system. The computing devices can be in communication with one another across any communication system or network. A cloud computing system can support a single cloud or service or any

number of discrete clouds or services. Services, applications and similar programs can be virtualized or executed as standard code. In one embodiment, cloud computing systems can support web services applications. Web services applications consist of a load balancing front end that dispatches requests to a pool of Web servers. The requests
5 originate from applications on remote machines on the Internet and therefore the security and privacy requirements are much looser than for applications in a private corporate network.

Cloud computer systems can also support secure multi-tenancy, in which the cloud computer system provider offers virtual private network (VPN)-like connections
10 between the clients' distributed office networks outside the cloud and a VPN within the cloud computing system. This allows the clients' applications within the cloud computing system to operate in a network environment that resembles a corporate WAN. For private data centers, in which services are only offered to customers within the corporation owning the data center, the security and privacy requirements for multi-
15 tenancy are relaxed. But for public data centers, the cloud operator must ensure that the traffic from multiple tenants is isolated and there is no possibility for traffic from one client to reach another client network.

Figure 10 is a diagram of one embodiment of a cloud computing system servicing a set of clients. A 'set,' as used herein, refers to any positive whole number of
20 items. In the embodiment, shown in **Figure 10**, two virtual private clouds (VPCs) are set up for two different external enterprise customers. A VPC consists of a collection of VMs, storage, and networking resources that provide secure multi-tenancy to the enterprises renting space in the cloud. The enterprise customers connect into the VPCs via VPNs over the Internet running on a public operator network.

Figure 11 is a diagram of another embodiment of a cloud computing system showing the process of adding a new service instance to a client's VPC. In this case, the in-cloud VPN is implemented using MAC layer Virtual LANs (VLANs). The VM is created on a hypervisor managed server within the VPC for the enterprise requesting the new service instance (step 1). The virtual switch VLAN is configured to include the new VM into the enterprise in-cloud VPN, thereby establishing service connectivity within the cloud (step 2). The virtual customer edge router (CE) is updated for the new service (step 3). The provider edge router in the operator network where the enterprise VPN is run is updated with the new service (step 4).

10 **3G PC Implementation in Cloud Computing System**

Figure 12 is a diagram of one embodiment of the EPC implemented in the cloud computing system. The 3G PC control plane entities (AUC, HLR, HSS) and the control plane parts of the support nodes (SGSN, GGSN), i.e. the parts that handle GTP signaling, are implemented in the cloud computing system as part of the OpenFlow controller. The control plane entities and the OpenFlow controller are packaged as VMs. The application programming interface (API) between the OpenFlow controller and the control plane entities can be a remote procedure call (RPC) interface or similar interface. This implementation technology is favorable for scalable management of the control plane entities within the cloud, since it allows execution of the control plane entities and the controller to be managed separately according to demand. The cloud manager can be a VM or an application executed within the cloud computing system.

The cloud manager monitors the central processor unit (CPU) utilization of the 3G PC control plane entities and the control plane traffic between the 3G

PC control plane entities 1207 within the cloud. It also monitors the control plane traffic between the end user devices (UEs) 1225 and NodeBs 1217, which do not have control plane entities in the cloud computing system 1201, and the 3G PC control plane entities 1207. If the 3G PC control plane entities 1207 begin to exhibit signs of

5 overloading, such as the utilization of too much CPU time, or the queuing up of too much traffic to be processed, the overloaded control plane entity 1207 requests that the cloud manager 1203 start up a new VM to handle the load. Additionally, the 3G PC control plane entities 1207 themselves can issue event notifications to the cloud manager 1203 if they detect internally that they are beginning to experience

10 overloading.

The cloud manager 1203 also provides reliability and failover by restarting a VM for a particular control plane entity 1207 or function if any of the 3G PC control plane entities 1207 should crash. During this restart process the cloud manager can collect diagnostic data, save any core files of the failed 3G PC control plane entity, and

15 inform the system administrators that a failure occurred. The control plane entities 1207 maintains the same protocol interface between themselves as in the 3GPP 3G PC architecture shown in **Figure 8**.

The OpenFlow control plane 1221, shown here as a dotted line, manages the routing and switching configuration in the network. The OpenFlow control plane 1221

20 connects the cloud computing system 1203 to the SGSN-Ds 1215 (i.e., the data plane of the SGSN), the standard OpenFlow switches 1213, and the GGSN-D 1211 (i.e., the data plane of the GGSN). The physical implementation of the OpenFlow control plane 1221 can be as a completely separate physical network, or it may be a virtual network running over the same physical network as the data plane, implemented with a

prioritized VLAN or with an MPLS label switched path or even with a generic routing encapsulation (GRE) or other IP tunnel. The OpenFlow control plane 1221 can in principle use the same physical control plane paths as the GTP-C and other mobile network signaling. The SGSN-Ds 1215 and the GGSN-Ds 1211 act as OpenFlow GTP-
5 extended gateways, encapsulating and decapsulating packets using the OpenFlow GTP switch extensions described further herein below.

The NodeBs 1217 have no control plane entities in the cloud because the radio access network (RAN) signaling required between the 3G PC and the NodeB includes radio parameters, and not just IP routing parameters. Therefore, there is no OpenFlow
10 control plane 1221 connection between the OpenFlow controller 1205 in the cloud computing system 1201 and the NodeBs 1217. The NodeBs 1217 can however act as OpenFlow GTP-extended gateways by implementing a local control to data plane connection using OpenFlow. This allows the packet switching side of the NodeBs 1217 to utilize the same OpenFlow GTP switching extensions as the packet gateways.

15 The operation of the 3G PC cloud computer system works as follows. The UE 1225, NodeB 1217, SGSN 1207 and GGSN signal 1207 to the HLR, HSS, AUC, SMS-GMSC 1207 using the standard 3G PC protocols, to establish, modify, and delete GTP tunnels. This signaling triggers procedure calls with the OpenFlow controller to modify the routing in the 3G PC as requested. The OpenFlow controller configures the
20 standard OpenFlow switches, the Openflow SGSN 1215, and GGSN 1211 with flow rules and actions to enable the routing requested by the control plane entities. Details of this configuration are described in further detail herein below.

Figure 13 is a flowchart of one embodiment of the basic operation of the 3G PC. In one embodiment, the process begins with the initialization of the control plane

modules of the 3G PC within the OpenFlow controller in the cloud computing system (Block 13401). Each control plane module in the plurality of control plane modules is initialized as a separate VM by the cloud manager. The cloud manager then monitors the resource utilization of each control plane module as well as the amount and type of control plane traffic handled by each control plane module (Block 1303). The cloud manager can directly monitor this data, receive reports from the control plane modules or any combination thereof.

If the cloud manager detects a threshold level of resource utilization or traffic load for any one of the plurality of control plane modules being monitored (Block 1305), the cloud manager can take steps to automatically respond to this scenario. The cloud manager can initialize a new control plane module or an instance of such a control plane module as a separate virtual machine (Block 1307). This new control plane module or instance can then share the load of existing control plane modules or instances of the same type thereby alleviating the load on these modules dynamically.

Similarly, the cloud manager may detect the failure or the underutilization of one of the plurality of control plane modules (Block 1309). The cloud manager can then restart a failed control plane module or terminate an underutilized control plane module (Block 1311). Restarting the control plane module ensures a level of load sharing for a pool of control plane modules. Deactivating a control plane module frees up the resources and reduces the overhead created by the control plane module. The cloud manager can perform these functions across VPCs and mobile operators using the cloud computing system resources, thereby maximizing the use of available resources and reducing the cost of operation while maintaining strict separation of data and traffic between mobile operators.

Figure 14 is a diagram of one embodiment of how the 3G PC in the cloud computing system enables a managed services company to manage multiple operator networks out of a single data center. The managed services cloud computing facility 1401 runs separate instances of the 3G PC control plane for every mobile operator with which the managed services company has a contract. Each 3G PC instance is in a VPC 1403A,B that isolates the mobile operator's traffic from other tenants in the cloud computing facility 1401 of the data center. The 3G PC control plane instance for a mobile operator is connected to the mobile operator's geographically distributed 3G PC OpenFlow data plane switching fabric 1407A,B and the mobile operator's base stations through a virtual edge router 1409A,B. The virtual edge router 1409A,B routes traffic from the data center to and from the appropriate mobile operator 3G PC data plane switching fabric 1407A,B. In some cases, the mobile operators may even share base stations and 3G PC switching fabrics, though the example embodiment in **Figure 14** shows a case where the two mobile operators have separate switching fabrics.

15 3G PC Peering and Differential Routing in a Cloud Computing System

Figure 15 is a diagram of one embodiment of a process for 3G PC peering and differential routing for specialized service treatment. The OpenFlow signaling, indicated by the solid lines and arrows 1501, sets up flow rules and actions on the switches and gateways within the 3G PC for differential routing. These flow rules direct GTP flows to particular locations. In this example, the operator in this case peers its 3G PC with two other fixed operators. Routing through each peering point is handled by the respective GGSN1 and GGSN2 1503A, B. The dashed lines and arrows 1505 show traffic from a UE 1507 that needs to be routed to another peering operator. The flow rules and actions to distinguish which peering point the traffic should traverse

are installed in the OpenFlow switches 1509 and gateways 1503A, B by the OpenFlow controller 1511. The OpenFlow controller 1511 calculates these flow rules and actions based on the routing tables it maintains for outside traffic, and the source and destination of the packets, as well as by any specialized forwarding treatment required for DSCP marked packets.

The long dash and dotted lines and arrows 1515 shows an example of a UE 1517 that is obtaining content from an external source. The content is originally not formulated for the UE's 1517 screen, so the OpenFlow controller 1511 has installed flow rules and actions on the GGSN1 1503B, SGSN 1519 and the OpenFlow switches 1509 to route the flow through a transcoding application 1521 in the cloud computing facility. The transcoding application 1521 reformats the content so that it will fit on the UE's 1517 screen. A MSC requests the specialized treatment at the time the UE sets up its session with the external content source via the IP Multimedia Subsystem (IMS) or another signaling protocol.

15 GTP TEID Routing

In one embodiment, OpenFlow is modified to provide rules for GTP TEID Routing. **Figure 16** is a diagram of one embodiment of the OpenFlow flow table modification for GTP TEID routing. An OpenFlow switch that supports TEID routing matches on the 2 byte (16 bit) collection of header fields and the 4 byte (32 bit) GTP TEID, in addition to other OpenFlow header fields, in at least one flow table (e.g., the first flow table). The GTP TEID flag can be wildcarded (i.e. matches are "don't care"). In one embodiment, the 3G PC protocols do not assign any meaning to TEIDs other than as an endpoint identifier for tunnels, like ports in standard UDP/TCP transport protocols. In other embodiments, the TEIDs can have a correlated meaning or

semantics. The GTP header flags field can also be wildcarded, this can be partially matched by combining the following bitmasks: 0xFF00 – Match the Message Type field; 0xe0 – Match the Version field; 0x10 – Match the PT field; 0x04 – Match the E field; 0x02 – Match the S field; and 0x01 – Match the PN field.

5 In one embodiment, OpenFlow can be modified to support virtual ports for fast path GTP TEID encapsulation and decapsulation. An OpenFlow mobile gateway can be used to support GTP encapsulation and decapsulation with virtual ports. The GTP encapsulation and decapsulation virtual ports can be used for fast encapsulation and decapsulation of user data packets within GTP-U tunnels, and can be designed simply
10 enough that they can be implemented in hardware or firmware. For this reason, GTP virtual ports may have the following restrictions on traffic they will handle: Protocol Type (PT) field = 1, where GTP encapsulation ports only support GTP, not GTP' (PT field = 0); Extension Header flag (E) = 0, where no extension headers are supported, Sequence Number flag (S) = 0, where no sequence numbers are supported; N-PDU flag
15 (PN) = 0; and Message type = 255, where Only G-PDU messages, i.e. tunneled user data, is supported in the fast path.

 If a packet either needs encapsulation or arrives encapsulated with nonzero header flags, header extensions, and/or the GTP-U packet is not a G-PDU packet (i.e. it is a GTP-U control packet), the processing must proceed via the gateway's slow path
20 (software) control plane. GTP-C and GTP' packets directed to the gateway's IP address are a result of mis-configuration and are in error. They must be sent to the OpenFlow controller, since these packets are handled by the SGSN and GGSN control plane entities in the cloud computing system or to the billing entity handling GTP' and not the SGSN and GGSN data plane switches.

GTP virtual ports are configured from the OpenFlow controller using a configuration protocol. The details of the configuration protocol are switch-dependent. The configuration protocol must support messages that perform the following functions: allow the controller to query for and return an indication whether the switch supports GTP fast path virtual ports and what virtual port numbers are used for fast path and slow path GTP-U processing; and allow the controller to instantiate a GTP-U fast path virtual port within a switch datapath for use in the OpenFlow table set-output-port action. The configuration command must be run in a transaction so that, when the results of the action are reported back to the controller, either a GTP-U fast path virtual port for the requested datapath has been instantiated or an error has returned indicating why the request could not be honored. The command also allows the OpenFlow controller to bind a GTP-U virtual port to a physical port. For decapsulation virtual ports, the physical port is an input port. For encapsulation virtual ports, the physical port is an output port.

The OpenFlow controller instantiates a virtual port for each physical port that may transmit or receive packets routed through a GTP tunnel, prior to installing any rules in the switch for GTP TEID routing.

In one embodiment, an OpenFlow GTP gateway maintains a hash table mapping GTP TEIDs into the tunnel header fields for their bearers. **Figure 17** is a diagram of the structure of a flow table row. The TEID hash keys are calculated using a suitable hash algorithm with low collision frequency, for example SHA-1. The gateway maintains one such flow table row for each GTP TEID/bearer. The TEID field contains the GTP TEID for the tunnel. The VLAN tags and MPLS labels fields contain an ordered list of VLAN tags and/or MPLS labels defining tunnels into which the

packet needs to be routed. The VLAN priority bits and MPLS traffic class bits are included in the labels. Such tunnels may or may not be required. If they are not required, then these fields are empty. The tunnel origin source IP address contains the address on the encapsulating gateway to which any control traffic involving the tunnel should be directed (for example, error indications). The tunnel end destination IP address field contains the IP address of the gateway to which the tunneled packet should be routed, at which the packet will be decapsulated and removed from the GTP tunnel. The QoS DSCP field contains the DiffServe Code Point, if any, for the bearer in the case of a dedicated bearer. This field may be empty if the bearer is a default bearer with best effort QoS, but will contain nonzero values if the bearer QoS is more than best effort.

In one embodiment, slow path support for GTP is implemented with an OpenFlow gateway switch. An OpenFlow mobile gateway switch also contains support on the software control plane for slow path packet processing. This path is taken by G-PDU (message type 255) packets with nonzero header fields or extension headers, and user data plane packets requiring encapsulation with such fields or addition of extension headers, and by GTP-U control packets. For this purpose, the switch supports three local ports in the software control plane: *LOCAL_GTP_CONTROL* - the switch fast path forwards GTP encapsulated packets directed to the gateway IP address that contain GTP-U control messages and the local switch software control plane initiates local control plane actions depending on the GTP-U control message; *LOCAL_GTP_U_DECAP* - the switch fast path forwards G-PDU packets to this port that have nonzero header fields or extension headers (i.e. E != 0, S != 0, or PN != 0). These packets require specialized handling. The local switch software slow path

processes the packets and performs the specialized handling; and

LOCAL_GTP_U_ENCAP - the switch fast path forwards user data plane packets to this port that require encapsulation in a GTP tunnel with nonzero header fields or extension headers (i.e. E != 0, S != 0, or PN != 0). These packets require specialized handling.

- 5 The local switch software slow path encapsulates the packets and performs the specialized handling. In addition to forwarding the packet, the switch fast path makes the OpenFlow metadata field available to the slow path software.

To support slow path encapsulation, the software control plane on the switch maintains a hash table with keys calculated from the GTP-U TEID. The TEID hash
10 keys are calculated using a suitable hash algorithm with low collision frequency, for example SHA-1. The flow table entries contain a record of how the packet header, including the GTP encapsulation header, should be configured. This includes: the same header fields as for the hardware or firmware encapsulation table in Figure 18; values for the GTP header flags (PT, E, S, and PN); the sequence number and/or the N-PDU
15 number if any; if the E flag is 1, then the flow table contains a list of the extension headers, including their types, which the slow path should insert into the GTP header.

In one embodiment, the system implements a GTP fast path encapsulation virtual port. When requested by the SGSN and GGSN control plane software running in the cloud computing system, the OpenFlow controller programs the gateway switch
20 to install rules, actions, and TEID hash table entries for routing packets into GTP tunnels via a fast path GTP encapsulation virtual port. The rules match the packet filter for the input side of GTP tunnel's bearer. Typically this will be a 4 tuple of: IP source address; IP destination address; UDP/TCP/SCTP source port; and UDP/TCP/SCTP destination port. The IP source address and destination address are typically the

addresses for user data plane traffic, i.e. a UE or Internet service with which a UE is transacting, and similarly with the port numbers. For a rule matching the GTP-U tunnel input side, the associated instructions and are the following:

Write-Metadata (*GTP-TEID*, 0xFFFFFFFF)

5 Apply-Actions (Set-Output-Port *GTP-Encap-VP*)

The switch also writes an entry in the TEID hash table containing the tunnel header fields for the packet. GTP-TEID is the GTP tunnel endpoint identifier. *GTP-Encap-VP* is the GTP fast path encapsulation virtual port bound to the physical port
10 out which the encapsulated packet will ultimately be routed.

When a packet header matches a rule associated with the virtual port, the GTP TEID is written into the lower 32 bits of the metadata and the packet is directed to the virtual port. The virtual port calculates the hash of the TEID and looks up the tunnel header information in the tunnel header table. If no such tunnel information is present,
15 the packet is forwarded to the controller with an error indication. Otherwise, the virtual port constructs a GTP tunnel header and encapsulates the packet. Any DSCP bits or VLAN priority bits are additionally set in the IP or MAC tunnel headers, and any
10 VLAN tags or MPLS labels are pushed onto the packet. The encapsulated packet is forwarded out the physical port to which the virtual port is bound.

20 In one embodiment, the system implements a GTP fast path decapsulation virtual port. When requested by the SGSN and GGSN control plane software running in the cloud computing system, the gateway switch installs rules and actions for routing GTP encapsulated packets out of GTP tunnels. The rules match the GTP header flags and the GTP TEID for the packet, in the modified OpenFlow flow table shown in

Figure 16 as follows: the IP destination address is an IP address on which the gateway is expecting GTP traffic; the IP protocol type is UDP (17); the UDP destination port is the GTP-U destination port (2152); and the header fields and message type field is wildcarded with the flag 0XFFF0 and the upper two bytes of the field match the G-
5 PDU message type (255) while the lower two bytes match 0x30, i.e. the packet is a GTP packet not a GTP' packet and the version number is 1. The virtual port simply removes the GTP tunnel header and forwards the enclosed user data plane packet out the bound physical port.

In one embodiment, the system implements handling of GTP-U control packets.
10 The OpenFlow controller programs the gateway switch flow tables with 5 rules for each gateway switch IP address used for GTP traffic. These rules contain specified values for the following fields: the IP destination address is an IP address on which the gateway is expecting GTP traffic; the IP protocol type is UDP (17); the UDP destination port is the GTP-U destination port (2152); the GTP header flags and
15 message type field is wildcarded with 0xFFFF0; the value of the header flags field is 0x30, i.e. the version number is 1 and the PT field is 1; and the value of the message type field is one of 1 (Echo Request), 2 (Echo Response), 26 (Error Indication), 31 (Support for Extension Headers Notification), or 254 (End Marker).

The instruction associated with a match to one of these rules is:

20 Apply-Actions (Set-Output-Port *LOCAL_GTP_CONTROL*)

This causes the packet to be forwarded to the gateway switch's local GTP-U control port for processing by the local software control plane. GTP-U control packets

that are originated by the switch are generated on the software control plane and are routed by the control plane.

In one embodiment, the system implements handling of G-PDU packets with extension headers, sequence numbers, and N-PDU numbers. G-PDU packets with
5 extension headers, sequence numbers, and N-PDU numbers need to be forwarded to the local switch software control plane for processing. The OpenFlow controller programs
3 rules for this purpose. They have the following common header fields: the IP destination address is an IP address on which the gateway is expecting GTP traffic; and the IP protocol type is UDP (17); the UDP destination port is the GTP-U destination
10 port (2152).

The header flags and message type fields for the three rules are wildcarded with the following bitmasks and match as follows: bitmask 0xFFF4 and the upper two bytes match the G-PDU message type (255) while the lower two bytes are 0x34, indicating that the version number is 1, the packet is a GTP packet, and there is an extension
15 header present; bitmask 0xFFF2 and the upper two bytes match the G-PDU message type (255) while the lower two bytes are 0x32, indicating that the version number is 1, the packet is a GTP packet, and there is a sequence number present; and bitmask
0xFF01 and the upper two bytes match the G-PDU message type (255) while the lower two bytes are 0x31, indicating that the version number is 1, the packet is a GTP packet,
20 and a N-PDU is present.

