

SYSTEMS AND METHODS FOR NETWORK SLICING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 62/645,484, filed March 20, 2018, titled “Systems and Methods for Non-Intrusive Network Slicing”, which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates generally to routing traffic in mobile networks, and in particular, to systems and methods of routing traffic using network slicing techniques.

BACKGROUND

[0003] Some wireless operators have deployed applications known as GTP Proxy functions. A GTP Proxy typically functions as an intermediary node between different mobile network operators (MNO) and is used to aid in interoperability. As an example, a GTP Proxy can be considered an entry point for an operator’s users that are roaming in other networks. A roaming network will send all the GTP related traffic, signaling and user plane, to the home GTP Proxy; from the roaming network the GTP Proxy is treated as the selected GGSN or PGW. The GTP Proxy terminates this roaming traffic and subsequently proxy the traffic to the real GGSN or PGW. In this mode of operation, the GTP Proxy hides the topology of the real GGSN/PGW(s), as well it can perform other functions of modifying and enforcement of the GTP related traffic. As the name describes the GTP Proxy always remains as an intermediary for the lifetime of a UE session; e.g. all the UE session traffic flows through the GTP Proxy.

SUMMARY

[0004] In some embodiments, a method includes receiving, at a network slicer in a networked system, a first message from a first network node of the networked system; determining, by the network slicer, a destination address from the first message; determining, by the network slicer, if a message type of the first message is a predetermined message type; if the determined message type of the first message is not the predetermined message type, routing the first message to a second network node of the networked system associated with the destination address; and if the determined message type of the first message is the

predetermined message type: determining a destination type from the first message, selecting, by the network slicer, a second destination address for the first message based on one or more of a policy of the network slicer or an external policy, the second destination being associated with a third network node of the networked system having a destination type that is the same as the determined destination type, creating, by the network slicer, a new message of the predetermined type including the second destination address, and routing the new message of the predetermined type to the third network node of the networked system associated with the second destination address.

[0005] In some embodiments, the predetermined message type is a creation request message.

[0006] In some embodiments, the first node is one or more of a mobility management entity (MME), a serving general packet radio services (GPRS) support node (SGSN), serving gateway (SGW), enhanced packet data gateway (ePDG), or a trusted WLAN.

[0007] In some embodiments, the determined destination type comprises a gateway general packet radio services (GPRS) support node (GGSN), a serving gateway (SGW), or a packet data network gateway (PGW).

[0008] In some embodiments, the selecting the second destination address for the first message based on the one or more of the policy of the network slicer or the external policy comprises: determining, by the network slicer, one or more of a throughput capability level of a user equipment (UE) associated with the first message, an enterprise associated with the UE associated with the first message, a traffic speed of the UE associated with the first message, a security level or requirement of the UE associated with the first message, or an employee status of the owner of the UE associated with the first message; and selecting, by the network slicer, the second destination address associated with one or more of the capability level of the UE, the enterprise of the UE, the traffic speed of the UE, the security level or requirement of the UE, or the employee status of the owner of the UE from a plurality of second destination addresses.

[0009] In some embodiments, the selecting the second destination address for the first message based on the one or more of the policy of the network slicer or the external policy comprises determining one or more of an IMSI of a user equipment associated with the first message, location of the user equipment associated with the first message, access point name

(APN) associated with the first message, an IMEI identification number of the user equipment associated with the first message, a Radio Access Technology type associated with the first message, or a timestamp of the first message.

[0010] In some embodiments, the method further includes receiving, by the network slicer, a response message from the third network node; and creating, by the network slicer, a new response message including: the destination address associated with the second network node as a source address of the new response message, and an indication of the second destination address associated with the third network node in the message content of the new response message.

[0011] In some embodiments, the method further includes storing, at the network slicer, one or more of an address associated with the first network node or the destination address prior to the receiving the response message from the third network node; and deleting, by the network slicer, the stored one or more of the address associated with the first network node or the destination address after the creating the new response message.

[0012] In some embodiments, the method further includes indicating to a next hop router by the network slicer that a router associated with the network slicer has a lower routing cost to the second network node than an expected routing cost from other routers in the networked system to the second network node prior to the receiving the first message, wherein the lower routing cost to the second network node is lower than an actual routing cost to the second network node.

[0013] In some embodiments, a method includes receiving, at a network slicer in a networked system, a creation request message from a first network node, the creation request message having as a first destination address an address associated with the network slicer; determining a destination type from the creation request message; selecting, by the network slicer, a second destination address for the creation request message based on one or more of a policy of the network slicer or an external policy, the destination being associated with a second network node of the networked system having a destination type that is the same as the determined destination type; creating, by the network slicer, a new creation request message including the second destination address and the message content from the creation request message; and routing the new creation request message to the second network node of the networked system associated with the second destination address.

[0014] In some embodiments, the method further includes receiving, by the network slicer, a response message from the second network node; and creating, by the network slicer, a new response message including the address associated with the second network node in the message content of the new response message and the address associated with the network slicer as the source address.

[0015] In some embodiments, the method further includes storing, at the network slicer, one or more of the address associated with the first network node or the second destination address prior to the receiving the response message from the second network node; and deleting, by the network slicer, the stored one or more of the address associated with the first network node or the second destination address after the creating the new response message.

[0016] In some embodiments, the first node is one or more of a mobility management entity (MME), a serving general packet radio services (GPRS) support node (SGSN), serving gateway (SGW), enhanced packet data gateway (ePDG), or a trusted WLAN.

[0017] In some embodiments, the determined destination type comprises a gateway general packet radio services (GPRS) support node (GGSN), a serving gateway (SGW), or a packet data network gateway (PGW).

[0018] In some embodiments, the selecting the second destination address for the creation request message based on the one or more of the policy of the network slicer or the external policy comprises: determining, by the network slicer, one or more of a throughput capability level of a user equipment (UE) associated with the first message, an enterprise associated with the UE associated with the first message, a traffic speed of the UE associated with the first message, a security level or requirement of the UE associated with the first message, or an employee status of the owner of the UE associated with the first message, and selecting, by the network slicer, the second destination address associated with one or more of the capability level of the UE, the enterprise of the UE, the traffic speed of the UE, the security level or requirement of the UE, or the employee status of the owner of the UE from a plurality of second destination addresses.

[0019] In some embodiments, the method further includes indicating to a domain name system of the networked system that the address associated with the network slicer has a higher priority than at least the second network node prior to the receiving the creation request message.

[0020] In some embodiments, a network slicer in a networked system includes: a processor; and a storage medium having instructions contained thereon that cause the processor to perform the following steps: receive, at the network slicer in the networked system, a first message from a first network node of the networked system; determine, by the network slicer, a destination address from the first message; determine, by the network slicer, if a message type of the first message is a predetermined message type; if the determined message type of the first message is not the predetermined message type, route the first message to a second network node of the networked system associated with the destination address; and if the determined message type of the first message is the predetermined message type: determine a destination type from the first message, select, by the network slicer, a second destination address for the first message based on one or more of a policy of the network slicer or an external policy, the second destination being associated with a third network node of the networked system having a destination type that is the same as the determined destination type, create, by the network slicer, a new message of the predetermined type including the second destination address, and route the new message of the predetermined type to the third network node of the networked system associated with the second destination address.

[0021] In some embodiments, the predetermined message type is a creation request message.

[0022] In some embodiments, the first node is one or more of a mobility management entity (MME), a serving general packet radio services (GPRS) support node (SGSN), serving gateway (SGW), enhanced packet data gateway (ePDG), or a trusted WLAN.

[0023] In some embodiments, the determined destination type comprises a gateway general packet radio services (GPRS) support node (GGSN), a serving gateway (SGW), or a packet data network gateway (PGW).

[0024] In some embodiments, the selecting the second destination address for the first message based on the one or more of the policy of the network slicer or the external policy comprises causing the processor to: determine, by the network slicer, one or more of a throughput capability level of a user equipment (UE) associated with the first message, an enterprise associated with the UE associated with the first message, a traffic speed of the UE associated with the first message, a security level or requirement of the UE associated with the first message, or an employee status of the owner of the UE associated with the first

message, and select, by the network slicer, the second destination address associated with one or more of the capability level of the UE, the enterprise of the UE, the traffic speed of the UE, the security level or requirement of the UE, or the employee status of the owner of the UE from a plurality of second destination addresses.

[0025] In some embodiments, the selecting the second destination address for the first message based on the one or more of the policy of the network slicer or the external policy comprises determining one or more of an IMSI of a user equipment associated with the first message, location of the user equipment associated with the first message, access point name (APN) associated with the first message, an IMEI identification number of the user equipment associated with the first message, a Radio Access Technology type associated with the first message, or a timestamp of the first message.

[0026] In some embodiments, the processor is further caused to: receive, by the network slicer, a response message from the third network node; and create, by the network slicer, a new response message including: the destination address associated with the second network node as a source address of the new response message, and an indication of the second destination address associated with the third network node in the message content of the new response message.

[0027] In some embodiments, the processor is further caused to: store, at the network slicer, one or more of an address associated with the first network node or the destination address prior to the receiving the response message from the third network node; and delete, by the network slicer, the stored one or more of the address associated with the first network node or the destination address after the creating the new response message.

[0028] In some embodiments, the processor is further caused to indicate to a next hop router by the network slicer that a router associated with the network slicer has a lower routing cost to the second network node than an expected routing cost from other routers in the networked system to the second network node prior to the receiving the first message, wherein the lower routing cost to the second network node is lower than an actual routing cost to the second network node.

[0029] In some embodiments, a network slicer in a networked system includes: a processor; and a storage medium having instructions contained thereon that cause the processor to perform the following steps: receive, at the network slicer in the networked system, a creation

request message from a first network node, the creation request message having as a first destination address an address associated with the network slicer; determine a destination type from the creation request message; select, by the network slicer, a second destination address for the creation request message based on one or more of a policy of the network slicer or an external policy, the destination being associated with a second network node of the networked system having a destination type that is the same as the determined destination type; create, by the network slicer, a new creation request message including the second destination address and the message content from the creation request message; and route the new creation request message to the second network node of the networked system associated with the second destination address.