The instruction for these rules is the following:

Apply-Actions (Set-Output-Port *LOCAL_GTP_U_DECAP*)

This sends the packet to the software slow path GTP-U decapsulation path for special processing.

In one embodiment, the system implements handling of user data plane packets requiring GTP-U encapsulation with extension headers, sequence numbers, and N-PDU numbers. User data plane packets that require extension headers, sequence numbers, or N-PDU numbers during GTP encapsulation require special handling by the software slow path. For these packets, the OpenFlow controller programs a rule matching the 4 tuple: IP source address; IP destination address; UDP/TCP/SCTP source port; and UDP/TCP/SCTP destination port. The instructions for matching packets are:

10 Write-Metadata (*GTP-TEID*, 0xFFFFFFFF)
Apply-Actions (Set-Output-Port *LOCAL_GTP_U_ENCAP*)

This sends the packet to the software slow path GTP encapsulation port and, in addition, makes the TEID available to the slow path.

15 The OpenFlow message programming the rule insertion also includes information on the values for the sequence number, N-PDU number, or the type and contents of the extension header, as well as the packet header fields designating the decapsulation gateway and bearer transport and the GTP TEID. This information is inserted by the switch's control plane software into the software encapsulation table,
20 keyed by the TEID.

In one embodiment, the system implements handling of GTP-C and GTP' control packets. Any GTP-C and GTP' control packets that are directed to IP addresses on a gateway switch are in error. These packets need to be handled by the SGSN, GGSN, and GTP' protocol entities in the cloud computing system, not the SGSN and

GGSN entities in the switches. To catch such packets, the OpenFlow controller must program the switch with the following two rules: the IP destination address is an IP address on which the gateway is expecting GTP traffic; the IP protocol type is UDP (17); for one rule, the UDP destination port is the GTP-U destination port (2152), for
5 the other, the UDP destination port is the GTP-C destination port (2123); the GTP header flags and message type fields are wildcarded.

These rules must be the lowest priority of all the GTP rules in the gateway switch's flow table. They will match any GTP packets that don't match other, more specific rules. The instruction for these rules is the following:

10 Apply-Actions (Set-Output-Port *CONTROLLER*)

This encapsulates the packet and sends it to the OpenFlow controller.

In one embodiment, the system implements non-gateway GTP routing. A GTP-extended Openflow switch can also accomplish GTP routing without performing the
15 gateway functions of encapsulation and decapsulation. The GTP routing function can be performed by a gateway switch in addition to its gateway function, or it can be performed by another switch that lacks a gateway function, within the distributed EPC switching fabric.

A GTP-extended Openflow switch contains at least one flow table that handles
20 rules matching the GTP header fields as in **Figure 16**. The Openflow controller programs the GTP header field rules in addition to the other fields to perform GTP routing and adds appropriate actions if the rule is matched. For example, the following rule matches a GTP-C control packet directed to a control plane entity (MSC, SGSN,GGSN) in the cloud computing system, which is not in the control plane VLAN:

the VLAN tag is not set to the control plane VLAN, the destination IP address field is set to the IP address of the targeted control plane entity, the IP protocol type is UDP (17), the UDP destination port is the GTP-C destination port (2123), the GTP header flags and message type is wildcarded with 0xF0 and the matched version and protocol type fields are 2 and 1, indicating that the packet is a GTPv1 control plane packet and not GTP'.

The following actions push a control plane VLAN tag onto the packet and forward it to the cloud for processing by the relevant control plane entity. The packet is forwarded without any L3 processing (i.e. not modifying the IP TTL):

10 Write-Action (Set-VLAN-ID *CP_VLAN_TAG*)

Write-Action (Set-Source-MAC-Address *SWITCH_MAC_ADDR*)

Write-Action (Set-Dest-MAC-Address *NEXT_HOP_MAC_ADDR*)

Set-Output-Port *NEXT_HOP_PORT*

15 GTP Extensions for OpenFlow

The OpenFlow protocol can be modified to provide extensions for GTP that enable the management of the 3G PC. OpenFlow utilizes data structures referred to as flow match structures that enable the protocol to define criteria for matching rules to particular flows. The OpenFlow flow match structure of *ofp_match*, contains two fields, type and length, that allow the flow match structure to be extended. The type field can be set to the type of the extension and the length field can be set to the length of the extended *ofp_match* structure. In one embodiment, a new type based on a random number for GTP flow matching is defined:

```
enum ofp_match_type_ext {  
    ERSMT_GTP = 48696,  
};
```

The type can be randomly generated so as not to interfere with other
5 extended types. There is currently no organizational mechanism to register type
identifiers in OpenFlow.

The `ersmt_gtp` structure defines the flow table fields for GTP flow routing as:

```
struct ersmt_gtp {  
    unit8_t gtp_wildcard;  
10    uint16_t gtp_type_n_flags;  
    uint16_t gtp_flag_mask;  
    uint32_t gtp_teid;  
};
```

15 The `gtp_type_n_flags` field contains the GTP message type in the upper 8 bits
and the GTP header flags in the lower 8 bits. The `gtp_teid` field contains the GTP
TEID. The `gtp_wildcard` field indicates whether the GTP type and flags and TEID
should be matched. If the lower four bits are 1, the type and flags field should be
ignored, while if the upper four bits are 1, the TEID should be ignored. If the lower bits
20 are 0, the type and fields flag should be matched subject to the flags in the
`gtp_flag_mask` field, while if the upper bits are 0 the TEID should be matched. The
mask is combined with the message type and header field of the packet using logical
AND; the result becomes the value of the match. Only those parts of the field in which
the mask has a 1 value are matched.

In addition to the flow table fields, an object is required to encode the encapsulation of the virtual port TEID hash table entry. The `ersmt_gtp_tuninfo` structure can be used to define this information:

```
struct ermst_mpls_lbl {  
5   uint8_t mpls_lbl_low;  
   uint8_t mpls_lbl_mid;  
   uint8_t mpls_lbl_high;  
};  
  
10 struct ersmt_gtp_tuninfo {  
   uint16_t gtp_tuninfo_length;  
   uint32_t gtp_tuninfo_saddr;  
   uint32_t gtp_tuninfo_daddr;  
   uint8_t gtp_tuninfo_dscp;  
15  uint8_t gtp_tuninfo_vlan_len;  
   uint16_t gtp_tuninfo_vlan_tags[0]; /*variable length*/  
   uint8_t gtp_tuninfo_mpls_len;  
   struct mpls_lbl gtp_tuninfo_mpls_labels[0]; /*variable length*/  
};  
20
```

The `ermst_mpls_lbl` struct provides a 24 bit data structure for encoding MPLS labels. The `ersmt_gtp_tunifo` structure contains fields describing a GTP tunnel. These are inserted into the encapsulation virtual port. The structure is variable length because it may contain a variable number of VLAN tags and/or MPLS labels. The

gtp_tuninfo_length field contains the length of the structure. The gtp_tuninfo_saddr, gtp_tuninfo_daddr, and gtp_tuninfo_dscp fields contain the source address of the tunnel (the address of the interface on the switch performing the encapsulation), the destination address of the tunnel (the switch to which the tunneled packet will be routed and that will decapsulate the packet), and the DiffServ Code Point (if any) assigned to the tunnel's bearer. The bearer DSCP will be nonzero if the bearer is a dedicated bearer and it is not a best effort bearer.

The gtp_tuninfo_vlan_len and gtp_tuninfo_mpls_len contain the length of the VLAN tags field and the MPLS labels field, respectively. The gtp_tuninfo_vlan_tags[0] and gtp_tuninfo_mpls_labels[0] fields contain the actual VLAN tags and/or the MPLS labels that need to be pushed onto the packet's tunnel header. These fields will be absent (and the corresponding length fields will be zero) if no VLAN or MPLS Label Switched Paths (LSPs) are used for the tunnel.

In one embodiment, OpenFlow is modified to add extension messages for adding, deleting, or modifying a 3G PC bearer or GTP tunnel. The OpenFlow signaling for adding, modifying, or deleting a 3G PC bearer or GTP tunnel consists of one OpenFlow message, the ofp_flow_mod message, containing an ersmt_gtp GTP flow definition. The standard OpenFlow ofp_flow_mod message can be used as long as the OpenFlow protocol parser can handle extended flows. If the flow modification requires a change to the encapsulation virtual port TEID hash table, the OpenFlow controller must issue a GTP OpenFlow extension message containing the TEID hash table entry. The OpenFlow controller must issue both messages sequentially, the ofp_flow_mod message first, then the TEID hash table modification message, then the OpenFlow

controller must issue an OFPT_BARRIER_REQUEST message to force processing of both messages by the OpenFlow switch.

The OpenFlow message extension header structure ofp_experimenter_header contains an experimenter id field, called experimenter. In one embodiment, this field
 5 can be set to the Ericsson IEEE OUI, 0x01ec or similar manufacturer or provider OUI. The rest of the structure contains the GTP extension messages. These messages can be identified by the following message codes:

```
enum ermst_gtp_message_code {
    GTP_ADD_TEID_TABLE_ENTRY = 0,
10    GTP_DEL_TEID_TABLE_ENTRY = 1,
};
```

The GTP OpenFlow extension contains a message for adding and for deleting a TEID hash table entry. Entries are modified by first deleting the entry for the TEID
 15 then adding a new entry for the same TEID. The GTP OpenFlow extension message for entering a new TEID entry in the encapsulation virtual port hash table is:

```
struct ermst_teid_table_add {
    ermst_gtp_message_code teid_table_add_type;
    uint16_t teid_table_add_teid;
20    struct ermst_gtp_tuninfo teid_table_add_entry;
};
```

The teid_table_add_type field is set to GTP_ADD_TEID_TABLE_ENTRY while the teid_table_add_teid field contains the TEID and the teid_table_add_entry

contains the table entry to be added. The GTP OpenFlow extension message for deleting a TEID entry from the encapsulation virtual port hash table is:

```
struct ermst_teid_table_del {  
    ermst_gtp_message_code teid_table_del_type;  
5    uint16_t teid_table_del_teid;  
};
```

The `teid_table_del_type` field is set to `GTP_DEL_TEID_TABLE_ENTRY` while the `teid_table_del_teid` field contains the TEID for the entry to be deleted.

10 In one embodiment, the extensions to OpenFlow for GTP also encompass OpenFlow switch configuration. Prior to accepting any GTP routing update RPCs from 3G PC cloud control plane entities, the OpenFlow controller must configure GTP encapsulation and/or decapsulation virtual ports on the GTP extended OpenFlow gateway switches. The configuration is accomplished using a switch-specific
15 configuration protocol, and is described above.

In addition to virtual port configuration on the GTP extended OpenFlow gateways, QoS queue configuration may be required on any OpenFlow switch that will be forwarding better than best effort GTP bearer traffic. The OpenFlow protocol contains no messages for configuring queues, this configuration is left up to the
20 configuration protocol, as is the case with virtual ports. Prior to installing any flow routes, the OpenFlow controller must configure any queues to connect with physical and/or virtual ports in switches that will route better than best effort GTP bearers. This configuration step must be done both for GTP extended OpenFlow switches and standard OpenFlow switches.

In one embodiment OpenFlow message flows for GTP operations are modified. As described above, the 3G PC control plane entities, including the 3G PC control plane parts of the SGSN and the SSSN reside in a cloud computing facility at a data center. The control plane of the SGSN and GGSN communicate via remote procedure calls (RPCs) or similar mechanism with the OpenFlow controller within the cloud when routing changes are triggered by the GTP signaling. The OpenFlow controller enacts the changes on the data plane to GTP extended OpenFlow enabled data plane gateways, the control plane of the SGSN and GGSN, and to OpenFlow switches that are extended for GTP routing, referred to herein as 'GxOFS,' through OpenFlow signaling on the control plane network connecting the cloud to the gateways and switches.

In general, no signaling is required to the GxOFS if no special routing treatment is required for GTP flows. Cases where such treatment might be required are, for example: where a where the operator's 3G PC has peering points with the Internet at more than one point and consequently has more than one gateway, routing to the optimal gateway may require steering traffic within the 3G PC at intermediate switches; and where the GTP flow must receive special treatment from an application somewhere within the operator's network, for example, within the cloud. An example of such special treatment is transcoding. The intermediate switches may require programming to route the user plane packets to the transcoding application. This list is not exhaustive, many other applications of GTP routing on intermediate switches are possible.

GTP PDP context tunnels can be set up using the GTP-C create PDP context request messages. This procedure is used in a variety of message sequences, for example, in an E-UTRAN initial attach procedure.

Figures 18A-C are flowcharts of session creation, modification and deletion in the 3G PC. The process for creating a session is illustrated in **Figure 18A**. The process is initiated in response to a request to create a GTP tunnel between a SGSN and a GGSN for a subscriber session (Block 1801). The subscriber session is initiated to connect user equipment in the 3G PC to another endpoint of the subscriber session, which can be another user equipment (UE), a server or similar endpoint. The GTP tunnel establishes the route of the subscriber session across the core network of the 3G PC network to a peering point, the Internet or similar endpoint. The controller configures a switch implementing the SGSN to encapsulate and decapsulate data packets for the subscriber session and to establish a first GTP tunnel endpoint (Block 1803). The controller also configures each switch in the route of the GTP tunnel to forward packets in each direction according to the route of the GTP tunnel (Block 1805). The controller configures a data plane of the GGSN to encapsulate and decapsulate the packets of the subscriber session to establish a second endpoint of the GTP tunnel at the GGSN (Block 1807).

The process for modifying a session is illustrated in **Figure 18B**. The process is initiated in response to a request to modify a GTP tunnel between a SGSN and a GGSN for a subscriber session (Block 1811). The subscriber session connects user equipment in the 3G PC to another endpoint of the subscriber session, which can be another user equipment (UE), a server or similar endpoint. The GTP tunnel is an established route of the subscriber session across the core network of the 3G PC network to a peering

point, the Internet or similar endpoint. A modification process can be utilize to re-route a subscriber session across the 3G PC network to the same endpoints or to different endpoints. Any combination of the endpoints of the GTP tunnel and the path of the GTP can be changed using this process. The controller modifies the configuration of the switch implementing the SGSN to encapsulate and decapsulate data packets for the subscriber session and to modify a first GTP tunnel endpoint (Block 1813). The controller also configures each switch in the current route and the new route of the GTP tunnel to forward packets in each direction according to the route of the GTP tunnel (Block 1815). The controller modifies the configuration of a data plane of the GGSN to encapsulate and decapsulate the packets of the subscriber session to establish a second endpoint of the GTP tunnel at the GGSN (Block 1817).

The process for deleting a session is illustrated in **Figure 18C**. The process is initiated in response to a request to delete a GTP tunnel between a SGSN and a GGSN for a subscriber session (Block 1821). The subscriber session connects user equipment in the 3G PC to another endpoint of the subscriber session, which can be another user equipment (UE), a server or similar endpoint. The GTP tunnel is an established route of the subscriber session across the core network of the 3G PC network to a peering point, the Internet or similar endpoint. The subscriber session and associated GTP tunnel are deleted when the associated application on the user equipment or the application on the other user equipment or the server application on the other end of the subscriber session terminate the connection. The subscriber session resources are then reclaimed by deleting the subscriber session and the GTP tunnel. The controller removes the configuration of the GTP tunnel at a switch implementing the SGSN that had been encapsulating and decapsulating data packets for the subscriber session and

thereby removes a first GTP tunnel endpoint (Block 1823). The controller also removes configuration for the GTP tunnel from each switch in the route of the GTP tunnel that had been forwarding packets in each direction according to the route of the GTP tunnel (Block 1815). The controller removes configuration for the GTP tunnel from a data plane of the GGSN that had been encapsulating and decapsulating the packets of the subscriber session and thereby removes a second endpoint of the GTP tunnel at the GGSN (Block 1827).

In **Figure 19**, an example of the OpenFlow message flows for the create session request procedure are shown. In the illustrated example, the SSGN control plane component sends a create session request to the GGSN control plane component in the cloud computing system, which then sends the request to the controller through a GTP routing update RPC call. The RPC call requests that the OpenFlow controller establish a new GTP tunnel endpoint at the SGSN and GGSN in the data plane, and to install routes for the new GTP bearer or tunnel on intermediate switches, if necessary.

Before returning a result to the control plane GGSN from the GTP routing update RPC, the OpenFlow controller issues a sequence of OpenFlow messages to the appropriate data plane gateway entity. In the example embodiment, the sequence begins with an `OFFP_BARRIER_REQUEST` to ensure that there are no pending messages that might influence processing of the following messages. Then an `OFPT_FLOW_MOD` message is issued, including the `ofp_match` structure with GTP extension as the match field and `OFFPFC_ADD` as the command field. The message specifies actions and instructions, as described above, to establish a flow route for the GTP tunnel that encapsulates and decapsulates the packets through the appropriate virtual port. In addition, immediately following the `OFPT_FLOW_MOD` message, the OpenFlow

controller issues a GTP_ADD_TEID_TABLE_ENTRY message to the gateways containing the TEID hash table entries for the encapsulation virtual port. As described above, the two OpenFlow messages are followed by an OFPT_BARRIER_REQUEST message to force the gateways to process the flow route and TEID hash table update
5 before proceeding.

Prior to returning from the GTP routing update RPC, the OpenFlow controller also issues GTP flow routing updates to any GTP extended OpenFlow Switches (GxOFSs) that need to be involved in customized GTP flow routing. The messages in these updates consist of an OFP_BARRIER_REQUEST followed by an
10 OFPT_FLOW_MOD message containing the ofp_match structure with GTP extension for the new GTP flow as the match field and OFPFC_ADD as the command field, and the actions and instructions described above for customized GTP flow routing. A final OFP_BARRIER_REQUEST forces the switch to process the change before
15 responding. The flow routes on any GxOFSs are installed after installing the GTP tunnel endpoint route on the SGSN in the data plane and prior to installing the GTP tunnel endpoint route on the GGSN in the data plane, as illustrated in **Figure 19**. The OpenFlow controller does not respond to the control plane GGSN RPC until all flow routing updates have been accomplished.

Once the RPCs have returned, the control plane GGSN and SGSN return create
20 PDP context response messages. Characteristics of the GTP bearer are changed using an update PDP context request process. Such changes may, for example, include the QoS assigned to the IP packets. This procedure is used in a variety of 3G PC message sequences, for example, a UE triggered service request.

Figure 20 is a diagram of one embodiment of the OpenFlow message sequence for the modify session request procedure. As with session creation, the 3G PC cloud control plane SGSN issues an update PDP context request message to the control plane GGSN and the control plane GGSN issues a GTP routing update RPC to the OpenFlow controller including the new tunnel update information. The OpenFlow controller then issues GTP extended OpenFlow messages to the data plane SGSN, GxOFSeS, and the data plane GGSN.

Before returning a result to the control plane GGSN from the GTP routing update RPC, the OpenFlow controller issues a sequence of OpenFlow messages to the appropriate data plane gateway entity. The sequence begins with an OFP_BARRIER_REQUEST to ensure that there are no pending messages that might influence processing of the following messages. Then an OFPT_FLOW_MOD message is issued, including the ofp_match structure with GTP extension as the match field and OFPFC_MODIFY or OFPFC_MODIFY_STRICT as the command field. If necessary, the message specifies actions and instructions, as described above, to establish a new flow route for the GTP tunnel that encapsulates and decapsulates the packets through the appropriate virtual port. In addition, if changes are required in the TEID hash table, immediately following the OFPT_FLOW_MOD message, the OpenFlow controller issues a TP_DEL_TEID_TABLE_ENTRY to delete the entry followed by a TP_ADD_TEID_TABLE_ENTRY message to install the new entry. As described above, the two OpenFlow messages are followed by an OFPT_BARRIER_REQUEST message to force the gateways to process the flow route and TEID hash table update before proceeding.

Prior to returning from the GTP routing update RPC, the OpenFlow controller also issues necessary GTP flow routing updates to any GTP extended OpenFlow Switches (GxOFSs) that need to be involved in customized GTP flow routing. The messages in these updates consist of an OFP_BARRIER_REQUEST followed by an
5 OFPT_FLOW_MOD message containing the ofp_match structure with GTP extension for the new GTP flow as the match field and OFPFC_MODIFY or OFPFC_MODIFY_STRICT as the command field, and, if necessary, the actions and instructions, as described above, for customized GTP flow routing. A final
10 OFP_BARRIER_REQUEST forces the switch to process the change before responding. The flow routes on any GxOFSs are installed after installing the GTP tunnel endpoint route on the data plane SGSN and prior to installing the GTP tunnel endpoint route on the data plane GGSN, as illustrated in **Figure 20**. The OpenFlow controller does not respond to the control plane GGSN RPC until all flow routing updates have been accomplished. Once the RPCs have returned, the control plane
15 GGSN and SGSN return update PDP context messages.