[0030] In some embodiments, the processor is further caused to: receive, by the network slicer, a response message from the second network node; and create, by the network slicer, a new response message including the address associated with the second network node in the message content of the new response message and the address associated with the network slicer as the source address.

[0031] In some embodiments, the processor is further caused to: store, at the network slicer, one or more of the address associated with the first network node or the second destination address prior to the receiving the response message from the second network node; and delete, by the network slicer, the stored one or more of the address associated with the first network node or the second destination address after the creating the new response message.

[0032] In some embodiments, the first node is one or more of a mobility management entity (MME), a serving general packet radio services (GPRS) support node (SGSN), serving gateway (SGW), enhanced packet data gateway (ePDG), or a trusted WLAN.

[0033] In some embodiments, the determined destination type comprises a gateway general packet radio services (GPRS) support node (GGSN), a serving gateway (SGW), or a packet data network gateway (PGW).

[0034] In some embodiments, the selecting the second destination address for the creation request message based on the one or more of the policy of the network slicer or the external policy comprises causing the processor to: determine, by the network slicer, one or more of a throughput capability level of a user equipment (UE) associated with the first message, an

enterprise associated with the UE associated with the first message, a traffic speed of the UE associated with the first message, a security level or requirement of the UE associated with the first message, or an employee status of the owner of the UE associated with the first message, and select, by the network slicer, the second destination address associated with one or more of the capability level of the UE, the enterprise of the UE, the traffic speed of the UE, the security level or requirement of the UE, or the employee status of the owner of the UE from a plurality of second destination addresses.

[0035] In some embodiments, the processor is further caused to indicate to a domain name system of the networked system that the address associated with the network slicer has a higher priority than at least the second network node prior to the receiving the creation request message.

BRIEF DESCRIPTION OF FIGURES

[0036] Various objectives, features, and advantages of the disclosed subject matter can be more fully appreciated with reference to the following detailed description of the disclosed subject matter when considered in connection with the following drawings, in which like reference numerals identify like elements.

[0037] FIG. 1 is a system diagram showing a networked system.

[0038] FIGS. 2A and 2B are system diagrams showing networked systems, according to some embodiments of the present disclosure.

[0039] FIG. 3 is a diagram showing a network slicer, according to some embodiments of the present disclosure.

[0040] FIGS. 4A and 4B are flow charts showing message flows for GTP-C messages that are not Create Session Requests or Create PDP Context Requests, according to some embodiments of the present disclosure.

[0041] FIGS. 5A and 5B are flow charts showing message flows for GTP-C messages that are Create Session Requests or Create PDP Context Requests, according to some embodiments of the present disclosure.

[0042] FIGS. 6A and 6B are flow charts showing processes of routing uplink packets in a network using a non-intrusive network slicer or minimally intrusive network slicer, according to some embodiments of the present disclosure.

[0043] FIG. 7 is a flow chart showing a process of routing downlink packets in a network using a non-intrusive network slicer or minimally intrusive network slicer, according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0044] Systems and methods are disclosed to provide wireless network operators the ability to slice the core network into different slices such that control plane and user plane functionality can be segregated for different group of mobile users. In some embodiments, the systems and methods described herein are accomplished without impacting the configurations of any of existing network functions (e.g., the function of an MME, SGSN, DNS server, serving gateway (SGW), packet data network gateway (PGW), GGSN, TWAN, and ePDG, described below). In some embodiments, only updates to the DNS servers are required without impacting other network functions.

[0045] In some embodiments, network slicing techniques described herein include a flexible application that determines the correct Gateway (SGW, PGW, and/or GGSN, collectively referred to as “right side” network elements) slice to use for the UE/device PDN/PDP sessions. The application can derive, through inspection and/or receipt of the operator GTP-C signaling traffic from the MME, SGSN, SGW, TWAN, and/or ePDG (collectively referred to as “left side” network elements) to the right side elements and based on how the application is configured, the correct network slice to which to route future communications. According to some embodiments, such functionality can be achieved when the network slicer is inserted between at least one left side element and associated right side elements. In some embodiments where the network slicer is located between the SGW (left) and the PGW (right), the network slicer only selects a PGW for future communications. Further, the SGW learns of the originally selected PGW from the MME/SGSN.

[0046] In contrast to the GTP Proxies described above, embodiments of the systems and methods described herein take a network slicing approach to flexibly classify a UE, selecting the correct network slice to anchor the UE session, establishing the Gateway association, and

subsequently removing itself from the UE session traffic flow. This process can be accomplished in a non-intrusive or minimally intrusive manner such that existing mobile operator core network nodes are unaware of the presence of this new slicing application.

[0047] In some non-intrusive embodiments, a network slicer can be introduced without changes to the left side network elements or the DNS server. The network slicer can be inserted between one or more left side network elements and one or more right side network elements. The network slicer can inform a next hop router between one or more of the one or more left side network elements that it has a lower routing cost to one or more of the one or more right side network elements than other routers in the networked system that are capable of routing to the one or more of the one or more right side network elements. Accordingly, the network slicer can intercept messages between the one or more of the one or more left side network elements and the one or more of the one or more right side network elements without updating a DNS server or instructing the one or more left side elements that messages should be directed to the network slicer. As described in more detail below, the non-intrusive network slicer can then inspect the messages to identify creation request messages. When creation request messages are received, the non-intrusive network slicer can select a network slice that can, but is not necessarily, be different from the network slice selected by a left side element in the original creation request message. The non-intrusive network slicer may also receive non-creation response messages, which the non-intrusive network slicer can route in a manner consistent with routers that do not perform network slicing functions.

[0048] In some minimally intrusive embodiments, an indication can be made to a DNS server that an IP address associated with the minimally intrusive network slicer has a higher priority than one or more right side elements. Accordingly, when one or more left side elements queries the DNS server, the one or more left side elements receive an indication that the minimally intrusive network slicer has a higher priority than the right side elements to which creation messages would otherwise be routed. As described in more detail below, the one or more left side elements can then direct creation request messages to the IP address associated with the minimally intrusive network slicer. The minimally intrusive network slicer can then apply techniques describe herein to forward a creation network to a selected network slice (e.g., one or more right side elements), and return a response to the one or more left side elements that informs the one or more left side elements of which slice to direct future messages. Such future messages can be routed directly to the selected slice in a way

that bypasses the network slicer. Thus, given the explanations herein, the terms “non-intrusive” and “minimally intrusive” are merely used to identify the various embodiments and do not impose any further limitations on the scope of the invention.

[0049] In some embodiments, a network slicer is introduced as a layer 3 router on its interface facing the MME, SGSN, SGW, TWAN, or ePDG, and on the interface towards the other network functions, the network slicer acts as both as a router as well as an IP host. As described in more detail below, this allows for the introduction of the network slicer without disturbing other network functions in some non-intrusive embodiments. In some minimally intrusive embodiments, the network slicer can be made visible on the DNS server to act as an SGW, GGSN, and/or a PGW.

[0050] In some non-intrusive embodiments, when session management packets (e.g., GTP-C) from the MME, SGSN are sent to the SGWs or the GGSNs, or from SGW, TWAN or ePDG to the PGW, the network slicer informs a next hop router in front of the MME, SGSN, SGW, TWAN, and ePDG or the MME, SGSN, SGW, TWAN, and ePDG directly that it can reach the SGW, PGW, or GGSN at a lower routing cost relative to other routers in the network capable of reaching the SGW, PGW, or GGSN.

[0051] In some embodiments, the network slicer inspects incoming packets and determine if it should apply a slicing function on the incoming packets. For GTP-C packets that are not Create Session Requests (e.g., GTPv2) or Create PDP Context Request (e.g., GTPv1), the network slicer routes the packets to their destination as a router. For Create Session Requests or Create PDP Context Requests, the network slicer applies information (such as UE identities, Location, APN etc.) from the request along with locally configured policies and policies retrieved from external policy database to select which network slice should be used for the particular session. DNS servers used by the MME, SGSN, TWAN, and/or ePDG can also be sources of policy information.

[0052] In some minimally intrusive embodiments, the MME/SGSN can discover the network slicer in the same way the SGW, PGW or GGSN might be discovered, for example from a DNS server. The TWAN or ePDG can discover the network slicer in the same way the PGW might be discovered, for example from the DNS servers. The network slicer can receive the Create Session Request (e.g., via GTPv2) or Create PDP Context Request (e.g., GTPv1) without leveraging a method of altering routing cost. Once the create request packet is received, the rest of the processing can be performed in a similar way to non-intrusive embodiments.

[0053] In some embodiments, once a slice is selected, the network slicer creates a new Create Session Request or Create PDP Context Request using a local host IP address and sends the request to the selected slice, which can be a different slice from what the MME, SGSN, TWAN, and/or ePDG has selected. This allows the response to come back to the network slicer, which forms a Create Session Response or Create PDP Context Response to send back to the MME, SGSN, TWAN, and/or ePDG. The source IP address used in this case is set to match the destination IP address of the original request.

[0054] FIG. 1 is a system diagram showing a networked system. FIG. 1 shows user equipment (UE) 102, RAN Node 104, mobility management entity (MME) 106, serving general packet radio services (GPRS) support node (SGSN) 108, serving gateway (SGW) 112, packet data network gateway (PGW) 114, gateway GPRS support node (GGSN) 116, domain name system (DNS) 120, Gi network 130, untrusted wireless access local area network (WLAN) 140, enhanced packet data gateway (ePDG) 142, trusted WLAN 150, and trusted WLAN access network (TWAN) 152.