GTP tunnels are deleted using the delete session request procedure. This procedure can be used in a variety of 3G PC message sequences, for example, a UE triggered detach request. **Figure 21** is a diagram of one embodiment of the OpenFlow message sequence for the delete session request procedure. Similar to session creation
20 and modification, the 3G PC cloud control plane SGSN issues a delete PDP context request message to the control plane GGSN and the control plane GGSN issues a GTP routing update RPC to the OpenFlow controller including the tunnel deletion information. The OpenFlow controller then issues GTP extended OpenFlow messages to the data plane SGSN, GxOFSes, and the data plane GGSN.

The OpenFlow signaling is conducted prior to returning the GTP routing update RPC to the calling party. The sequence begins with an `OFFP_BARRIER_REQUEST` to ensure that there are no pending messages that might influence processing of the following messages. Then an `OFPT_FLOW_MOD` message is issued, including the
5 `ofp_match` structure with GTP extension as the match field and `OFFPFC_DELETE` or `OFFPFC_DELETE_STRICT` as the command field. In addition, immediately following the `OFPT_FLOW_MOD` message, the OpenFlow controller issues a
`GTP_DEL_TEID_TABLE_ENTRY` to delete the TEID hash table entry. As described above, this OpenFlow message is followed by an `OFPT_BARRIER_REQUEST`
10 message to force the gateways to process the flow route and TEID hash table update before proceeding.

Prior to returning from the GTP routing update RPC, the OpenFlow controller also issues necessary GTP flow routing updates to any GTP extended OpenFlow Switches that need to be involved in customized GTP flow routing. The messages in
15 these updates consist of an `OFFP_BARRIER_REQUEST` followed by an `OFPT_FLOW_MOD` message containing the `ofp_match` structure with GTP extension for the new GTP flow as the match field and `OFFPFC_DELETE` or `OFFPFC_DELETE_STRICT` as the command field. A final
`OFFP_BARRIER_REQUEST` forces the switch to process the change before
20 responding. The flow routes on any GxOFSs are installed after installing the GTP tunnel endpoint route on the data plane SGSN and prior to installing the GTP tunnel endpoint route on the data plane GGSN, as illustrated in Figure 21. The OpenFlow controller does not respond to the calling entity until all flow routing updates have been accomplished.

Alternate Implementations

In other embodiments, the split 3G PC architecture can be implemented in non-cloud and non-virtualized systems. The control plane entities of the 3G PC architecture can be stored and executed on a single server or distributed across any number of
5 servers or similar computing devices. Similarly, the control plane entities can be executed as standard software code and modules without virtualization or similar systems. These control plane entities can communicate with one another through local
10 system or procedure calls, remote procedure calls or similar mechanisms. In further embodiments, a subset of the control plane entities can be virtualized or executed in a cloud computing system while another subset of the control plane entities can be
executed in a server, distributed server system or similar system. The control plane
entities can communicate with the data plane through the use of the OpenFlow protocol
as described herein above or through other control protocols as described herein below.

The cloud computing system described herein above is provided by way of
15 example and not by way of limitation. One skilled in the art would understand that the principles and features described above in relation to the cloud computing system can also be implemented in other configurations such as single servers or distributed server
systems. Similar principles and features to those described above can be implemented
in single server systems, distributed server systems and similar computing
20 environments. These principles and features can also be implemented using a non-virtualized environment including non-virtualized control plane entities that are
executed in any combination of cloud computing systems, single servers, distributed
server systems and similar systems.

In other embodiments, other control protocols can be utilized in place of OpenFlow as described herein. The use of OpenFlow is presented by way of example and not limitation. Other control protocols can also be utilized to manage the communication between the control plane and data plane and configuration of the data plane of the split 3G PC architecture. An example of such a protocol is FORCES, an IETF standard protocol for splitting the control plane and forwarding plane in networks. The FORCES protocol specification is described in RFC 5810. RFC 5812 describes the architecture of a FORCES forwarding element, the equivalent of an OpenFlow switch. The FORCES protocol itself does not directly support programming routes into the forwarding element, it is, instead, a framework for handling the interaction between the FORCES controller and a FORCES forwarding element. The forwarding element architecture describes how to design the protocol that actually allows a FORCES controller to program a FORCES forwarding element. One skilled in the art would understand that a FORCES based system could include features described herein above in relation to the OpenFlow embodiment, such as the GTP OpenFlow extension, to allow the controller to program the switches for GTP TEID routing.

FORCES and OpenFlow are provided by way of example and not limitation. One skilled in the art would understand that the principles and features described above in relation to the FORCES and OpenFlow protocols can also be implemented in other similar control protocols.

As used herein a network device (ND) is an electronic device that communicatively interconnects other electronic devices on the network (e.g., other network devices, end-user devices). Some network devices are “multiple services network devices” that provide support for multiple networking functions (e.g., routing,

bridging, switching, Layer 2 aggregation, session border control, Quality of Service, and/or subscriber management), and/or provide support for multiple application services (e.g., data, voice, and video). Network devices can include the devices implementing the OpenFlow switches of the data plane, the devices implementing the SGSN-D, GGSN-D and similar devices in the network. The

Figure 22A illustrates connectivity between network devices (NDs) within an exemplary network, as well as three exemplary implementations of the NDs, according to some embodiments of the invention. Figure 22A shows NDs 2200A-H, and their connectivity by way of lines between A-B, B-C, C-D, D-E, E-F, F-G, and A-G, as well as between H and each of A, C, D, and G. These NDs are physical devices, and the connectivity between these NDs can be wireless or wired (often referred to as a link). An additional line extending from NDs 2200A, E, and F illustrates that these NDs act as ingress and egress points for the network (and thus, these NDs are sometimes referred to as edge NDs; while the other NDs may be called core NDs).

Two of the exemplary ND implementations in Figure 22A are: 1) a special-purpose network device 2202 that uses custom application-specific integrated-circuits (ASICs) and a proprietary operating system (OS); and 2) a general purpose network device 2204 that uses common off-the-shelf (COTS) processors and a standard OS.

The special-purpose network device 2202 includes networking hardware 2210 comprising compute resource(s) 2212 (which typically include a set of one or more processors), forwarding resource(s) 2214 (which typically include one or more ASICs and/or network processors), and physical network interfaces (NIs) 2216 (sometimes called physical ports), as well as non-transitory machine readable storage media 2218 having stored therein networking software 2220. A physical NI is hardware in a ND

through which a network connection (e.g., wirelessly through a wireless network interface controller (WNIC) or through plugging in a cable to a physical port connected to a network interface controller (NIC)) is made, such as those shown by the connectivity between NDs 2200A-H. During operation, the networking software 2220
5 may be executed by the networking hardware 2210 to instantiate a set of one or more networking software instance(s) 2222. Each of the networking software instance(s) 2222, and that part of the networking hardware 2210 that executes that network software instance (be it hardware dedicated to that networking software instance and/or time slices of hardware temporally shared by that networking software instance with
10 others of the networking software instance(s) 2222), form a separate virtual network element 2230A-R. Each of the virtual network element(s) (VNEs) 2230A-R includes a control communication and configuration module 2232A-R (sometimes referred to as a local control module or control communication module) and forwarding table(s) 2234A-R, such that a given virtual network element (e.g., 2230A) includes the control
15 communication and configuration module (e.g., 2232A), a set of one or more forwarding table(s) (e.g., 2234A), and that portion of the networking hardware 2210 that executes the virtual network element (e.g., 2230A).

In the embodiments described herein above, the network devices can implement any of the data plane functionality of the 3G PC as a set of data plane modules 2233A.
20 The data plane modules 2233A can be configured via a ND control plane 2224 and interface with the forwarding plane 2226 of the ND. In some embodiments, the data plane modules can implement the functionality of the SGSN data plane (SGSN-D) or GGSN data plane (GGSN-D) or similar components in the 3G PC.

The special-purpose network device 2202 is often physically and/or logically considered to include: 1) the ND control plane 2224 (sometimes referred to as a control plane) comprising the compute resource(s) 2212 that execute the control communication and configuration module(s) 2232A-R; and 2) a ND forwarding plane 2226 (sometimes referred to as a forwarding plane, a data plane, or a media plane) comprising the forwarding resource(s) 2214 that utilize the forwarding table(s) 2234A-R and the physical NIs 2216. By way of example, where the ND is a router (or is implementing routing functionality), the ND control plane 2224 (the compute resource(s) 2212 executing the control communication and configuration module(s) 2232A-R) is typically responsible for participating in controlling how data (e.g., packets) is to be routed (e.g., the next hop for the data and the outgoing physical NI for that data) and storing that routing information in the forwarding table(s) 2234A-R, and the ND forwarding plane 2226 is responsible for receiving that data on the physical NIs 2216 and forwarding that data out the appropriate ones of the physical NIs 2216 based on the forwarding table(s) 2234A-R.

Figure 22B illustrates an exemplary way to implement the special-purpose network device 2202 according to some embodiments of the invention. Figure 22B shows a special-purpose network device including cards 2238 (typically hot pluggable). While in some embodiments the cards 2238 are of two types (one or more that operate as the ND forwarding plane 2226 (sometimes called line cards), and one or more that operate to implement the ND control plane 2224 (sometimes called control cards)), alternative embodiments may combine functionality onto a single card and/or include additional card types (e.g., one additional type of card is called a service card, resource card, or multi-application card). A service card can provide specialized processing

(e.g., Layer 4 to Layer 7 services (e.g., firewall, Internet Protocol Security (IPsec) (RFC 4301 and 4309), Secure Sockets Layer (SSL) / Transport Layer Security (TLS), Intrusion Detection System (IDS), peer-to-peer (P2P), Voice over IP (VoIP) Session Border Controller, Mobile Wireless Gateways (Gateway General Packet Radio Service (GPRS) Support Node (GGSN), Evolved Packet Core (EPC) Gateway)). By way of example, a service card may be used to terminate IPsec tunnels and execute the attendant authentication and encryption algorithms. These cards are coupled together through one or more interconnect mechanisms illustrated as backplane 2236 (e.g., a first full mesh coupling the line cards and a second full mesh coupling all of the cards).

10 Returning to Figure 22A, the general purpose network device 2204 includes hardware 2240 comprising a set of one or more processor(s) 2242 (which are often COTS processors) and network interface controller(s) 2244 (NICs; also known as network interface cards) (which include physical NIs 2246), as well as non-transitory machine readable storage media 2248 having stored therein software 2250. During operation, the processor(s) 2242 execute the software 2250 to instantiate a hypervisor 15 2254 (sometimes referred to as a virtual machine monitor (VMM)) and one or more virtual machines 2262A-R that are run by the hypervisor 2254, which are collectively referred to as software instance(s) 2252. A virtual machine is a software implementation of a physical machine that runs programs as if they were executing on a 20 physical, non-virtualized machine; and applications generally do not know they are running on a virtual machine as opposed to running on a “bare metal” host electronic device, though some systems provide para-virtualization which allows an operating system or application to be aware of the presence of virtualization for optimization purposes. Each of the virtual machines 2262A-R, and that part of the hardware 2240

that executes that virtual machine (be it hardware dedicated to that virtual machine and/or time slices of hardware temporally shared by that virtual machine with others of the virtual machine(s) 2262A-R), forms a separate virtual network element(s) 2260A-R.

The virtual network element(s) 2260A-R perform similar functionality to the
5 virtual network element(s) 2230A-R. For instance, the hypervisor 2254 may present a virtual operating platform that appears like networking hardware 2210 to virtual machine 2262A, and the virtual machine 2262A may be used to implement functionality similar to the control communication and configuration module(s) 2232A and forwarding table(s) 2234A (this virtualization of the hardware 2240 is sometimes
10 referred to as network function virtualization (NFV)). Thus, NFV may be used to consolidate many network equipment types onto industry standard high volume server hardware, physical switches, and physical storage, which could be located in Data centers, NDs, and customer premise equipment (CPE). However, different embodiments of the invention may implement one or more of the virtual machine(s)
15 2262A-R differently. For example, while embodiments of the invention are illustrated with each virtual machine 2262A-R corresponding to one VNE 2260A-R, alternative embodiments may implement this correspondence at a finer level granularity (e.g., line card virtual machines virtualize line cards, control card virtual machine virtualize control cards, etc.); it should be understood that the techniques described herein with
20 reference to a correspondence of virtual machines to VNEs also apply to embodiments where such a finer level of granularity is used.

In certain embodiments, the hypervisor 2254 includes a virtual switch that provides similar forwarding services as a physical Ethernet switch. Specifically, this virtual switch forwards traffic between virtual machines and the NIC(s) 2244, as well

as optionally between the virtual machines 2262A-R; in addition, this virtual switch may enforce network isolation between the VNEs 2260A-R that by policy are not permitted to communicate with each other (e.g., by honoring virtual local area networks (VLANs)).

5 In the embodiments described herein above, the virtual machines 2262A-R can implement any of the data plane functionality of the 3G PC as the set of data plane modules 2263A. The data plane modules 2263A can control the forwarding plane of the local ND or any of the other ND in the network of the 3G PC. In some embodiments, the data plane modules can implement the entire functionality or a portion of the
10 functionality of the SGSN data plane (SGSN-D) or GGSN data plane (GGSN-D) or similar components in the 3G PC in a distributed scheme within the set of ND of the network of the 3G PC.

 The third exemplary ND implementation in Figure 22A is a hybrid network device 2206, which includes both custom ASICs/proprietary OS and COTS
15 processors/standard OS in a single ND or a single card within an ND. In certain embodiments of such a hybrid network device, a platform VM (i.e., a VM that that implements the functionality of the special-purpose network device 2202) could provide for para-virtualization to the networking hardware present in the hybrid network device 2206.

20 Regardless of the above exemplary implementations of an ND, when a single one of multiple VNEs implemented by an ND is being considered (e.g., only one of the VNEs is part of a given virtual network) or where only a single VNE is currently being implemented by an ND, the shortened term network element (NE) is sometimes used to refer to that VNE. Also in all of the above exemplary implementations, each of the

VNEs (e.g., VNE(s) 2230A-R, VNEs 2260A-R, and those in the hybrid network device 2206) receives data on the physical NIs (e.g., 2216, 2246) and forwards that data out the appropriate ones of the physical NIs (e.g., 2216, 2246). For example, a VNE implementing IP router functionality forwards IP packets on the basis of some of the IP header information in the IP packet; where IP header information includes source IP address, destination IP address, source port, destination port (where “source port” and “destination port” refer herein to protocol ports, as opposed to physical ports of a ND), transport protocol (e.g., user datagram protocol (UDP) (RFC 768, 2460, 2675, 4113, and 5405), Transmission Control Protocol (TCP) (RFC 793 and 1180), and differentiated services (DSCP) values (RFC 2474, 2475, 2597, 2983, 3086, 3140, 3246, 3247, 3260, 4594, 5865, 3289, 3290, and 3317).

Figure 22C illustrates various exemplary ways in which VNEs may be coupled according to some embodiments of the invention. Figure 22C shows VNEs 2270A.1-2270A.P (and optionally VNEs 2280A.Q-2280A.R) implemented in ND 2200A and VNE 2270H.1 in ND 2200H. In Figure 22C, VNEs 2270A.1-P are separate from each other in the sense that they can receive packets from outside ND 2200A and forward packets outside of ND 2200A; VNE 2270A.1 is coupled with VNE 2270H.1, and thus they communicate packets between their respective NDs; VNE 2270A.2-2270A.3 may optionally forward packets between themselves without forwarding them outside of the ND 2200A; and VNE 2270A.P may optionally be the first in a chain of VNEs that includes VNE 2270A.Q followed by VNE 2270A.R (this is sometimes referred to as dynamic service chaining, where each of the VNEs in the series of VNEs provides a different service – e.g., one or more layer 4-7 network services). While Figure 22C illustrates various exemplary relationships between the VNEs, alternative embodiments

may support other relationships (e.g., more/fewer VNEs, more/fewer dynamic service chains, multiple different dynamic service chains with some common VNEs and some different VNEs).

The NDs of Figure 22A, for example, may form part of the Internet or a private
5 network; and other electronic devices (not shown; such as end user devices including workstations, laptops, netbooks, tablets, palm tops, mobile phones, smartphones, multimedia phones, Voice Over Internet Protocol (VOIP) phones, terminals, portable media players, GPS units, wearable devices, gaming systems, set-top boxes, Internet enabled household appliances) may be coupled to the network (directly or through
10 other networks such as access networks) to communicate over the network (e.g., the Internet or virtual private networks (VPNs) overlaid on (e.g., tunneled through) the Internet) with each other (directly or through servers) and/or access content and/or services. Such content and/or services are typically provided by one or more servers (not shown) belonging to a service/content provider or one or more end user devices
15 (not shown) participating in a peer-to-peer (P2P) service, and may include, for example, public webpages (e.g., free content, store fronts, search services), private webpages (e.g., username/password accessed webpages providing email services), and/or corporate networks over VPNs. For instance, end user devices may be coupled (e.g., through customer premise equipment coupled to an access network (wired or
20 wirelessly)) to edge NDs, which are coupled (e.g., through one or more core NDs) to other edge NDs, which are coupled to electronic devices acting as servers. However, through compute and storage virtualization, one or more of the electronic devices operating as the NDs in Figure 22A may also host one or more such servers (e.g., in the case of the general purpose network device 2204, one or more of the virtual machines

2262A-R may operate as servers; the same would be true for the hybrid network device 2206; in the case of the special-purpose network device 2202, one or more such servers could also be run on a hypervisor executed by the compute resource(s) 2212); in which case the servers are said to be co-located with the VNEs of that ND.

5 A virtual network is a logical abstraction of a physical network (such as that in Figure 22A) that provides network services (e.g., L2 and/or L3 services). A virtual network can be implemented as an overlay network (sometimes referred to as a network virtualization overlay) that provides network services (e.g., layer 2 (L2, data link layer) and/or layer 3 (L3, network layer) services) over an underlay network (e.g.,
10 an L3 network, such as an Internet Protocol (IP) network that uses tunnels (e.g., generic routing encapsulation (GRE), layer 2 tunneling protocol (L2TP), IPSec) to create the overlay network).

 A network virtualization edge (NVE) sits at the edge of the underlay network and participates in implementing the network virtualization; the network-facing side of
15 the NVE uses the underlay network to tunnel frames to and from other NVEs; the outward-facing side of the NVE sends and receives data to and from systems outside the network. A virtual network instance (VNI) is a specific instance of a virtual network on a NVE (e.g., a NE/VNE on an ND, a part of a NE/VNE on a ND where that NE/VNE is divided into multiple VNEs through emulation); one or more VNIs can be
20 instantiated on an NVE (e.g., as different VNEs on an ND). A virtual access point (VAP) is a logical connection point on the NVE for connecting external systems to a virtual network; a VAP can be physical or virtual ports identified through logical interface identifiers (e.g., a VLAN ID).

Examples of network services include: 1) an Ethernet LAN emulation service (an Ethernet-based multipoint service similar to an Internet Engineering Task Force (IETF) Multiprotocol Label Switching (MPLS) or Ethernet VPN (EVPN) service) in which external systems are interconnected across the network by a LAN environment
5 over the underlay network (e.g., an NVE provides separate L2 VNIs (virtual switching instances) for different such virtual networks, and L3 (e.g., IP/MPLS) tunneling encapsulation across the underlay network); and 2) a virtualized IP forwarding service (similar to IETF IP VPN (e.g., Border Gateway Protocol (BGP)/MPLS IPVPN RFC 4364) from a service definition perspective) in which external systems are
10 interconnected across the network by an L3 environment over the underlay network (e.g., an NVE provides separate L3 VNIs (forwarding and routing instances) for different such virtual networks, and L3 (e.g., IP/MPLS) tunneling encapsulation across the underlay network)). Network services may also include quality of service capabilities (e.g., traffic classification marking, traffic conditioning and scheduling),
15 security capabilities (e.g., filters to protect customer premises from network -- originated attacks, to avoid malformed route announcements), and management capabilities (e.g., full detection and processing).

Fig. 22D illustrates a network with a single network element on each of the NDs of Figure 22A, and within this straight forward approach contrasts a traditional
20 distributed approach (commonly used by traditional routers) with a centralized approach for maintaining reachability and forwarding information (also called network control), according to some embodiments of the invention. Specifically, Figure 22D illustrates network elements (NEs) 2270A-H with the same connectivity as the NDs 2200A-H of Figure 22A.

Figure 22D illustrates that the distributed approach 2272 distributes responsibility for generating the reachability and forwarding information across the NEs 2270A-H; in other words, the process of neighbor discovery and topology discovery is distributed.

5 For example, where the special-purpose network device 2202 is used, the control communication and configuration module(s) 2232A-R of the ND control plane 2224 typically include a reachability and forwarding information module to implement one or more routing protocols (e.g., an exterior gateway protocol such as Border
10 Gateway Protocol (BGP) (RFC 4271), Interior Gateway Protocol(s) (IGP) (e.g., Open Shortest Path First (OSPF) (RFC 2328 and 5340), Intermediate System to Intermediate System (IS-IS) (RFC 1142), Routing Information Protocol (RIP) (version 1 RFC 1058, version 2 RFC 2453, and next generation RFC 2080)), Label Distribution Protocol (LDP) (RFC 5036), Resource Reservation Protocol (RSVP) (RFC 2205, 2210, 2211, 2212, as well as RSVP-Traffic Engineering (TE): Extensions to RSVP for LSP Tunnels
15 RFC 3209, Generalized Multi-Protocol Label Switching (GMPLS) Signaling RSVP-TE RFC 3473, RFC 3936, 4495, and 4558)) that communicate with other NEs to exchange routes, and then selects those routes based on one or more routing metrics. Thus, the NEs 2270A-H (e.g., the compute resource(s) 2212 executing the control
20 communication and configuration module(s) 2232A-R) perform their responsibility for participating in controlling how data (e.g., packets) is to be routed (e.g., the next hop for the data and the outgoing physical NI for that data) by distributively determining the reachability within the network and calculating their respective forwarding information. Routes and adjacencies are stored in one or more routing structures (e.g., Routing Information Base (RIB), Label Information Base (LIB), one or more adjacency

structures) on the ND control plane 2224. The ND control plane 2224 programs the ND forwarding plane 2226 with information (e.g., adjacency and route information) based on the routing structure(s). For example, the ND control plane 2224 programs the adjacency and route information into one or more forwarding table(s) 2234A-R (e.g., Forwarding Information Base (FIB), Label Forwarding Information Base (LFIB), and one or more adjacency structures) on the ND forwarding plane 2226. For layer 2 forwarding, the ND can store one or more bridging tables that are used to forward data based on the layer 2 information in that data. While the above example uses the special-purpose network device 2202, the same distributed approach 172 can be implemented on the general purpose network device 2204 and the hybrid network device 2206.

Figure 22D illustrates that a centralized approach 2274 (also known as software defined networking (SDN)) that decouples the system that makes decisions about where traffic is sent from the underlying systems that forwards traffic to the selected destination. The illustrated centralized approach 2274 has the responsibility for the generation of reachability and forwarding information in a centralized control plane 2276 (sometimes referred to as a SDN control module, controller, network controller, OpenFlow controller, SDN controller, control plane node, network virtualization authority, or management control entity), and thus the process of neighbor discovery and topology discovery is centralized. The centralized control plane 2276 has a south bound interface 2282 with a data plane 2280 (sometimes referred to the infrastructure layer, network forwarding plane, or forwarding plane (which should not be confused with a ND forwarding plane)) that includes the NEs 2270A-H (sometimes referred to as switches, forwarding elements, data plane elements, or nodes). In one embodiment, the centralized control plane 2276 includes the set of control plane modules 2281 as

described herein above, which implement the functions of the 3G PC. As discussed further herein below and above, these functions can be implemented also at the application layer 228 6 or via a distributed approach 2272 or any combination thereof with different control plane module 2281 functions of the 3G PC handled with any of these approaches. The centralized control plane 2276 includes a network controller 2278, which includes a centralized reachability and forwarding information module 2279 that determines the reachability within the network and distributes the forwarding information to the NEs 2270A-H of the data plane 2280 over the south bound interface 2282 (which may use the OpenFlow protocol). Thus, the network intelligence is centralized in the centralized control plane 2276 executing on electronic devices that are typically separate from the NDs.

For example, where the special-purpose network device 2202 is used in the data plane 2280, each of the control communication and configuration module(s) 2232A-R of the ND control plane 2224 typically include a control agent that provides the VNE side of the south bound interface 2282. In this case, the ND control plane 2224 (the compute resource(s) 2212 executing the control communication and configuration module(s) 2232A-R) performs its responsibility for participating in controlling how data (e.g., packets) is to be routed (e.g., the next hop for the data and the outgoing physical NI for that data) through the control agent communicating with the centralized control plane 2276 to receive the forwarding information (and in some cases, the reachability information) from the centralized reachability and forwarding information module 2279 (it should be understood that in some embodiments of the invention, the control communication and configuration module(s) 2232A-R, in addition to communicating with the centralized control plane 2276, may also play some role in

determining reachability and/or calculating forwarding information -- albeit less so than in the case of a distributed approach; such embodiments are generally considered to fall under the centralized approach 2274, but may also be considered a hybrid approach).

While the above example uses the special-purpose network device 2202, the same centralized approach 174 can be implemented with the general purpose network device 2204 (e.g., each of the VNE 2260A-R performs its responsibility for controlling how data (e.g., packets) is to be routed (e.g., the next hop for the data and the outgoing physical NI for that data) by communicating with the centralized control plane 2276 to receive the forwarding information (and in some cases, the reachability information) from the centralized reachability and forwarding information module 2279; it should be understood that in some embodiments of the invention, the VNEs 2260A-R, in addition to communicating with the centralized control plane 2276, may also play some role in determining reachability and/or calculating forwarding information -- albeit less so than in the case of a distributed approach) and the hybrid network device 2206. In fact, the use of SDN techniques can enhance the NFV techniques typically used in the general purpose network device 2204 or hybrid network device 2206 implementations as NFV is able to support SDN by providing an infrastructure upon which the SDN software can be run, and NFV and SDN both aim to make use of commodity server hardware and physical switches.

Figure 22D also shows that the centralized control plane 2276 has a north bound interface 2284 to an application layer 2286, in which resides application(s) 2288. The centralized control plane 2276 has the ability to form virtual networks 2292 (sometimes referred to as a logical forwarding plane, network services, or overlay networks (with the NEs 2270A-H of the data plane 2280 being the underlay network)) for the

application(s) 2288. Thus, the centralized control plane 2276 maintains a global view of all NDs and configured NEs/VNEs, and it maps the virtual networks to the underlying NDs efficiently (including maintaining these mappings as the physical network changes either through hardware (ND, link, or ND component) failure, addition, or removal).

While Figure 22D shows the distributed approach 2272 separate from the centralized approach 2274, the effort of network control may be distributed differently or the two combined in certain embodiments of the invention. For example: 1) embodiments may generally use the centralized approach (SDN) 2274, but have certain functions delegated to the NEs (e.g., the distributed approach may be used to implement one or more of fault monitoring, performance monitoring, protection switching, and primitives for neighbor and/or topology discovery); or 2) embodiments of the invention may perform neighbor discovery and topology discovery via both the centralized control plane and the distributed protocols, and the results compared to raise exceptions where they do not agree. Such embodiments are generally considered to fall under the centralized approach 2274, but may also be considered a hybrid approach.

While Figure 22D illustrates the simple case where each of the NDs 2200A-H implements a single NE 2270A-H, it should be understood that the network control approaches described with reference to Figure 22D also work for networks where one or more of the NDs 2200A-H implement multiple VNEs (e.g., VNEs 2230A-R, VNEs 2260A-R, those in the hybrid network device 2206). Alternatively or in addition, the network controller 2278 may also emulate the implementation of multiple VNEs in a single ND. Specifically, instead of (or in addition to) implementing multiple VNEs in a single ND, the network controller 2278 may present the implementation of a VNE/NE

in a single ND as multiple VNEs in the virtual networks 2292 (all in the same one of the virtual network(s) 2292, each in different ones of the virtual network(s) 2292, or some combination). For example, the network controller 2278 may cause an ND to implement a single VNE (a NE) in the underlay network, and then logically divide up the resources of that NE within the centralized control plane 2276 to present different VNEs in the virtual network(s) 2292 (where these different VNEs in the overlay networks are sharing the resources of the single VNE/NE implementation on the ND in the underlay network).

On the other hand, Figures 22E and 22F respectively illustrate exemplary abstractions of NEs and VNEs that the network controller 2278 may present as part of different ones of the virtual networks 2292. Figure 22E illustrates the simple case of where each of the NDs 2200A-H implements a single NE 2270A-H (see Figure 22D), but the centralized control plane 2276 has abstracted multiple of the NEs in different NDs (the NEs 2270A-C and G-H) into (to represent) a single NE 2270I in one of the virtual network(s) 2292 of Figure 22D, according to some embodiments of the invention. Figure 22E shows that in this virtual network, the NE 2270I is coupled to NE 2270D and 2270F, which are both still coupled to NE 2270E.

Figure 22F illustrates a case where multiple VNEs (VNE 2270A.1 and VNE 2270H.1) are implemented on different NDs (ND 2200A and ND 2200H) and are coupled to each other, and where the centralized control plane 2276 has abstracted these multiple VNEs such that they appear as a single VNE 2270T within one of the virtual networks 2292 of Figure 22D, according to some embodiments of the invention. Thus, the abstraction of a NE or VNE can span multiple NDs.

While some embodiments of the invention implement the centralized control plane 2276 as a single entity (e.g., a single instance of software running on a single electronic device), alternative embodiments may spread the functionality across multiple entities for redundancy and/or scalability purposes (e.g., multiple instances of software running on different electronic devices).

Similar to the network device implementations, the electronic device(s) running the centralized control plane 2276, and thus the network controller 2278 including the centralized reachability and forwarding information module 2279, may be implemented a variety of ways (e.g., a special purpose device, a general-purpose (e.g., COTS) device, or hybrid device). These electronic device(s) would similarly include compute resource(s), a set or one or more physical NICs, and a non-transitory machine-readable storage medium having stored thereon the centralized control plane software. For instance, Figure 23 illustrates, a general purpose control plane device 2304 including hardware 2340 comprising a set of one or more processor(s) 2342 (which are often COTS processors) and network interface controller(s) 2344 (NICs; also known as network interface cards) (which include physical NIs 2346), as well as non-transitory machine readable storage media 2348 having stored therein centralized control plane (CCP) software 2350. The control plane device 2304 can be one of many such devices networked together to form a cloud computing system or similar system as described herein above and in particular for implementing the control plane of a 3G PC.

In embodiments that use compute virtualization, the processor(s) 2342 typically execute software to instantiate a hypervisor 2354 (sometimes referred to as a virtual machine monitor (VMM)) and one or more virtual machines 2362A-R that are run by the hypervisor 2354; which are collectively referred to as software instance(s) 2352. A

virtual machine is a software implementation of a physical machine that runs programs as if they were executing on a physical, non-virtualized machine; and applications generally are not aware they are running on a virtual machine as opposed to running on a “bare metal” host electronic device, though some systems provide para-virtualization which allows an operating system or application to be aware of the presence of virtualization for optimization purposes. Again, in embodiments where compute virtualization is used, during operation an instance of the CCP software 2350 (illustrated as CCP instance 2376A) on top of an operating system 2364A are typically executed within the virtual machine 2362A. In embodiments where compute virtualization is not used, the CCP instance 2376A on top of operating system 2364A is executed on the “bare metal” general purpose control plane device 2304.

The operating system 2364A provides basic processing, input/output (I/O), and networking capabilities. In some embodiments, the CCP instance 2376A includes a network controller instance 2378. The network controller instance 2378 includes a centralized reachability and forwarding information module instance 2379 (which is a middleware layer providing the context of the network controller 2278 to the operating system 2364A and communicating with the various NEs), and an CCP application layer 2380 (sometimes referred to as an application layer) over the middleware layer (providing the intelligence required for various network operations such as protocols, network situational awareness, and user – interfaces). At a more abstract level, this CCP application layer 2380 within the centralized control plane 2276 works with virtual network view(s) (logical view(s) of the network) and the middleware layer provides the conversion from the virtual networks to the physical view.

In one embodiment, any number or combination of the control plane modules 2381 that implement the functions of the control plane of the 3G PC can be executed by a virtual machine 2362A as part of a centralized control plane (CCP) instance 2376A (of the CCP software) or in a similar virtualized configuration. The control plane modules 2381 can include the AUC, HLR, HSS, SGSN-C, GGSN-C or similar control plane modules 2381 of the 3G PC. The control plane modules 2381 of a particular 3G PC can be executed by the same virtual machine 2362A or distributed over any number of virtual machines and over any number of control plane devices 2304. This distribution can be dynamic to enable load balancing and similar resource management within a cloud computing system. In one embodiment, the control plane modules 2381 can be part of the CCP application layer 2381 of a network controller instance 778.

The centralized control plane 2276 transmits relevant messages to the data plane 2280 based on CCP application layer 2380 calculations and middleware layer mapping for each flow. A flow may be defined as a set of packets whose headers match a given pattern of bits; in this sense, traditional IP forwarding is also flow-based forwarding where the flows are defined by the destination IP address for example; however, in other implementations, the given pattern of bits used for a flow definition may include more fields (e.g., 10 or more) in the packet headers. Different NDs/NEs/VNEs of the data plane 2280 may receive different messages, and thus different forwarding information. The data plane 2280 processes these messages and programs the appropriate flow information and corresponding actions in the forwarding tables (sometime referred to as flow tables) of the appropriate NE/VNEs, and then the NEs/VNEs map incoming packets to flows represented in the forwarding tables and forward packets based on the matches in the forwarding tables.

Standards such as OpenFlow define the protocols used for the messages, as well as a model for processing the packets. The model for processing packets includes header parsing, packet classification, and making forwarding decisions. Header parsing describes how to interpret a packet based upon a well-known set of protocols. Some
5 protocol fields are used to build a match structure (or key) that will be used in packet classification (e.g., a first key field could be a source media access control (MAC) address, and a second key field could be a destination MAC address).

Packet classification involves executing a lookup in memory to classify the packet by determining which entry (also referred to as a forwarding table entry or flow
10 entry) in the forwarding tables best matches the packet based upon the match structure, or key, of the forwarding table entries. It is possible that many flows represented in the forwarding table entries can correspond/match to a packet; in this case the system is typically configured to determine one forwarding table entry from the many according to a defined scheme (e.g., selecting a first forwarding table entry that is matched).

15 Forwarding table entries include both a specific set of match criteria (a set of values or wildcards, or an indication of what portions of a packet should be compared to a particular value/values/wildcards, as defined by the matching capabilities – for specific fields in the packet header, or for some other packet content), and a set of one or more actions for the data plane to take on receiving a matching packet. For example, an
20 action may be to push a header onto the packet, for the packet using a particular port, flood the packet, or simply drop the packet. Thus, a forwarding table entry for IPv4/IPv6 packets with a particular transmission control protocol (TCP) destination port could contain an action specifying that these packets should be dropped.

Making forwarding decisions and performing actions occurs, based upon the forwarding table entry identified during packet classification, by executing the set of actions identified in the matched forwarding table entry on the packet.

However, when an unknown packet (for example, a “missed packet” or a “match-miss” as used in OpenFlow parlance) arrives at the data plane 2280, the packet (or a subset of the packet header and content) is typically forwarded to the centralized control plane 2276. The centralized control plane 2276 will then program forwarding table entries into the data plane 2280 to accommodate packets belonging to the flow of the unknown packet. Once a specific forwarding table entry has been programmed into the data plane 2280 by the centralized control plane 2276, the next packet with matching credentials will match that forwarding table entry and take the set of actions associated with that matched entry.

A network interface (NI) may be physical or virtual; and in the context of IP, an interface address is an IP address assigned to a NI, be it a physical NI or virtual NI. A virtual NI may be associated with a physical NI, with another virtual interface, or stand on its own (e.g., a loopback interface, a point-to-point protocol interface). A NI (physical or virtual) may be numbered (a NI with an IP address) or unnumbered (a NI without an IP address). A loopback interface (and its loopback address) is a specific type of virtual NI (and IP address) of a NE/VNE (physical or virtual) often used for management purposes; where such an IP address is referred to as the nodal loopback address. The IP address(es) assigned to the NI(s) of a ND are referred to as IP addresses of that ND; at a more granular level, the IP address(es) assigned to NI(s) assigned to a NE/VNE implemented on a ND can be referred to as IP addresses of that NE/VNE.

Next hop selection by the routing system for a given destination may resolve to one path (that is, a routing protocol may generate one next hop on a shortest path); but if the routing system determines there are multiple viable next hops (that is, the routing protocol generated forwarding solution offers more than one next hop on a shortest path – multiple equal cost next hops), some additional criteria is used - for instance, in a connectionless network, Equal Cost Multi Path (ECMP) (also known as Equal Cost Multi Pathing, multipath forwarding and IP multipath) (RFC 2991 and 2992) may be used (e.g., typical implementations use as the criteria particular header fields to ensure that the packets of a particular packet flow are always forwarded on the same next hop to preserve packet flow ordering). For purposes of multipath forwarding, a packet flow is defined as a set of packets that share an ordering constraint. As an example, the set of packets in a particular TCP transfer sequence need to arrive in order, else the TCP logic will interpret the out of order delivery as congestion and slow the TCP transfer rate down.

While the invention has been described in terms of several embodiments, those skilled in the art will recognize that the invention is not limited to the embodiments described, can be practiced with modification and alteration within the spirit and scope of the appended claims. The description is thus to be regarded as illustrative instead of limiting.

CLAIMS

What is claimed is:

1. A control plane device (2304) in a cloud computing system to execute a plurality of virtual machines for implementing network function virtualization (NFV), the
5 control plane device operable to manage implementation of a general packet radio service (GPRS) tunnel protocol (GTP) in a packet core (PC) of a third generation (3G) network having a split architecture where a control plane of the PC of the 3G network is in the cloud computing system, the control plane to communicate with a data plane of the PC through a control plane protocol, the data plane implemented in a plurality of
10 network devices of the 3G network, the control plane device and the plurality of virtual machines operable to communicate with other control plane devices in the cloud computing system and with the plurality of network devices of the data plane, the control plane device comprising:
 - a storage medium (2348) to store a centralized control plane software (2350)
15 including control plane modules for implementing the control plane of the PC; and
 - a processor (2342) communicatively coupled to the storage medium, the processor operable to execute the plurality of virtual machines, where at least one of the plurality of virtual machines is operable to execute the centralized control plane (CCP) software including at least one of the control plane modules, each control plane module
20 to provide a set of control plane functions for managing the data plane, the CCP software operable to receive a request to create a GTP tunnel in the PC of the 3G network between a serving GPRS support node (SGSN) and a gateway GPRS support node (GGSN) for a subscriber session, the CCP software operable to configure a switch implementing a data plane of the SGSN, via the control plane protocol, to encapsulate

and decapsulate packets of the subscriber session and to establish a first GTP tunnel endpoint, the CCP software operable to configure at least one switch in a route of the GTP tunnel, via the control plane protocol, to forward packets of the subscriber session according to the GTP tunnel, and the CCP software operable to configure a switch
5 implementing a data plane of the GGSN, via the control plane protocol, to encapsulate and decapsulate the packets of the subscriber session and to establish a second GTP tunnel endpoint.

2. The control plane device of claim 1, wherein the processor is operable to
10 execute the CCP software to receive the request to create a GTP tunnel as a protocol data packet (PDP) context creation request from a control plane component of the SGSN.

3. The control plane device of claim 1, wherein the processor is operable to
15 execute the CCP software to receive the request to create a GTP tunnel as a GTP routing update request from a control plane component of the GGSN.

4. The control plane device of claim 1, wherein the processor is operable to
execute the CCP software to receive a request to modify the GTP tunnel in the PC of
20 the 3G network between the SGSN and GGSN for the subscriber session, the CCP is operable to modify configuration of the switch implementing the data plane of the SGSN, via the control plane protocol, to encapsulate and decapsulate packets of the subscriber session and to modify the first GTP tunnel endpoint, and the CCP software is operable to modify configuration of the switch implementing the data plane of the

GGSN, via the control plane protocol, to encapsulate and decapsulate the packets of the subscriber session and to modify the second GTP tunnel endpoint.

5. The control plane device of claim 4, wherein the processor is operable to
5 execute the CCP software to remove configuration of the at least one switch in the route of the GTP tunnel, via the control plane protocol, to end forwarding of packets of the subscriber session according to the GTP tunnel.

6. The control plane device of claim 5, wherein the processor is operable to
10 execute the CCP software to configure a switch in a new route of the GTP tunnel, via the control plane protocol, to forward packets of the subscriber session according to the GTP tunnel.

7. The control plane device of claim 1, wherein the processor is operable to
15 execute the CCP software to receive a request to delete the GTP tunnel in the PC of the 3G network between the SGSN and GGSN for the subscriber session, to remove configuration of the switch implementing the data plane of the SGSN, via the control plane protocol, to end encapsulation and decapsulation of packets of the subscriber session and to remove the first GTP tunnel endpoint, and to remove configuration of
20 the switch implementing the data plane of the GGSN, via the control plane protocol, to end encapsulation and decapsulation of the packets of the subscriber session and to remove the second GTP tunnel endpoint.