[0055] UE 102 connects to a networked system 100 through the RAN Node 104, which can be an evolved NodeB (eNodeB) in 4G network, or a Radio Network Controller (RNC) for 3G or a Base Station Subsystem (BSS) for 2G. UE 102 includes computing devices configured to connect to a mobile data network (e.g., mobile phones, tablets, laptops). RAN Node 104 is a radio part of a cell site. A single RAN Node 104 may contain several radio transmitters, receivers, control sections and power supplies. RAN Node 104 can be backhauled to MME 106 and SGSN 108. Backhaul is a process of transferring packets or communication signals over relatively long distances to a separate location for processing, for example at a data center maintained by a network operator. SGSN 108 routes and forwards user data packets while also acting as the mobility anchor for a user plane during inter-RAN Node handovers. MME 106 is a control node in the networked system 100. MME 106 handles mobility and session management for the UE 102 while it is in the 4G network. SGSN 108 and GGSN 116 are used to route traffic in a 2G or 3G system while MME 106, SGW 112, and PGW 114 are used to route traffic in a 4G/EPC system. UE 102 can obtain network access via two different types of WLAN. As shown in FIG. 1, WLAN 150 is trusted by the network operator, while WLAN 140 is not trusted by the network operator. WLANs 140, 150 link devices using wireless communications to form local area networks (LANs), for example, in limited areas. The WLANs 140, 150 can be integrated into the networked

system 100 via ePDG 142 and TWAN 152, respectively, for communications between the WLANs 140, 150 and PGW 114 over the S2b and S2a interfaces, respectively. For the trusted WLAN 150, UE 102 can obtain network access using TWAN 152 and PGW 114. For the untrusted WLAN 140, UE 102 can obtain network access using ePDG 142 and PGW 114.

[0056] When UE 102 attaches to the network, multiple control messages are exchanged between network elements in order to create a data session (e.g., 3G session, 4G session) and provide data connectivity to the UE 102. As explained above, RAN Node 104 can be backhauled to MME 106/SGSN 108. MME 106/SGSN 108 routes and forwards packets (e.g., user plane packets from SGSN 108 to GGSN 116) and/or sends other messages to SGW 112, PGW 114, and GGSN 116. In some network implementations, GGSN acts a serving node in a general packet radio service (GPRS) network, and PGW acts as a serving node in an evolved packet core (EPC) network. PGW 114 and GGSN 116 route packets to and from Gi Network 130. PGW 114 can also route packets to and from the Trusted WLAN 150 and Untrusted WLAN 140 via TWAN 152 and ePDG 142, respectively. Gi Network 130 can include any enterprise or origin servers. Gi Network 130 can connect to the internet or any other third-party server.

[0057] DNS 120 translates domain names into IP addresses and can be queried by any of the network nodes described above. DNS 120 can be implemented in servers connected to network nodes over an internal or external network connection. Each of the network nodes described above can be configured to communicate with DNS 120.

[0058] As discussed above, a GTP Proxy can sit between, for example, the SGW 112 and the PGW 114 and/or the SGSN 108 and the GGSN 116. However, as discussed above, GTP proxies can be intrusive.

[0059] FIG. 2A is a system diagram showing a networked system with a non-intrusive network slicer, according to some embodiments of the present disclosure. FIG. 2A shows mobile management entity 202, Serving GPRS Support Node (SGSN) 204, TWAN 205a, ePDG 205b, network slicer 206, SGW 207, router 208, router 210, host IP addresses 212, policy server/database 220, domain name system (DNS) server/database 120 slice 1 230, serving gateway (SGW) 232, packet data network gateway (PGW) 234, Gateway GPRS Support Node (GGSN) 236, slice 2 240, serving gateway (SGW) 242, packet data network gateway (PGW) 244, Gateway GPRS Support Node (GGSN) 246, slice 3 250, serving

gateway (SGW) 252, packet data network gateway (PGW) 254, and Gateway GPRS Support Node (GGSN) 256.

[0060] MME 202 functions similarly to or the same as MME 106 as shown in FIG. 1 and described in the accompanying text and SGSN 204 functions similarly to SGSN 108 as shown in FIG. 1 and described in the accompanying text. SGW 207, 232, 242, 252; PGW 234, 244, 254; GGSN 236, 246, 256 also function similarly to SGW 112, PGW 114, and GGSN 116 as shown in FIG. 1 and described in the accompanying text, and TWAN 205a and ePDG 205b function similarly to TWAN 152 and ePDG 142, respectively, as shown in FIG. 1 and described in the accompanying text. Note that although only singular MME 202, SGSN 204, SGW 207, TWAN 205a and ePDG 205b are shown in FIG. 2A that sit at the left side of the network slicer 206a, multiple of each of these elements can be included and can be connected to one or more SGWs 232, 242, 252, PGWs 234, 244, 254, and/or GGSNs 236, 246, 256. As shown in FIG. 2A, an SGW can be on both sides of the network slicer due to the interface it supports (S4/S11 on one side and S5/S8 on the other side). A network slicer 206a can be introduced into the network to route traffic to and from “access-network facing nodes to data-network network facing nodes” (e.g., from left side elements to right side elements).