8. The control plane device of claim 1, wherein the control plane protocol is the OpenFlow protocol.

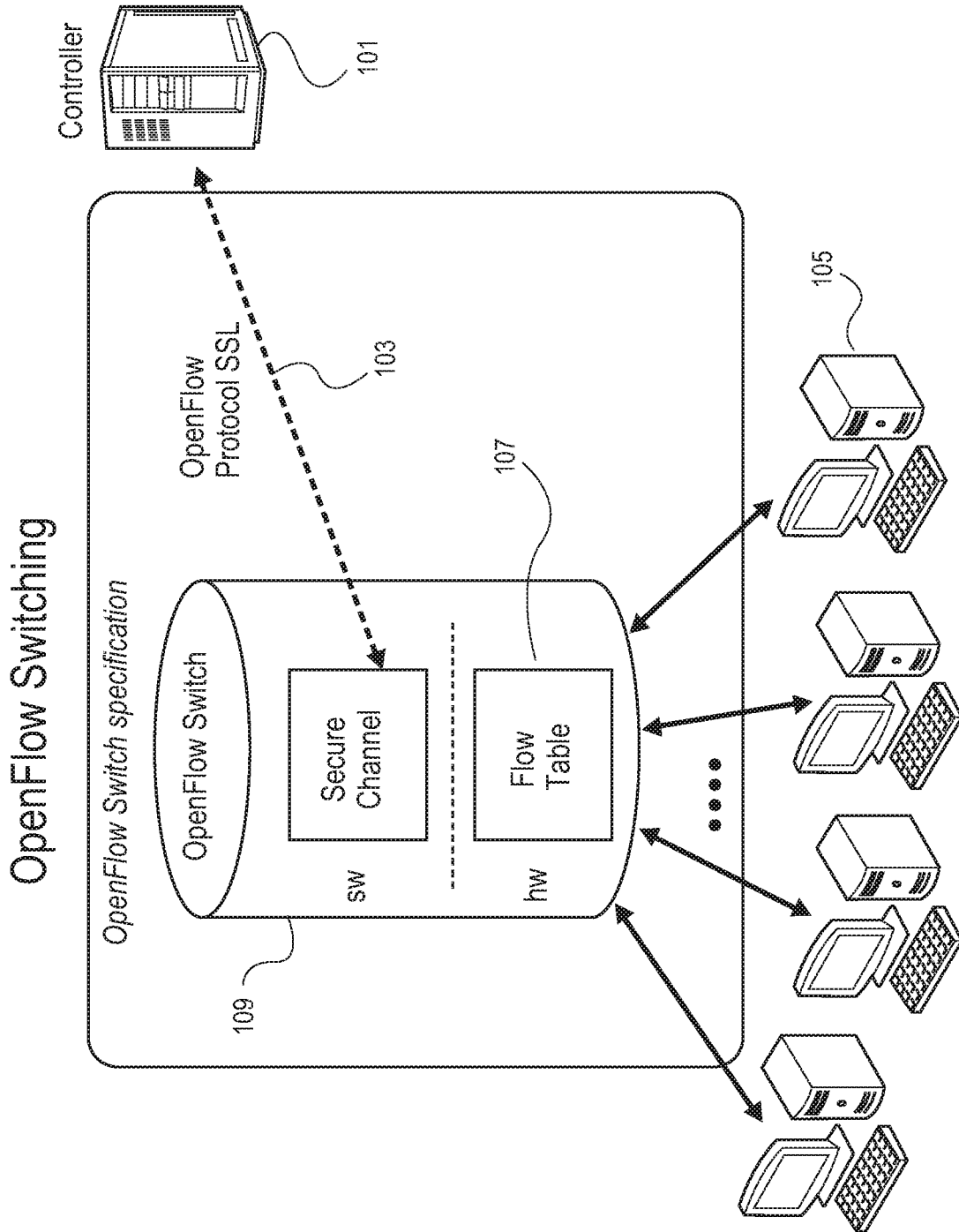


FIG. 1

Flow Table Entry

"Type 0" OpenFlow Switch

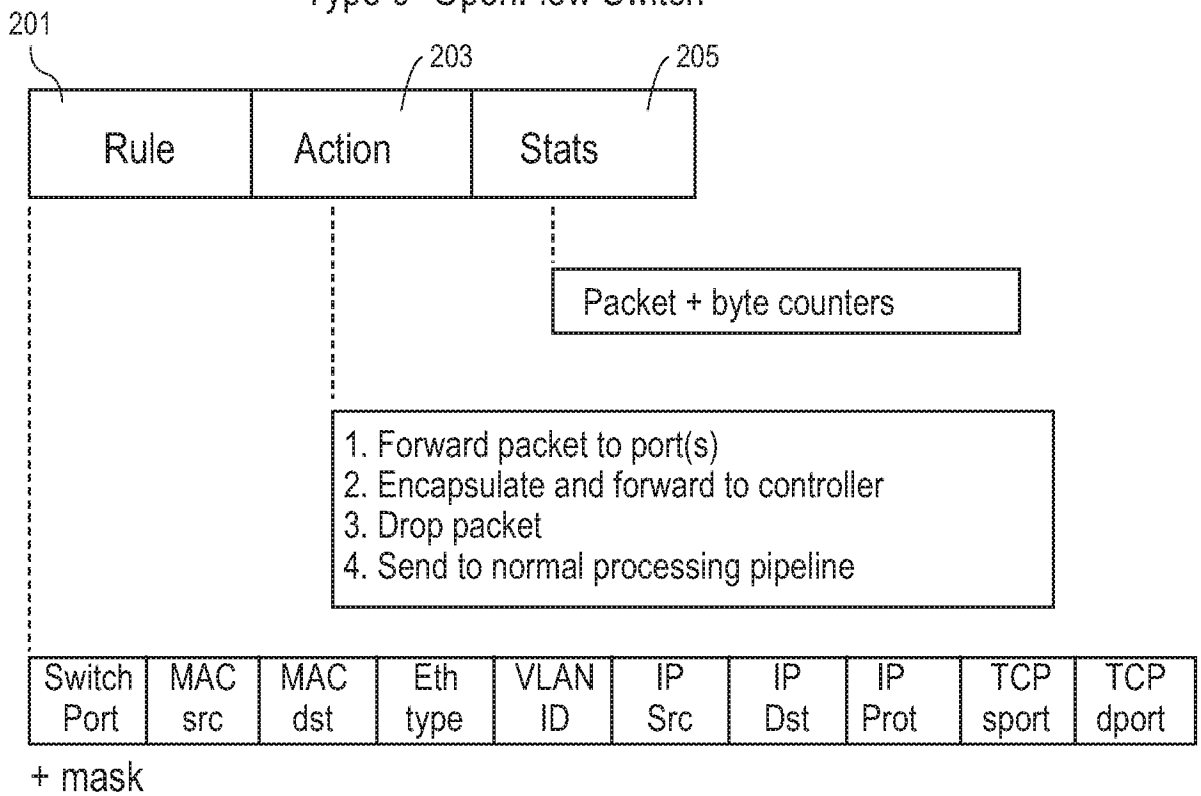


FIG. 2

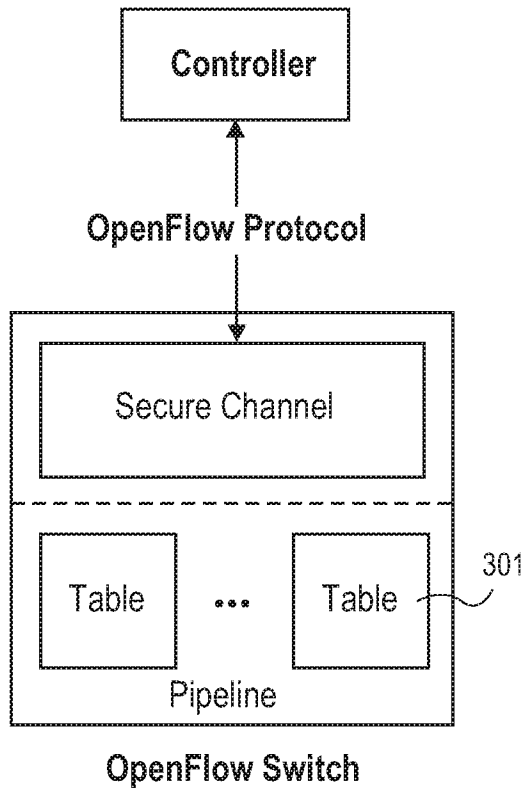


FIG. 3

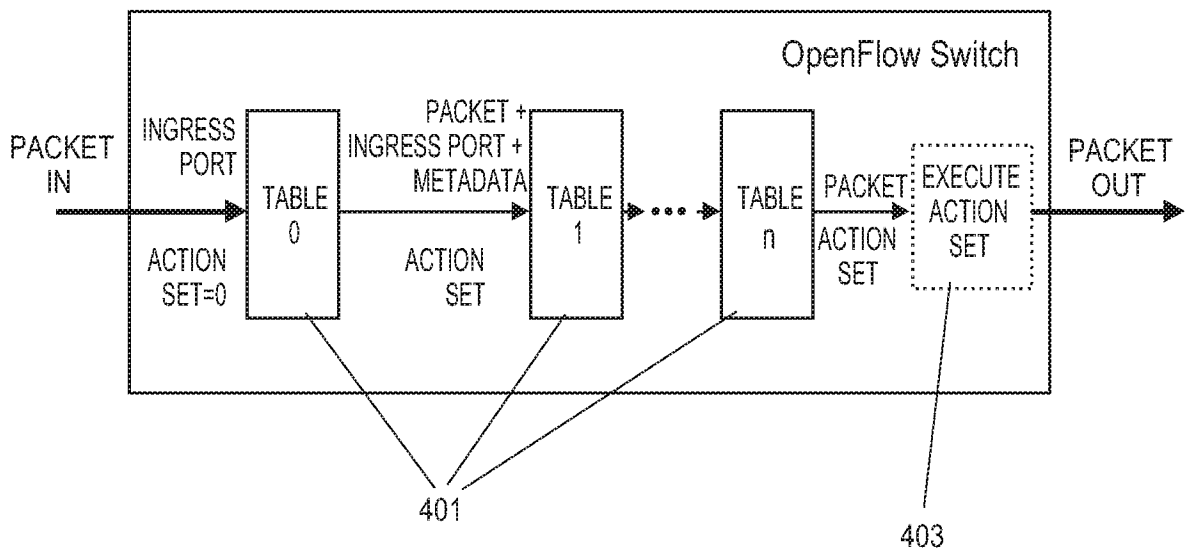


FIG. 4

Ingress Port
Metadata
Ether src
Ether dst
Ether type
VLAN id
VLAN priority
MPLS label
MPLS traffic class
IP src
IP dst
IP proto / ARP opcode
IP ToS bits
TCP / UDP / SCTP src port
ICMP Type
TCP / UDP / SCTP dst port
ICMP Code

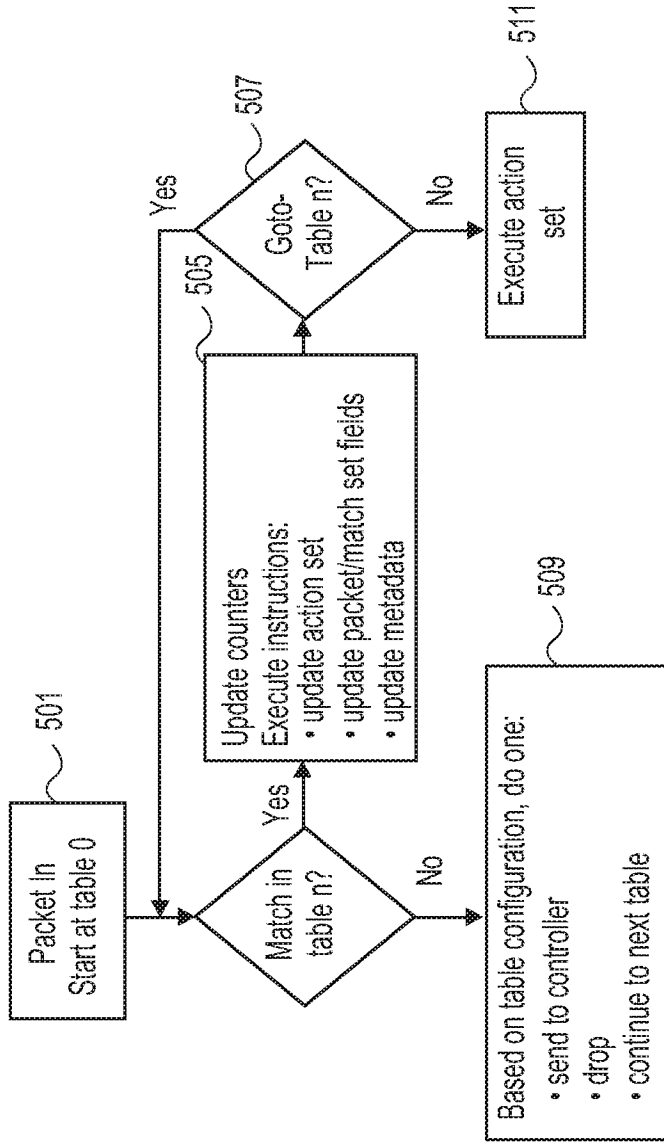


FIG. 5

FIG. 6

FIG. 7A

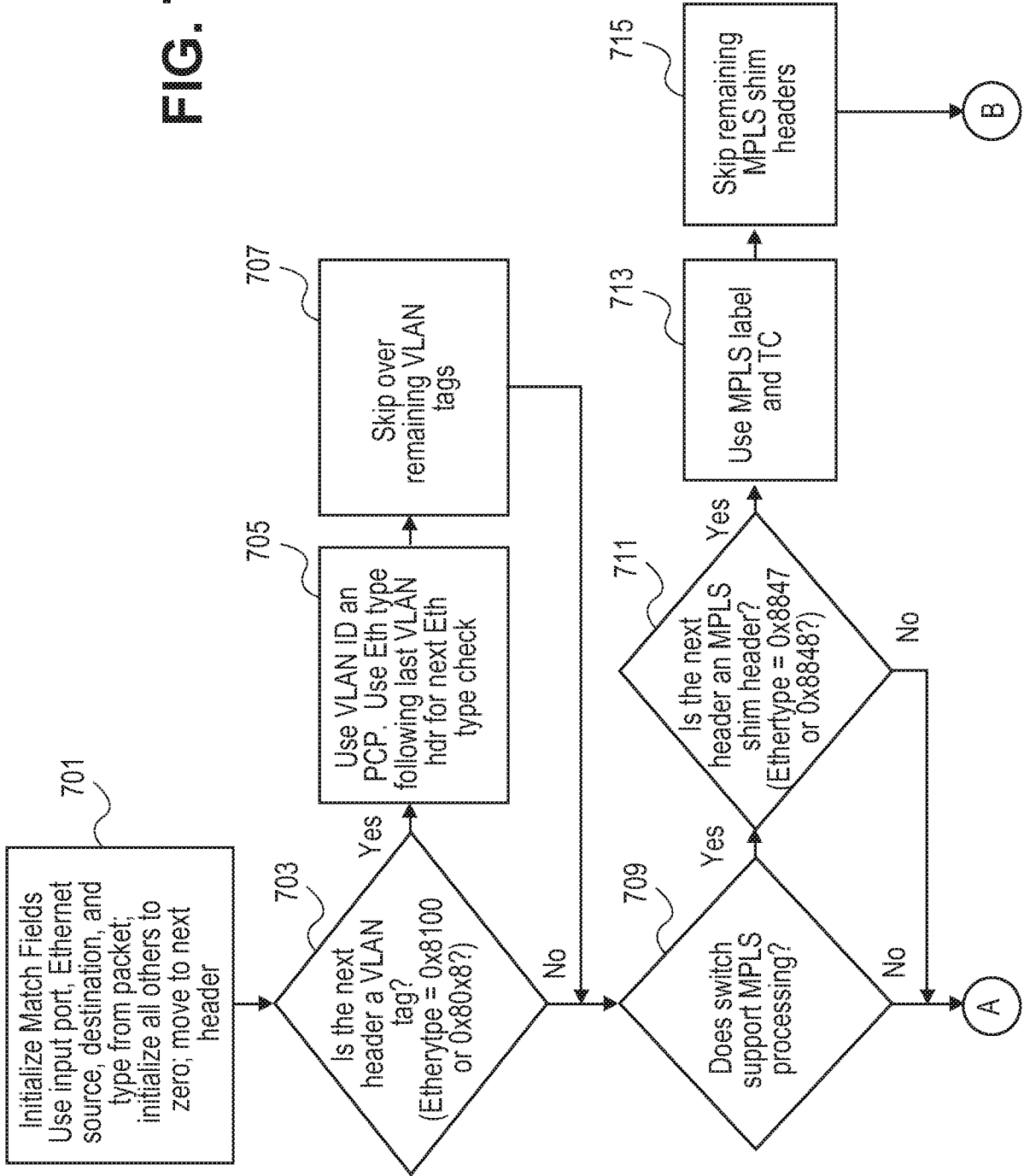
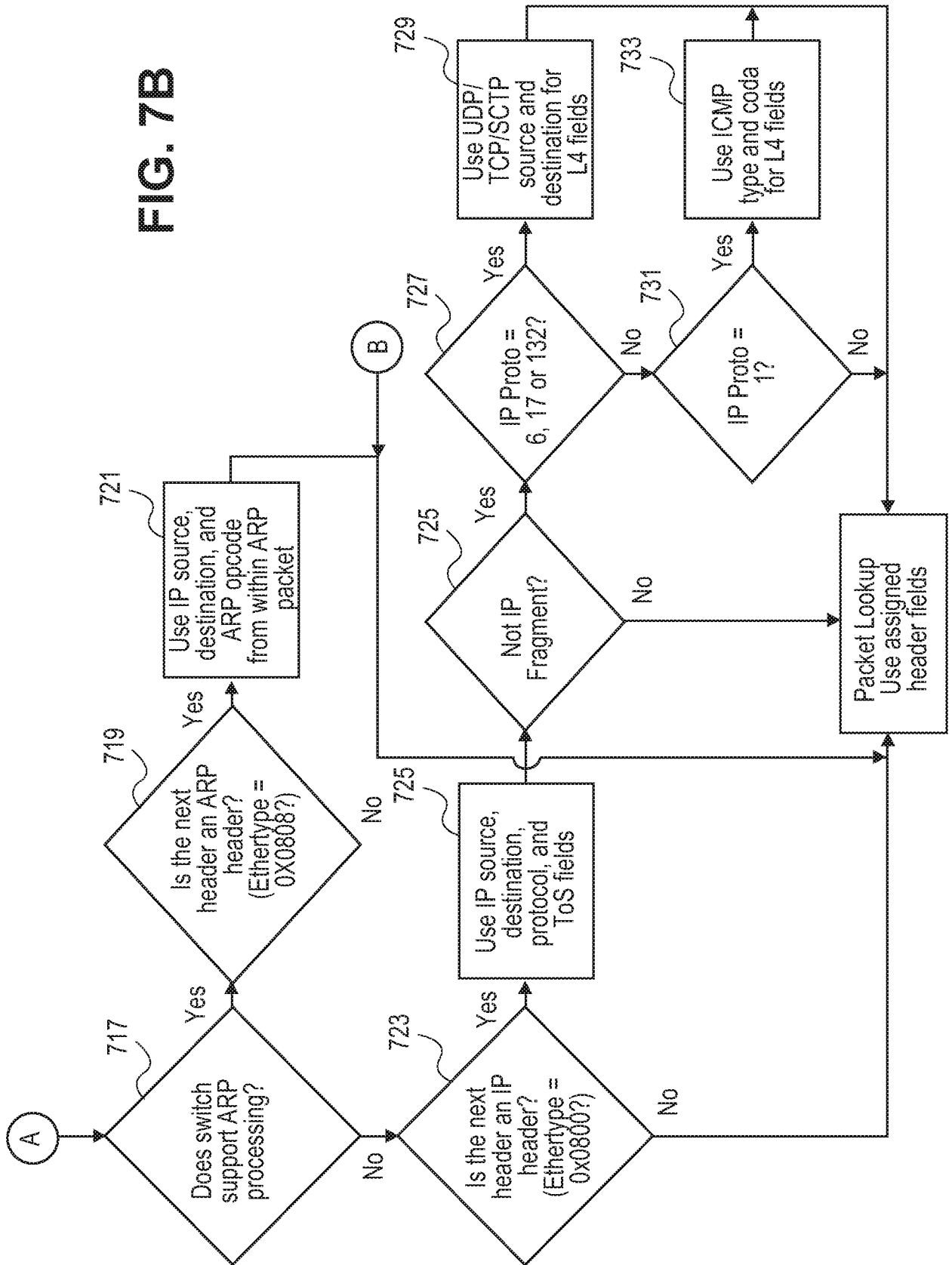


FIG. 7B



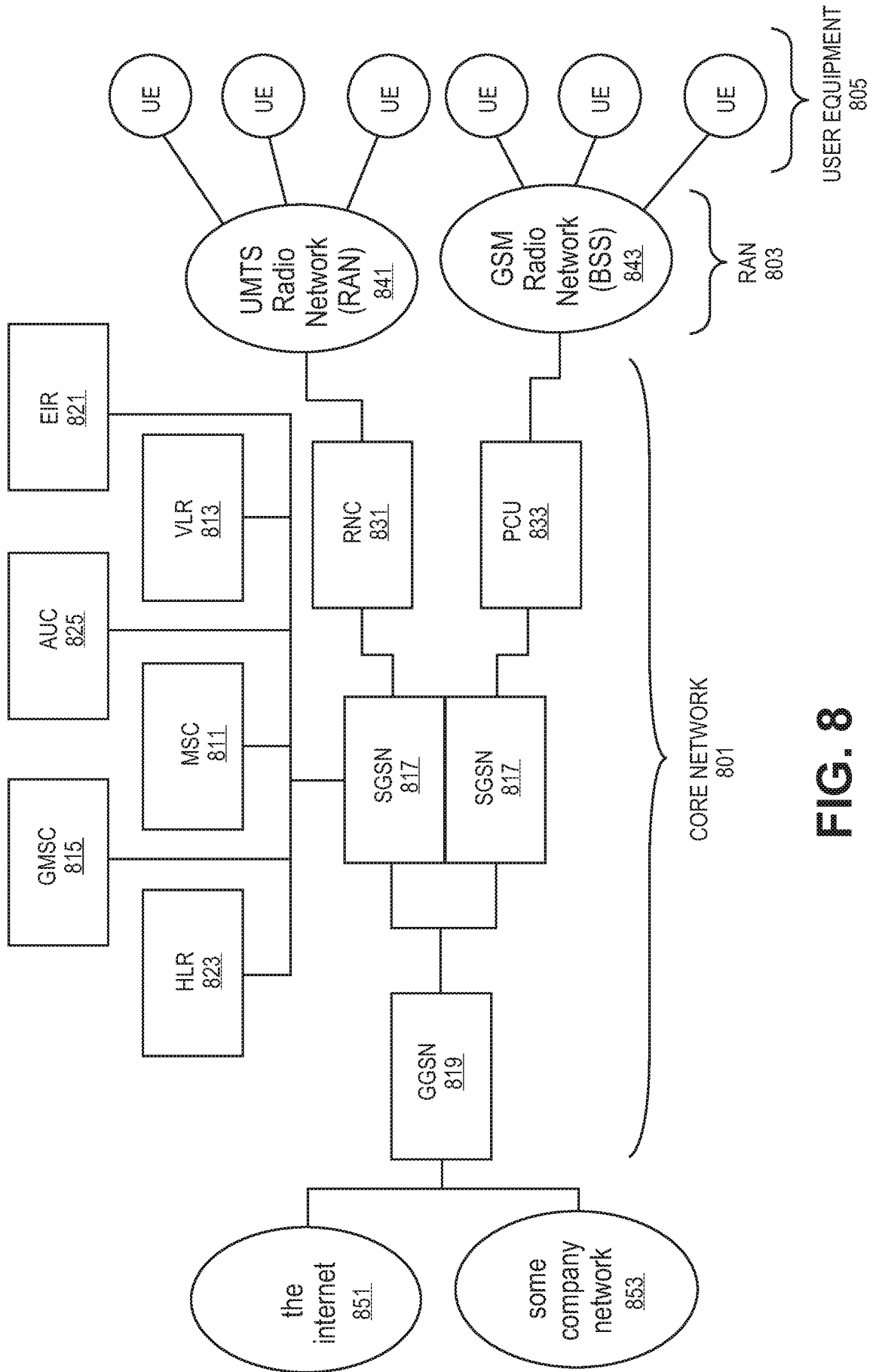
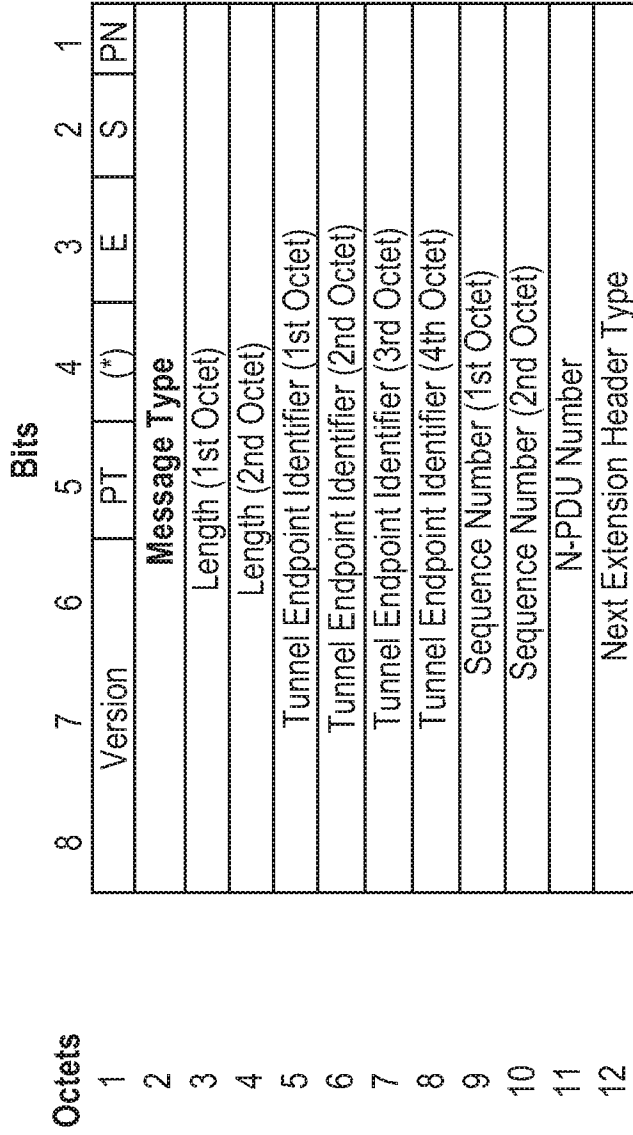


FIG. 8



- NOTE 0: (*) This bit is a spare bit. It shall be sent as '0'. The receiver shall not evaluate this bit,
- NOTE 1: 1) This field shall only be evaluated when indicated by the S flag set to 1.
- NOTE 2: 2) This field shall only be evaluated when indicated by the PN flag set to 1.
- NOTE 3: 3) This field shall only be evaluated when indicated by the E flag set to 1.
- NOTE 4: 4) This field shall be present if and only if any one or more of the S, PN and E flags are set.