[0061] Network slicer 206a can include router 208 for transmitting and receiving packets between the network slicer 206a and MME 202, SGSN 204, SGW 207, TWAN 205a, ePDG 205b, and router 210 for transmitting and receiving packets between the network slicer 206a and SGWs 232/242/252, PGWs 234/244-254, and GGSNs 236/246/256. For network slicer 206a, the router 208 may inform the next hop router or MME, SGSN, SGW, TWAN, and ePDG directly that it has the lowest routing costs as discussed above. Thus, all traffic, regardless of type, may be routed to network slicer 206a regardless of type. As shown in FIG. 2A, network slicer can consult one of policy 220 and DNS 120 for a policy associated with a received packet. In particular, network slicer 206a can inspect Create Session Requests or Create PDP Context Requests and their corresponding responses to determine if modification is required. Network slicer 206a routes other GTP-C packets through to the network slices without any inspection or processing. Packets from router 208 are sent to router 210 and vice versa. Policy 220 can be local to the network slicer 206a, an external database, or a combination of the two. Policy 220, which is explained in more detail

throughout the present disclosure, can include various rules applied by the network slicer 206a.

[0062] For example, and as described in more detail below, for an incoming Create Session Request (or Create PDP Context Request), the request can include the MME, SGSN, SGW, TWAN, or ePDG's IP as the source IP address and the IP address of the SGW, PGW or GGSN selected by the MME, SGSN, TWAN, or ePDG as the destination IP address. In some embodiments, the network slicer 206a may determine that a different SGW, PGW, or GGSN should be used based on policy 220. Examples of such scenarios include, but are not limited to when an operator has a dedicated slice for a particular enterprise customer (when a UE is detected to belong to the enterprise, the network slicer directs the request to the designated slice for the customer), when one slice or more slices are used for high security users (the network slicer can be configured to detect such users, for example, based on identifying characteristics, and direct request therefrom to the one or more secure slices), and when a new slice is considered to be experimental for a particular function such as high throughput (the operator may want to test that slice with particular UEs, such as internal employee devices, and the network slicer can direct employee devices to that slice using the slicer). An exemplary gateway-selection policy can list a specific right side element or a list of right side elements associated with a predetermined group of users. In another exemplary gateway-selection policy, the policy can list one or more characteristics of a right side element to be selected (e.g., support for high speed data) for particular users, session types, etc. An exemplary admission-priority policy can prioritize users or groups of users when the network slicer is overloaded. For example, Create Session Requests from a particular group or groups of lower priority users may be ignored if the network slicer is overloaded (or above a threshold usage) such that the network slicer does not select new network slices for the lower priority users. When the network slicer 206a determines a different SGW, PGW or GGSN should be used, the network slicer 206a replaces the source IP using its own host IP 212 and the destination IP using that of the SGW or GGSN selected by the Network Slicer based on its local policy. Since the source IP is that of the network slicer 206a, the response from the SGW, PGW or GGSN will come back to the network slicer 206a. At that time, it replaces the destination IP using that of the MME, SGSN, SGW, TWAN, or ePDG from which the original request was received, and the source IP using the IP Address of the SGW or GGSN selected by the MME, SGSN, TWAN, or ePDG such that the source IP address of the response is the destination IP address of the original request.

[0063] Slices 230, 240, 250 refer to network slices, each of which is a partitioning of a network infrastructure into multiple network infrastructures. For example, slices 230, 240, 250 can all be running on the same hardware platform, but implemented as separate logical networks in software. The slices can also be implemented in different hardware platform or different virtualization infra-structures.

[0064] FIG. 2B is a system diagram showing a networked system with a minimally-intrusive network slicer, according to some embodiments of the present disclosure. Elements in FIG. 2B with similar numbers to FIG. 2A function in similar ways as described herein. In some embodiments, FIG. 2B differs from FIG. 2A in that a minimally-intrusive network slicer 206b includes a host IP address 214. As shown in FIG. 2B, host IP address 214 of the minimally-intrusive network slicer 206b is visible to MME 202, SGSN 204, SGW 207, TWAN 205a, and ePDG 205b. Unlike the embodiment shown in FIG. 2A, MME 202, SGSN 204, SGW 207, TWAN 205a, and ePDG 205b may be modified to transmit creation session request messages directly to the host IP address 214 of the minimally-intrusive network slicer 206b. As discussed above with reference to non-intrusive network slicer 206a, minimally-intrusive network slicer 206b can determine the appropriate anchor node from the available network slices (e.g., by applying various policies 220 described throughout the present disclosure), and forward the message to the appropriate node by replacing its own IP 214 in the message with the destination IP of the appropriate network element in the network slices 230, 240, 250. The network slicer 206b can be referred to as minimally intrusive because the MMEs, SGSNs, SGW, TWANs, and ePDGs first send a create session request directed to the host IP address 214 of the network slicer 206b, whereas the initial create session requests discussed with reference to FIG. 2A are directed at network element(s) in the network slices 230, 240, or 250 but routed to network slicer because the network slicer informs the next hop routers or MME, SGSN, SGW, TWAN, and ePDG that it has lower routing costs relative to other routers to the right side elements. However, the minimally-intrusive network slicer 206b may nonetheless be considered stateless because session state information is not maintained on the network slicer, like in non-intrusive embodiments. This is because both the non-intrusive network slicer 206a and the minimally intrusive network slicer 206b do not need to maintain any stored information about received or forwarded messages after sending a return message to the original left side element. Thus, in some embodiments, only the left and/or right side network elements maintain session state information, while the network slicer does not. However, in other embodiments, the network slicer may record, keep, or

maintain session state information, or may otherwise function as a “stateful” network element. In some embodiments, for non-creation requests, the network slicer 206b does not receive any traffic because MMEs, SGSNs, SGWs, TWANs, and ePDGs route directly to the selected slice that is indicated to the left side element in a return message from the network slicer 206b, rather than directly to the host IP address 214 of the network slicer 206b and because the network slicer 206b may be implemented in a manner that does not inform next hop routers that the network slicer 206b has lower routing costs to the right side elements.