FIG. 9

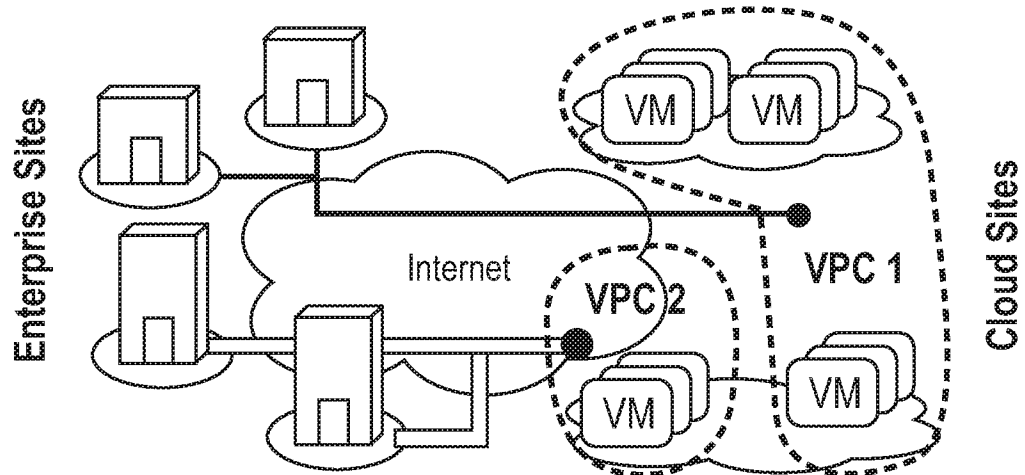


FIG. 10

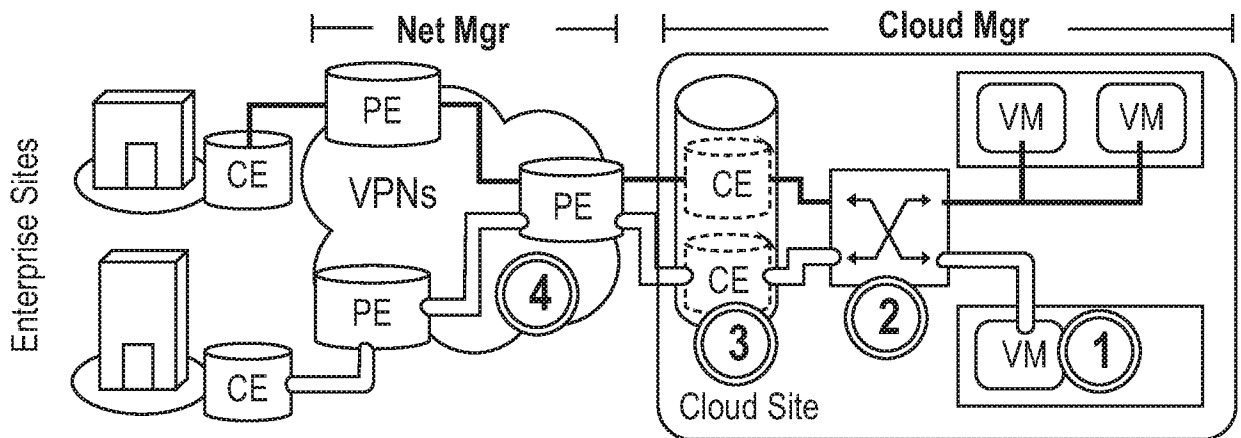


FIG. 11

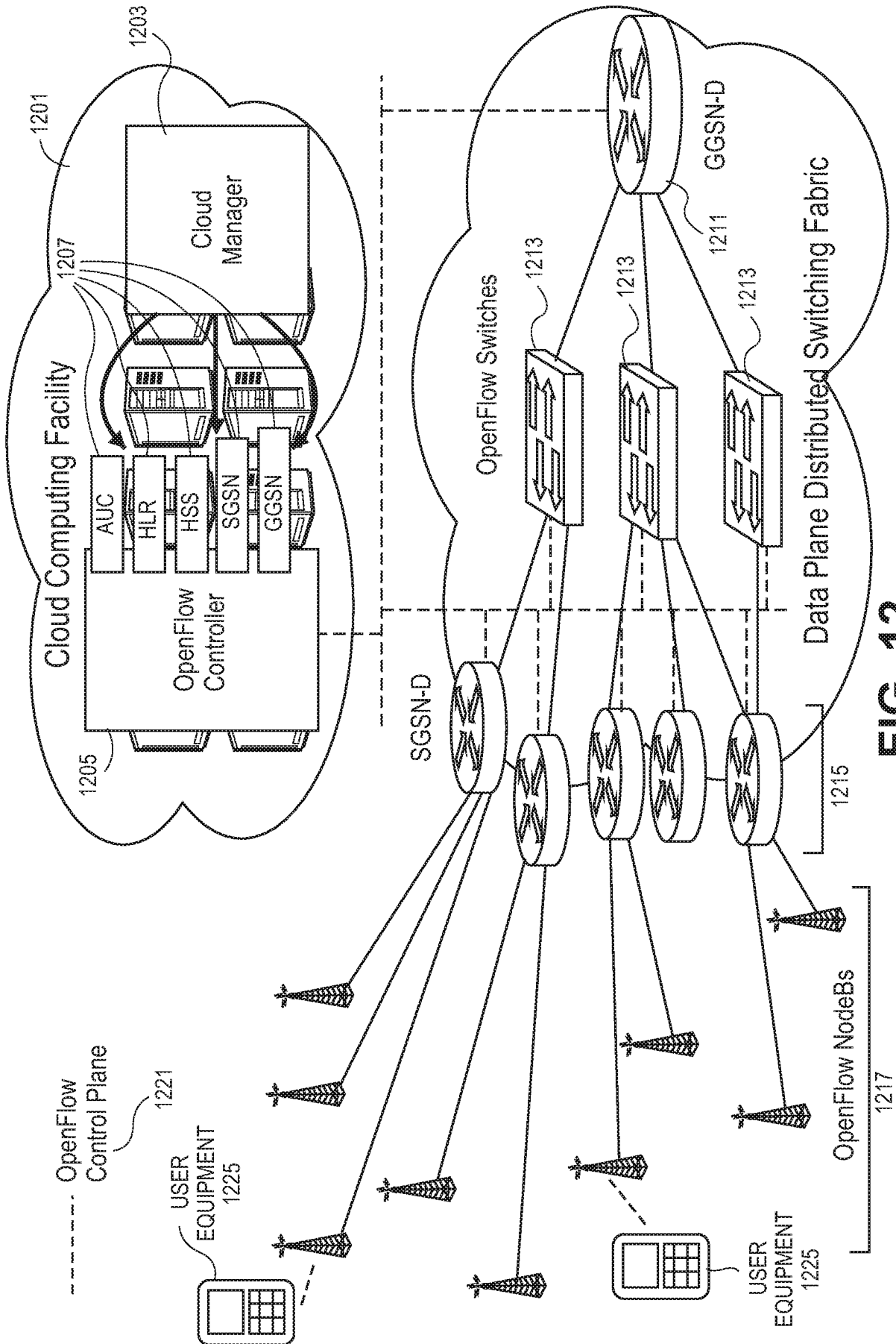


FIG. 12

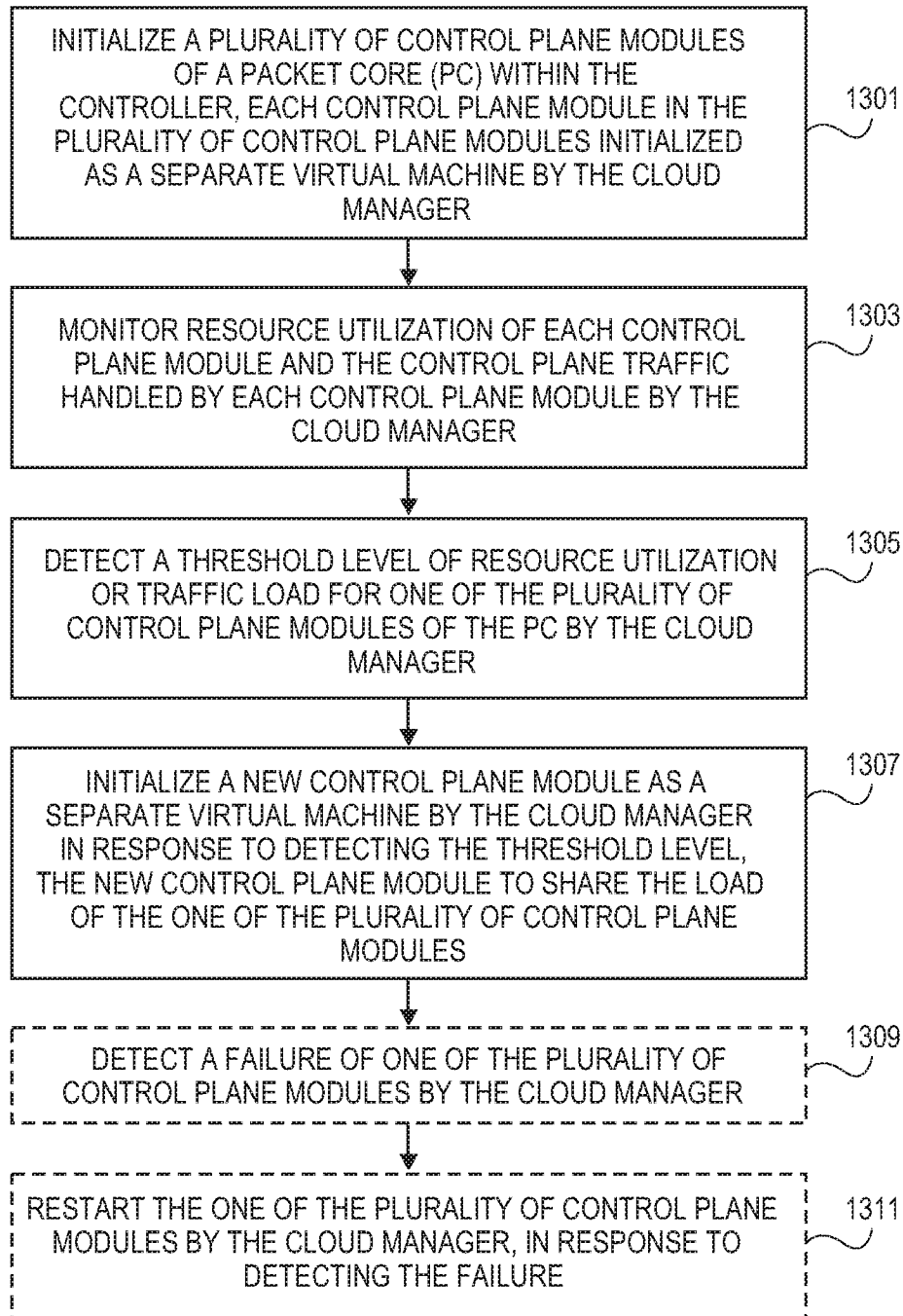


FIG. 13

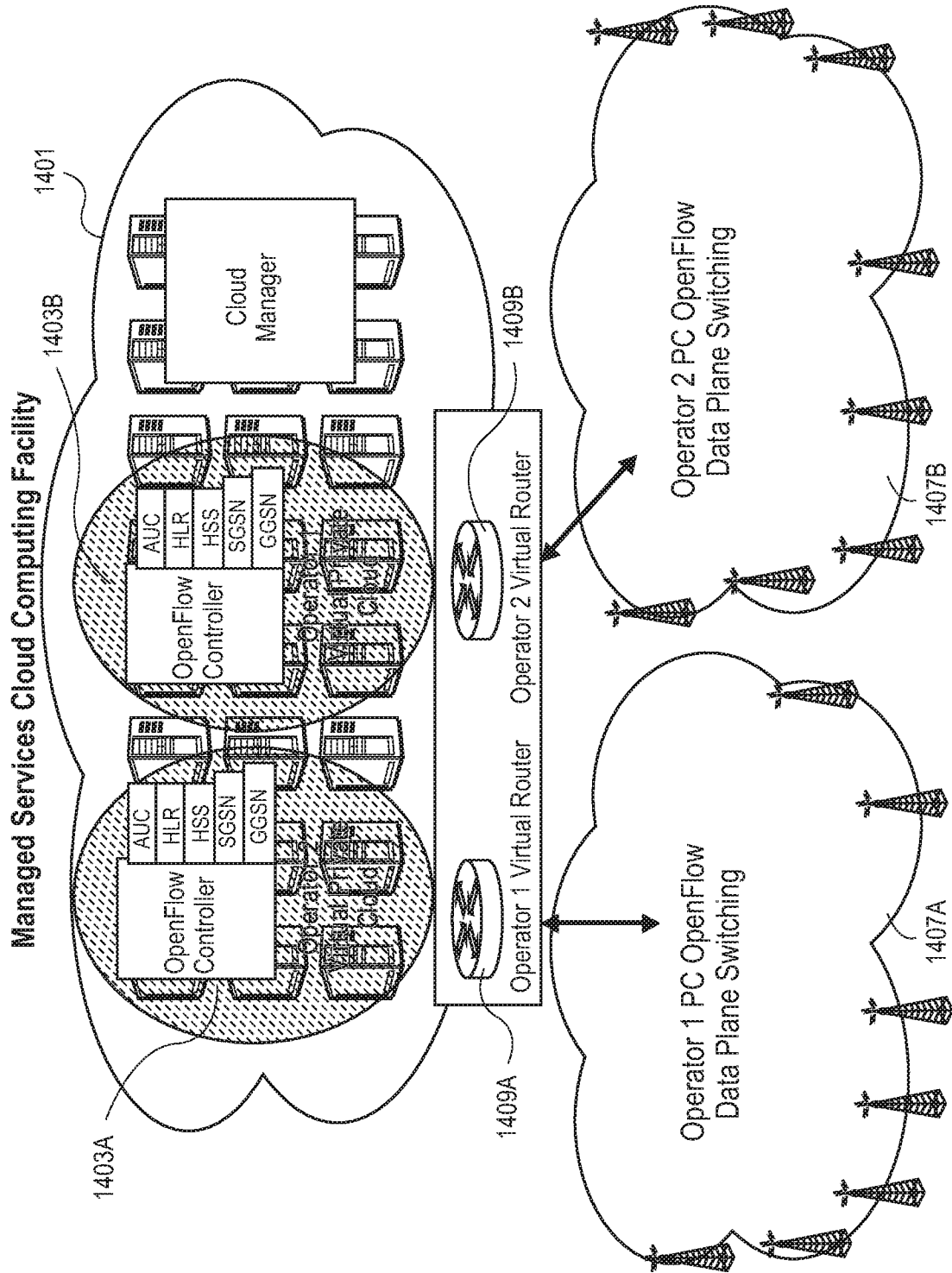


FIG. 14

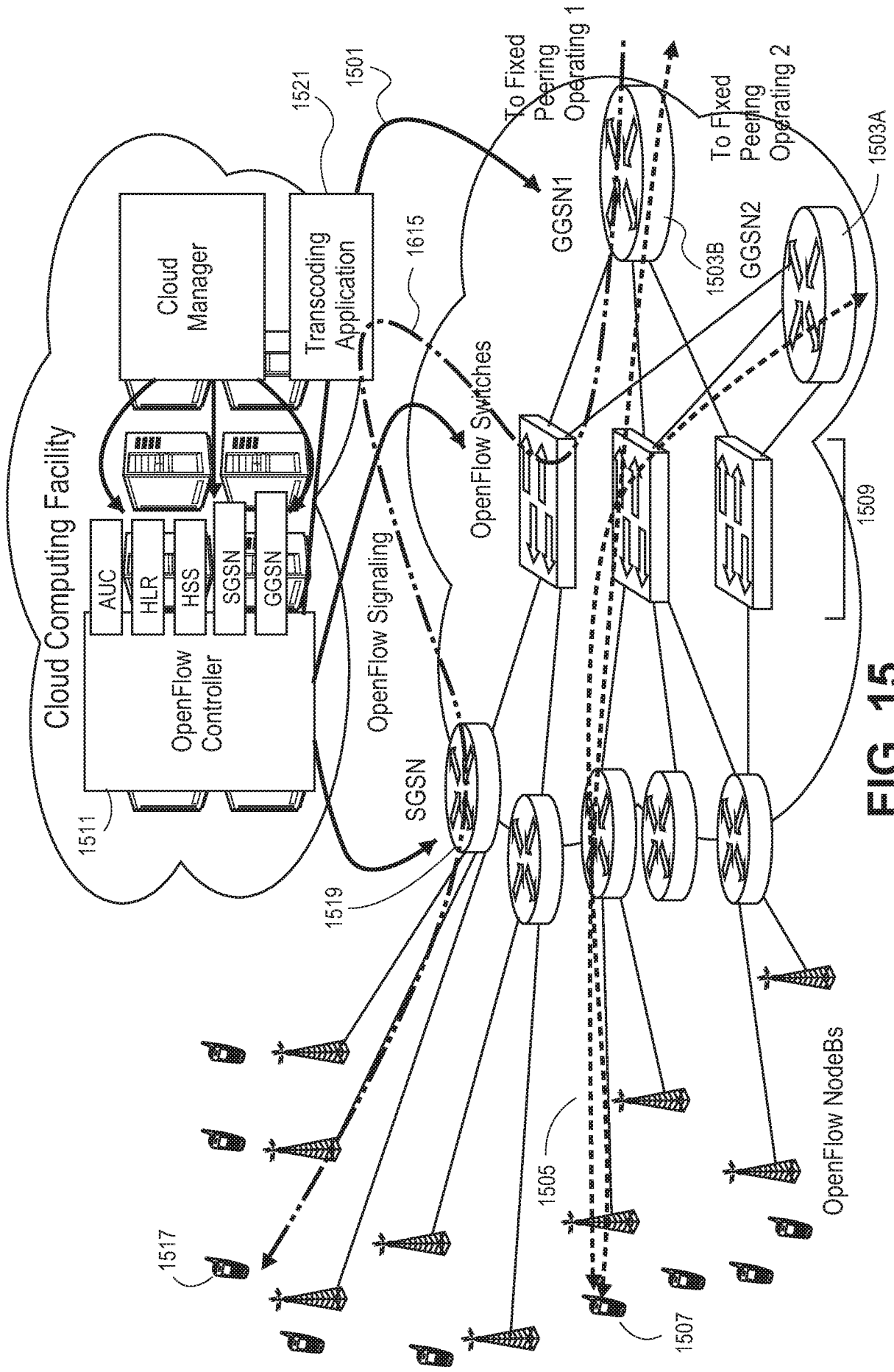


FIG. 15

Ingress Port
Metadata
Ether src
Ether dst
Ether type
VLAN id
VLAN priority
MPLS label
MPLS traffic class
IP src
IP dst
IP proto / ARP opcode
IP ToS bits
TCP / UDP / SCTP src port ICMP Type
TCP / UDP / SCTP dst port ICMP Code
2 byte GTP header flags field
4 byte GTP TEID

FIG. 16

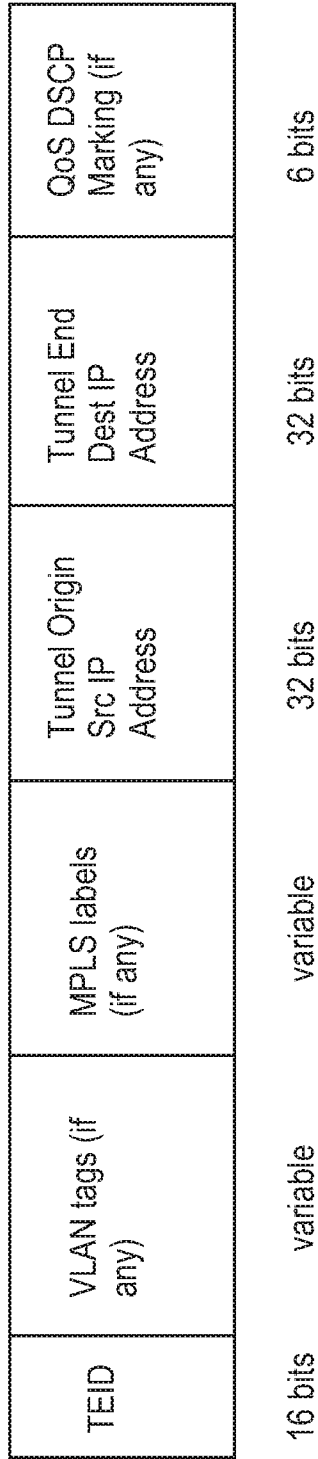


FIG. 17

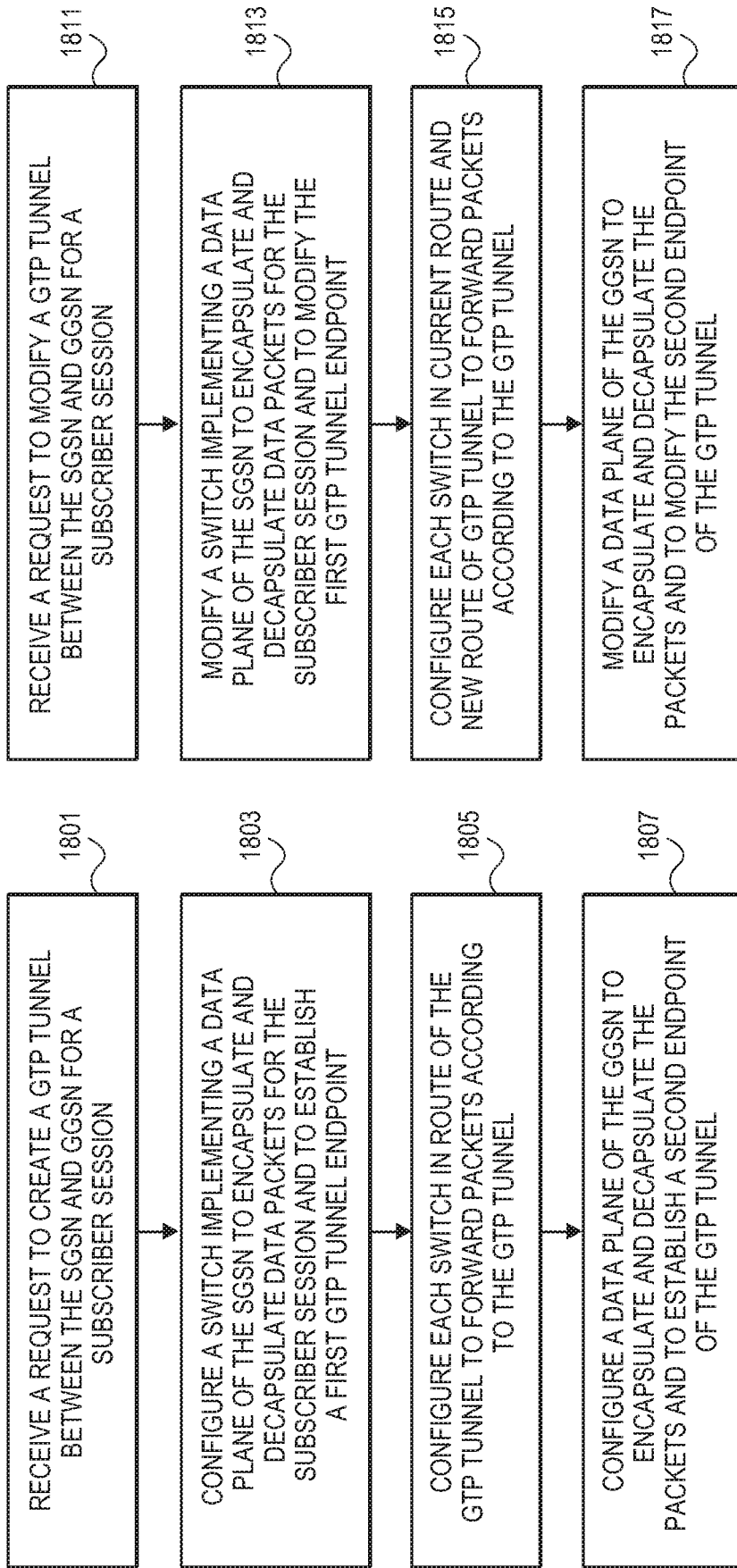
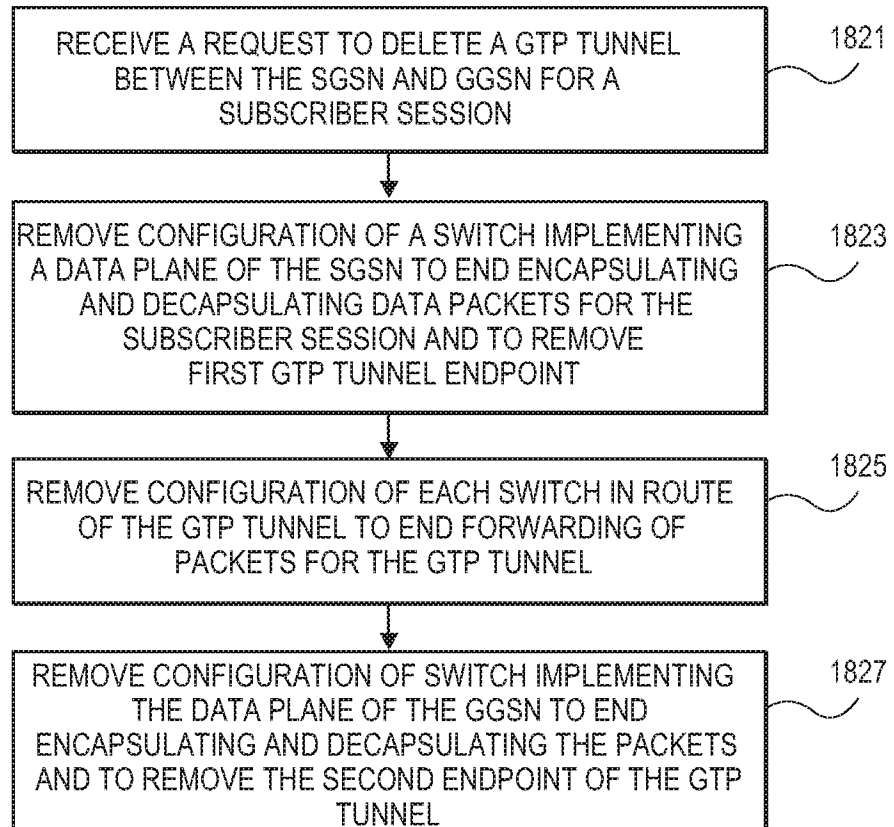