[0065] FIG. 3 is a diagram showing a minimally or non-intrusive network slicer, according to some embodiments of the present disclosure. FIG. 3 shows a Management module 301, a Policy Server Interface module 302, a Policy Application module 303, a Routing Control module 304 and a Packet Forwarding Module 305.

[0066] Management module 301 provides configuration management, operation management, performance management (e.g., tracking a number of requests received, number or responses received, number of requests that have been redirected, number of responses that have been redirected, etc.) and fault management function (e.g., monitoring the health of the nodes belong to various slices, monitoring the connectivity to the policy database, generating alarms when peer nodes in the slices are down or the DB is not accessible, etc.) for the network slicer 206a or 206b. Management module 301 allows a user to control and monitor the network slicer 206a or 206b, including the various modules within the network slicer 206a or 206b.

[0067] Policy application module 303 performs the application of the network slicing policy to determine the right SGW, PGW and GGSN to use for a Create Session Request or Create PDP Context Request being processed. In some embodiments, policy application module 303 uses locally configured policy, which can be stored in the policy application module 303 or a storage in the network slicer 206a or 206b, as well as policies received from external servers via the Policy Server Interface module 302 to external servers, such as a subscriber policy repository (SPR), which can store subscribers’ privileges and grouping information such as enterprise customer status, security level, etc.

[0068] Packet Forwarding module 305 inspects incoming packets. For Create Session Request or Create PDP Context Request messages and their corresponding response messages, the Packet Forwarding module sends them to the Policy application module 303.

In some non-invasive embodiments, for other packets, the Packet Forwarding module 305 simply forwards them to the appropriate network interface based on forwarding rules determined by the Routing Control module 304, such as IP 10.10.2.4 mask 255.255.0.0 interface port1 or IP 10.20.3.1 mask 255.255.0.0 interface port2.

[0069] FIGS. 4A and 4B are flow charts showing message flows for GTP-C messages that are not Create Session Request or Create PDP Context Request, according to some embodiments of the present disclosure. The message flow in FIG. 4A can be associated with the non-intrusive network slicer 206a shown in FIG. 2A, according to some embodiments.

[0070] Referring to step 402, MME 202 routes a GTP-C that is not a Create Session Request to the network slicer 206a. The MME 202 may route the request to network slicer 206a because network slicer 206a informs the next hop router or MME, SGSN, SGW, TWAN, and ePDG directly that router 208 has the cheapest routing costs. Thus, all packets regardless of type may be sent to the network slicer 206a. The received request has a source IP address associated with the MME 202 and a destination IP address of an SGW 232 in slice 1 230.

[0071] Referring to step 404, network slicer 206a determines that the packet is not a create session request (also referred to herein “other control messages”) and routes the packet to the destination IP associated with the request, which is SGW 232 in slice 1 230. In this way, the routers 208 and 210 act as normal routers.

[0072] Referring to step 406, network slicer 206a receives a response from SGW 232 in slice 1 230, which has a source IP address associated with from SGW 232 in slice 1 and a destination address associated with MME 202.

[0073] Referring to step 408, network slicer 206a determines that the packet is not a create session response and routes the packet to the destination IP associated with the response, which is the IP address associated with MME 202. In this way, the routers 208 and 210 act as normal routers.

[0074] The message flow in FIG. 4B can be associated with the minimally intrusive network slicer 206b shown in FIG. 2B, according to some embodiments. As shown in FIG. 4B, referring to step 422, MME 202 routes a GTP-C that is not a Create Session Request to

directly to the SGW 232 in slice 1 230. This is because, as described in more detail below, the left side element, such as the MME 202, will have received an indication in a response message of a selected slice, and will route traffic directly to that selected slice. Referring to step 424, network slicer 206b is not involved in traffic between MME 202 and SGW 232 in slice 1 230. Similarly, in step 432, the network slicer 206b is bypassed in communications between MME 202 and SGW 232 in slice 1 230.

[0075] FIGS. 5A and 5B are flow charts showing message flow for GTP-C messages that are Create Session Requests or Create PDP Context Requests for the non-intrusive network slicer 206a (FIG. 5A) and the minimally intrusive network slicer (FIG. 5B), according to some embodiments of the present disclosure. Similarly labeled steps function similarly between FIGS. 5A and 5B unless otherwise noted below.