FIG. 18A

FIG. 18B

**FIG. 18C**

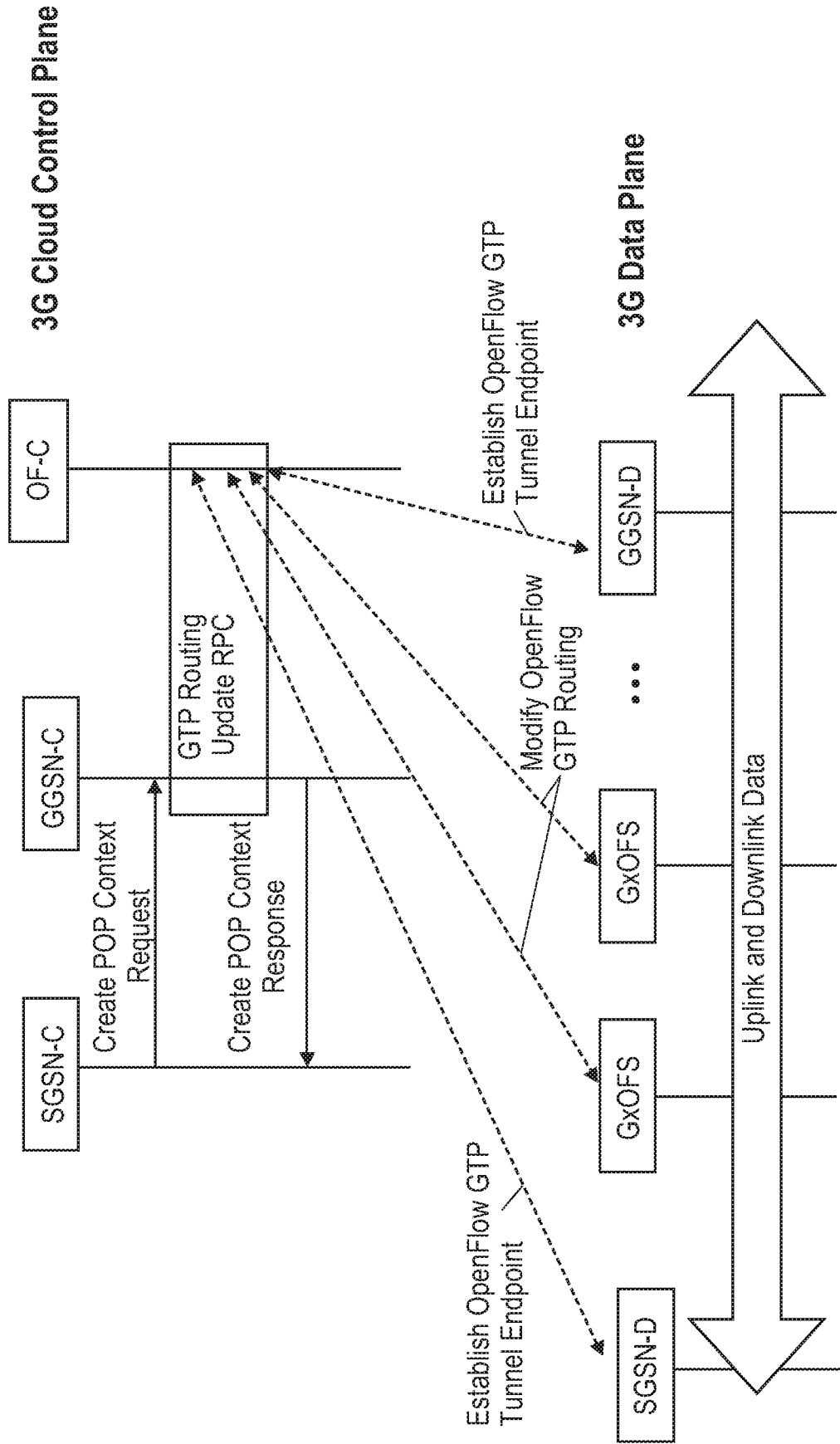


FIG. 19

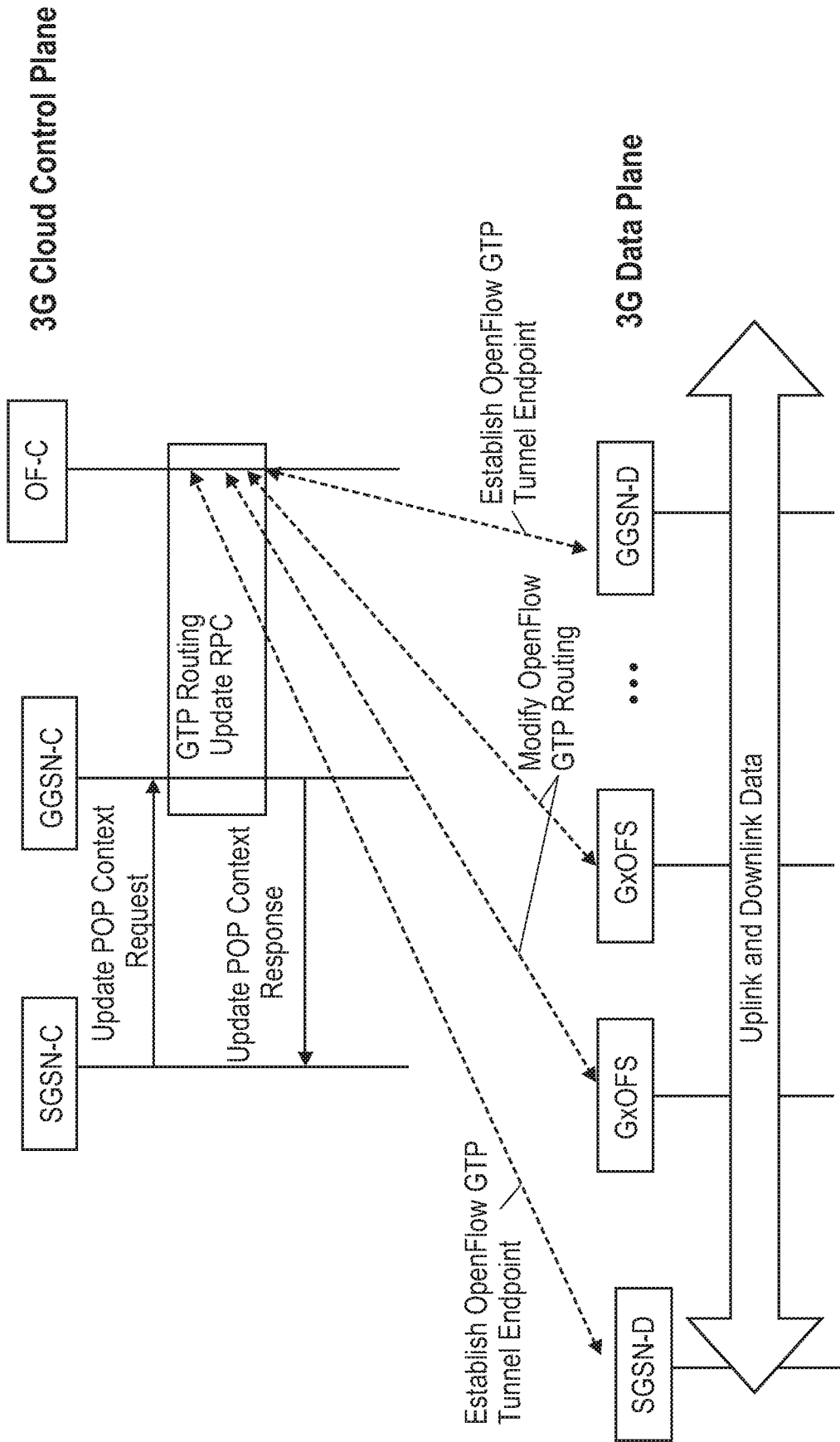


FIG. 20

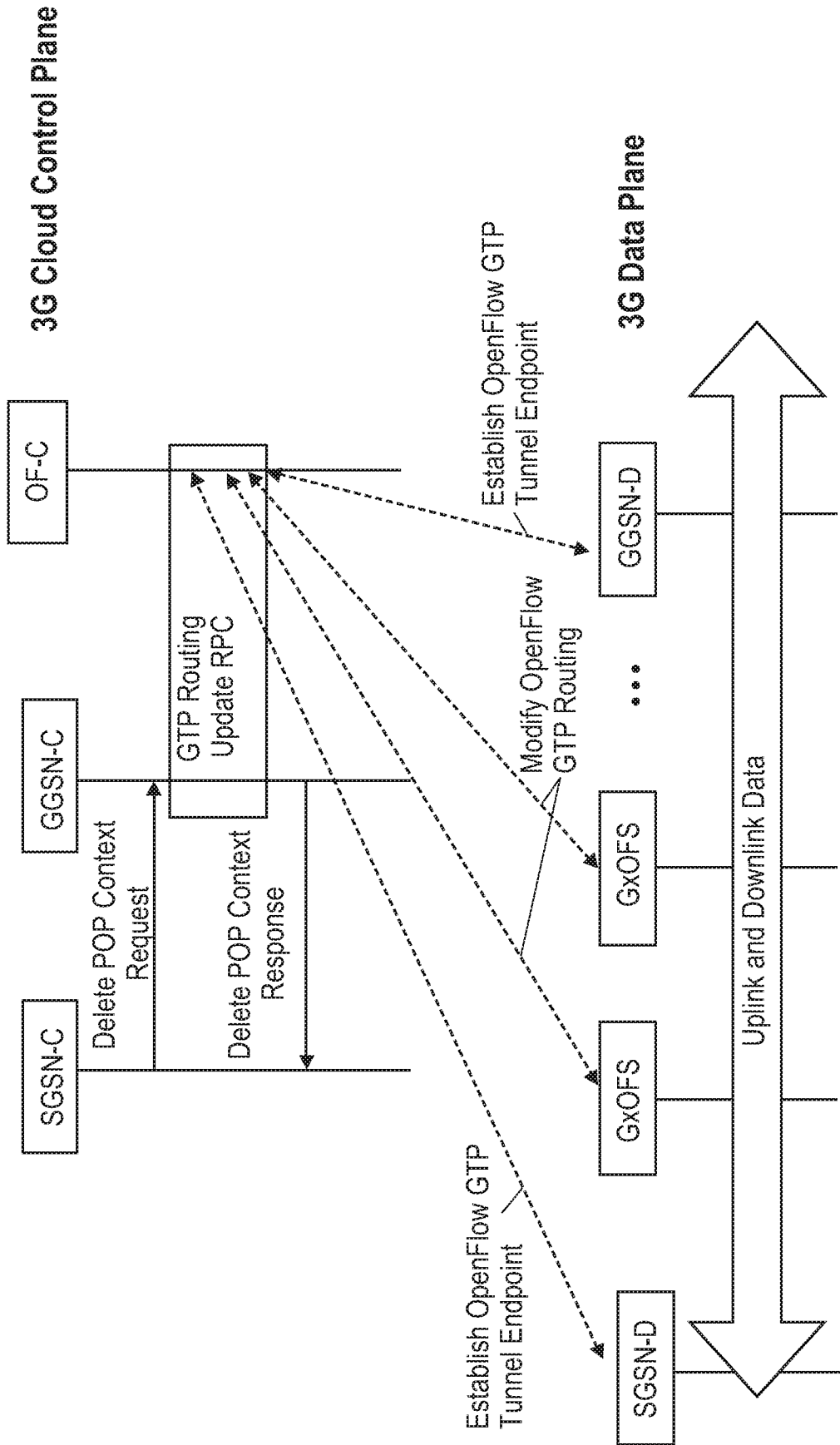


FIG. 21

FIG. 22A

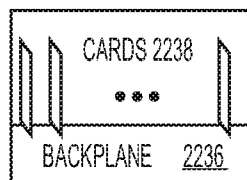
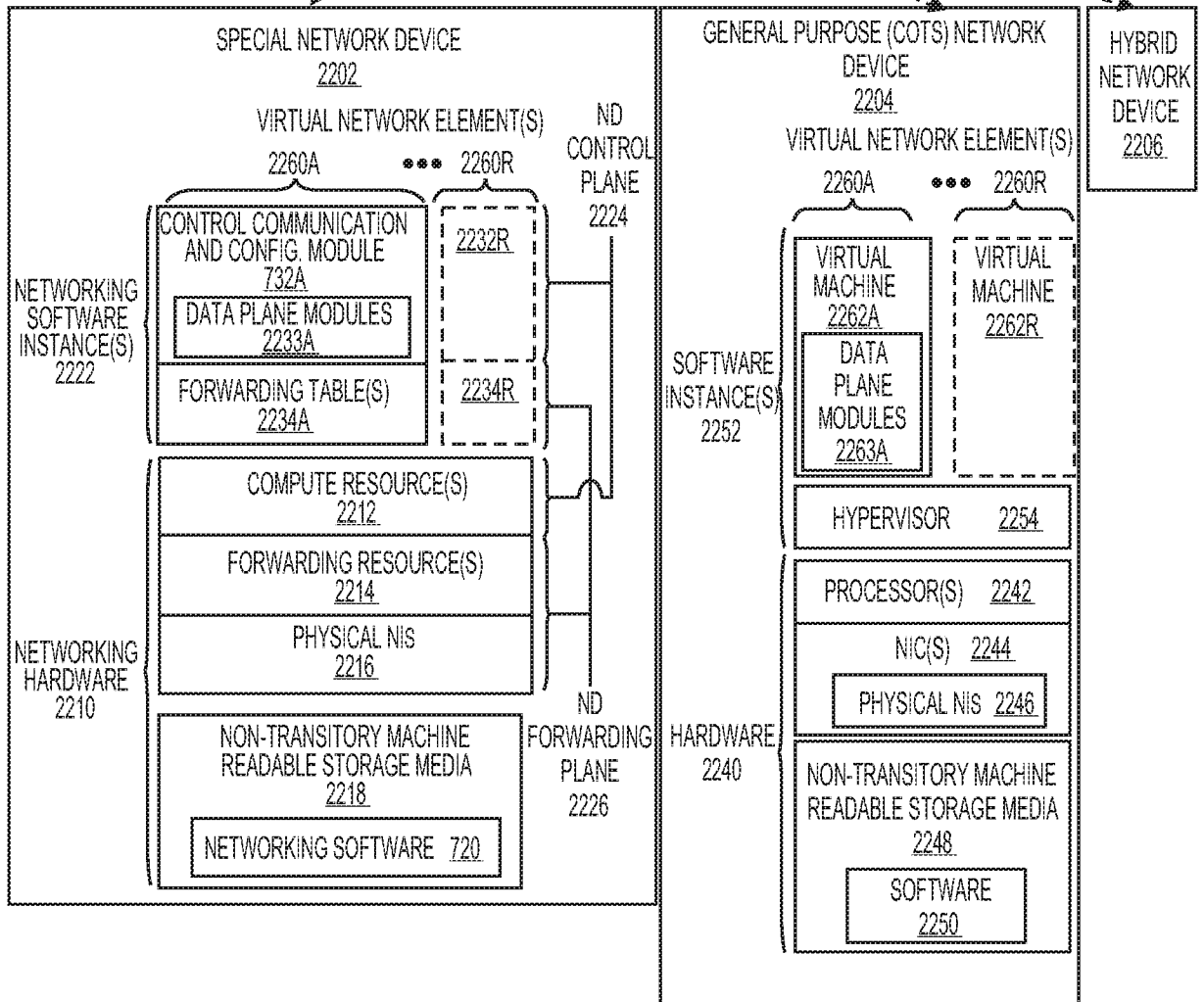
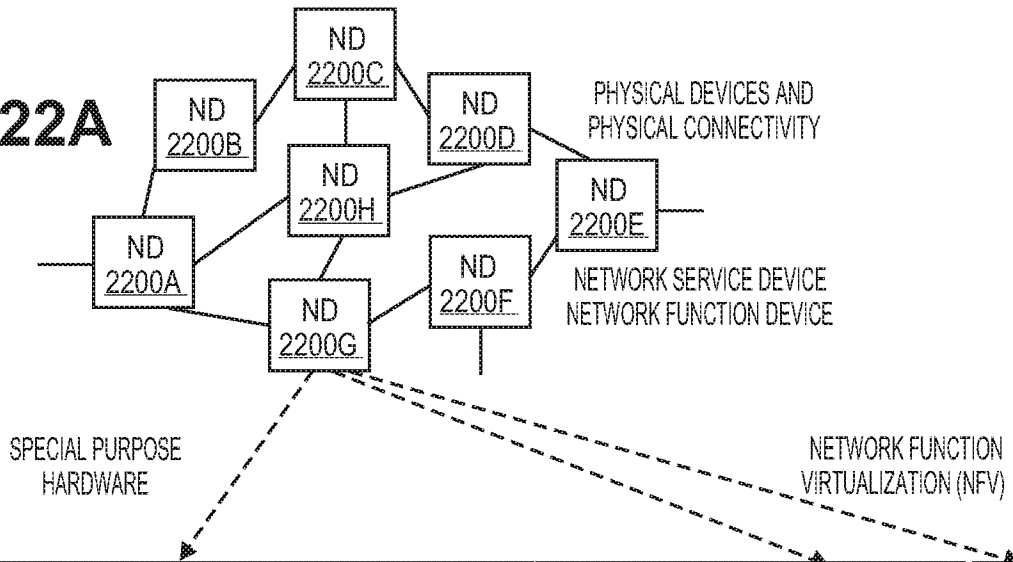


FIG. 22B

22/23

FIG. 22C

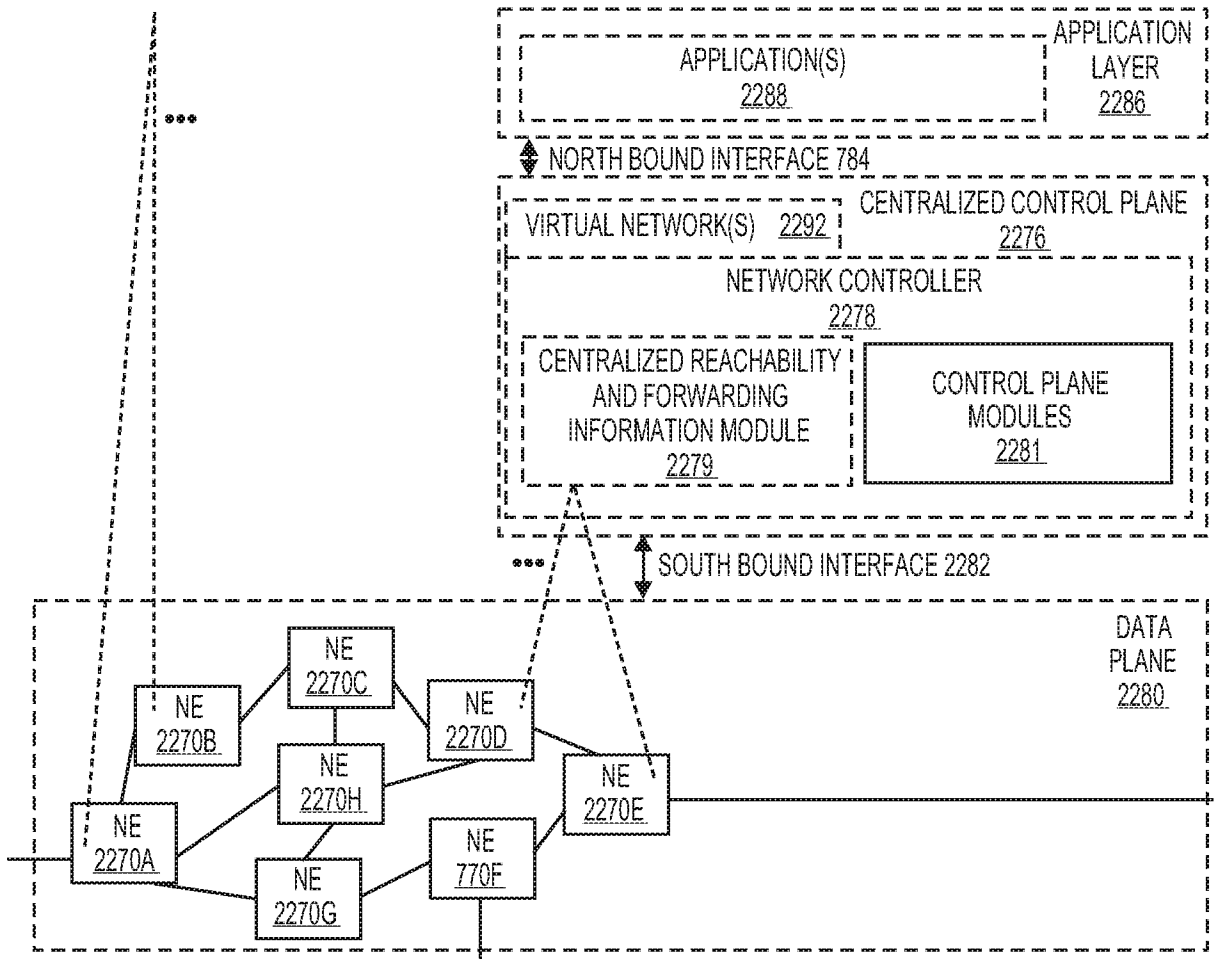
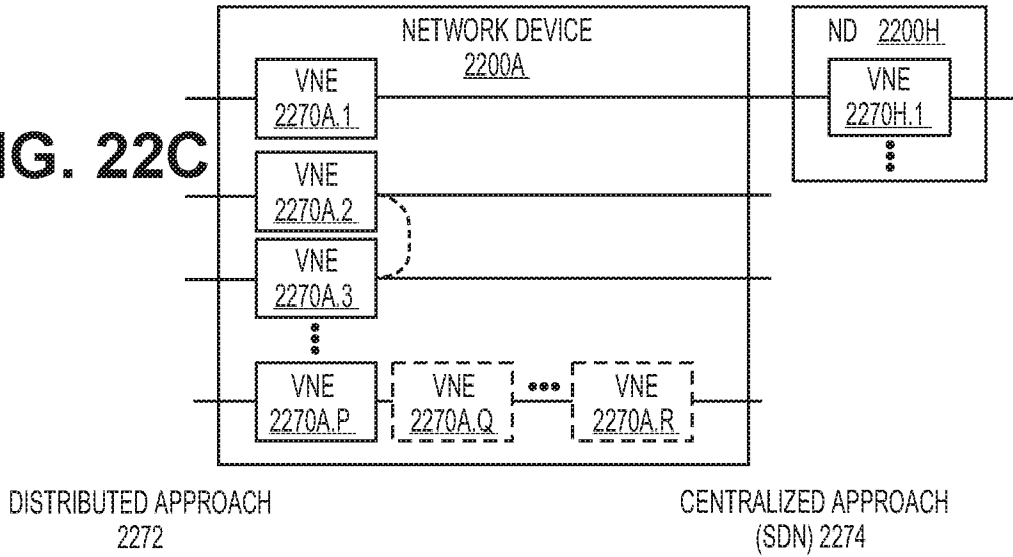


FIG. 22D

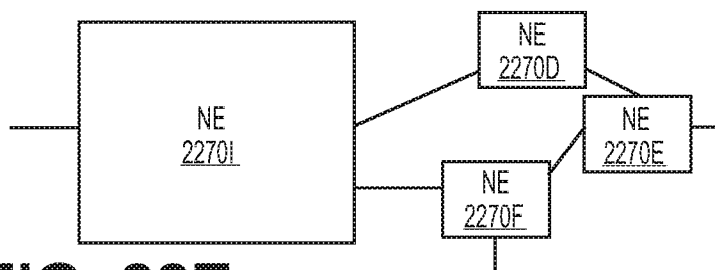


FIG. 22E

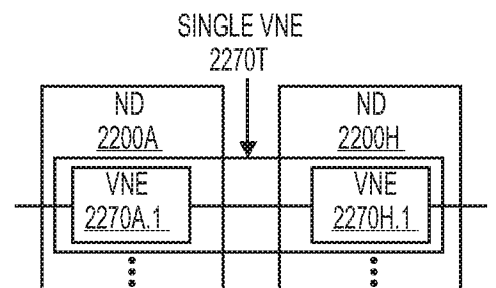
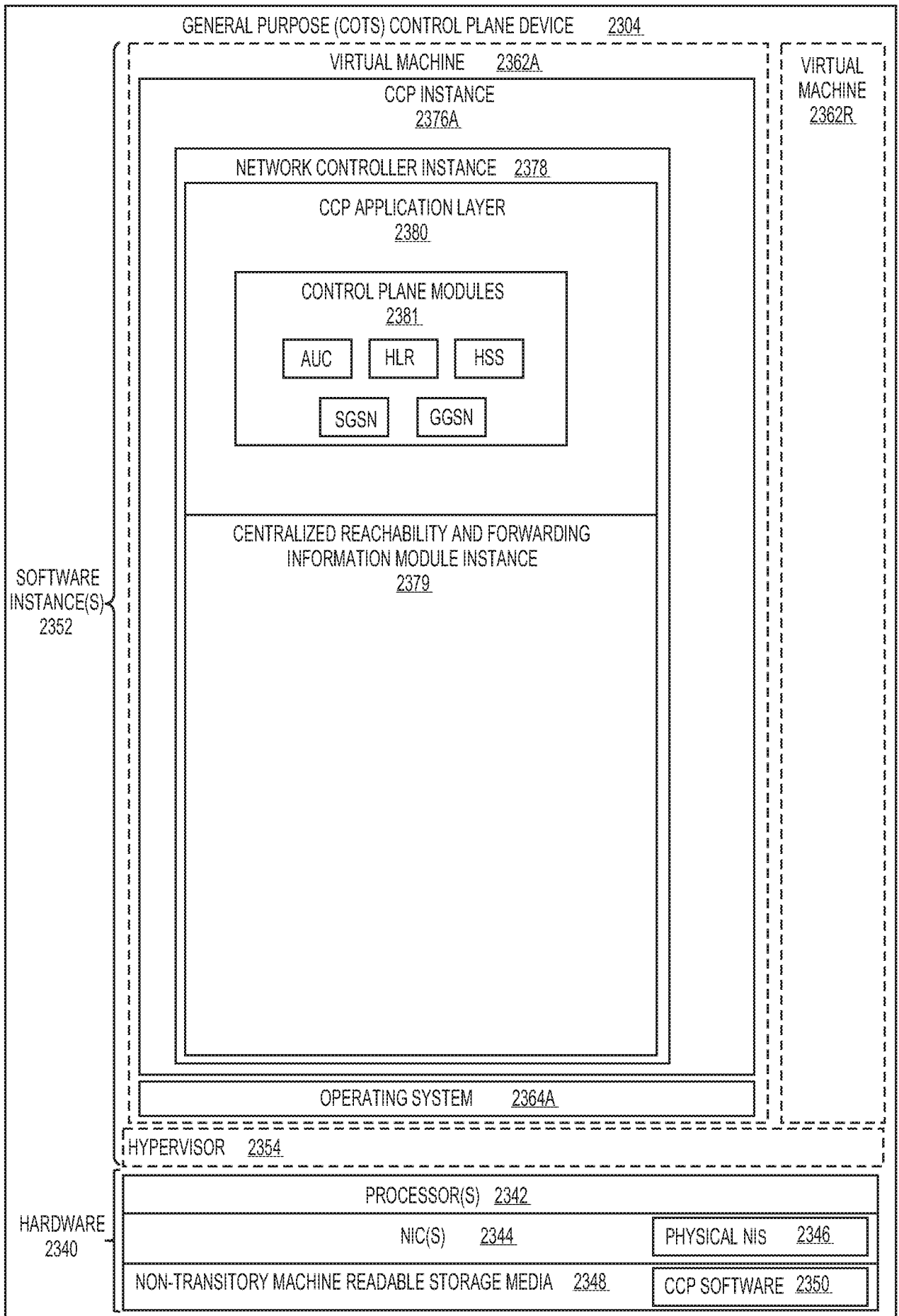


FIG. 22F

FIG. 23



INTERNATIONAL SEARCH REPORT

International application No PCT/IB2015/052447

A. CLASSIFICATION OF SUBJECT MATTER INV. H04L12/717 H04W76/02 H04L12/911 ADD. H04L12/721				
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 H04W				
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	US 2013/054761 A1 (KEMPf JAMES [US] ET AL) 28 February 2013 (2013-02-28) paragraphs [0005], [0006], [0031], [0038], [0040], [0042], [0043], [0064] - [0066], [0071] - [0074], [0133] - [0138], [0149] claims 1-8 figures 12, 14, 15	1-8		
X,P	US 2014/241247 A1 (KEMPf JAMES [US] ET AL) 28 August 2014 (2014-08-28) the whole document	1-8		
----- -/--				
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.				
* Special categories of cited documents : <table style="width: 100%; border: none;"> <tr> <td style="width: 50%; border: none; vertical-align: top;"> "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 </td> <td style="width: 50%; border: none; vertical-align: top;"> "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 </td> </tr> </table>			"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
"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			
7 July 2015	13/07/2015			
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Ramenzoni, Stefano			

INTERNATIONAL SEARCH REPORT

International application No PCT/IB2015/052447

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>KEMPF J ET AL: "Moving the Mobile Evolved Packet Core to the Cloud", 2012 IEEE 8TH INTERNATIONAL CONFERENCE ON WIRELESS AND MOBILE COMPUTING, NETWORKING AND COMMUNICATIONS (WIMOB 2012) : BARCELONA, SPAIN, 8 - 10 OCTOBER 2012, IEEE, PISCATAWAY, NJ, 8 October 2012 (2012-10-08), pages 784-791, XP002717394, ISBN: 978-1-4673-1430-5 the whole document</p>	1-8
A	<p>----- US 2014/078988 A1 (KANT NISHI [US] ET AL) 20 March 2014 (2014-03-20) paragraphs [0014] - [0020], [0022], [0023] -----</p>	1-8

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2015/052447

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2013054761 A1	28-02-2013	AU 2012303738 A1	20-03-2014
		CA 2847103 A1	07-03-2013
		CN 103931149 A	16-07-2014
		EP 2751964 A1	09-07-2014
		JP 2014531792 A	27-11-2014
		KR 20140054357 A	08-05-2014
		US 2013054761 A1	28-02-2013
		WO 2013030693 A1	07-03-2013

US 2014241247 A1	28-08-2014	NONE	

US 2014078988 A1	20-03-2014	NONE	