[0076] Referring to steps 502a and 205b, a network element, such as an MME 202 routes a GTP-C that is a create session request to the network slicer 206a/206b. The request has a source IP address associated with the MME 202 and a destination IP address of an SGW 232 in slice 1 230. In some embodiments, as shown in the non-intrusive embodiment of FIG. 5A, the MME 202 request was routed to the network slicer 206a because the network slicer 206a informed the next hop router (e.g., a router between the left side network element and the network slicer 206a) or MME, SGSN, SGW, TWAN, and ePDG directly that the network slicer 206a has a cheaper routing cost than other routers in the network. For example, the network slicer 206a may be configured to indicate a routing cost that is zero or below an expected threshold that other routers are expected to indicate. Accordingly, routers attempting to route from the left side elements to the right side elements in FIG. 2A will route messages, including creation messages and non-creation messages, through the network slicer 206a because the routers see the network slicer 206a as having lower routing costs to the network slices relative to other routers. As shown in FIG. 5B, during deployment of the network slicer 206b, the DNS 120 can be updated to include the network slicer 206b of FIG. 2B as the highest priority right side node. This can be accomplished by a manual reconfiguration during deployment, or can be performed automatically by the network slicer 206b. Thus, when a left side element queries the DNS for a right side element to send the creation request, the DNS 120 will return the IP address 214 as the highest priority element even though the network slicer is not a PGW, SGW, or GGSN. Accordingly, the network

slicer 206b receives initial creation messages from the left side elements to the right side elements directly at the IP address 214.

[0077] Referring to step 504a, network slicer 206a determines that the packet is a create session request and applies a policy to determine a better slice than the one indicated in the create session request. In step 504b, the network slicer 206b may not need to determine that the packet is a create request because network elements such as MME 202 only route creation requests to the network slicer 206b explicitly using IP address 214 (this can be because future messages can be routed directly to the right side element selected by the network slicer 206b). In steps 504a and 504b, network slicers 206a/206b apply one or more policies to identify if a different network slice should be used. In some embodiments, the network slicers 206a/206b use the information elements from the Create Session Requests and/or the source IP of the MME/SGSN as the input. The information used includes (but is not limited to) any one or more of IMSI, Location, Access Point Name (APN), Device id IMEI, Radio Access Technology type and Time of Day. Network slicer 206a/206b then searches through the locally configured policies for the best match based on the above-identified information. For example, a best match can be specified by an operator that wants to use a dedicated SGW and a dedicated PGW for a particular group of users, such as those at a certain location, those with a certain IMSI, those with a certain APN type, etc. The locally configured policy can identify the IMSI of the users and link them to a specified SGW and PGW. The policy can also refer to a linkage to an external policy server. The network slicer 206a/206b then queries the external server to determine the best slice to use. External servers can contain similar policies that instruct the network slicer 206a/206b to make similar selections to the locally configured policies based on similar information. Network slicer 206a/206b can also assign its own IP address as the source IP address for the modified Create Session Request or Create PDP Context Request. A person of ordinary skill in the art would understand from the present disclosure that, rather than modifying the existing Create Session Request or Create PDP Context Request, a new Create Session Request or Create PDP Context Request could be created with the required information.

[0078] Referring to step 506, network slicer 206a/206b transmits a GTP-C create session request to the slice selected based on the policy (which is, for example, SGW 242 in slice 2 (240) in this figure) and sets the source address as the network slicer's own IP address.

[0079] Referring to step 508, network slicer 206a/206b receives a response from SGW 242 in slice 2 240, which has a source IP address associated with from SGW 242 in slice 2 and a destination address associated with the network slicer 206a/206b (e.g., IP address 212).

[0080] Referring to steps 510a and 510b, network slicer 206a/206b determines that the packet is a create session response and creates a new response (or modifies the existing response) with a source IP address associated with slice 1 230 (the source address expected by the MME in FIG. 5A) or network slicer 206b (the source address expected by the MME in FIG. 5B) and the destination IP address associated with MME 202. In some embodiments, the message content from the SGW in slice 2 can already contain the IP address that the MME 202 shall use for future communications with the SGW. Exemplary content and relevant handling by the MME 202 are defined, for example, by 3GPP (e.g., 29.274 and 29.060).

[0081] Referring to steps 512a and 512b, network slicer 206a transmits the response to MME 202 with an updated source IP (SGW 232 of slice 1 230 in FIG. 5A or network slicer 206b in FIG. 5B) and a destination address of the MME 202. Even though the source IP indicates that the response came from slice 1 230, the content of the response is from SGW 242 of slice 2 240. The content includes information informing the MME 202 of the identity of the selected slice and gateway by the network slicer 206a/206b such that MME 202 can communicate with the selected slice and gateway in future messaging related to the UE session. In some embodiments, the IP address of the selected slice can be included in the message content in compliance with, for example, 3GPP standards 29.274 for GTPv2 messages, or 29.060 for GTPv1 messages. In some embodiments, Slice 1 and Slice 2 are two different sets of SGW, PGW and GGSNs. The slices may offer different capabilities. For example, Slice 1 may only have an SGW, PGW and GGSN that can handle regular speed user traffic whereas Slice 2 may have an SGW, PGW and GGSN that can support higher speed user traffic. As another example, a network slicer 206a can be used to direct UEs that are capable of a 5G level of throughput to a high-speed slice such that the network can provide the best service to the UEs.

[0082] In some embodiments, multiple instances of network slicers can be used in a network to provide redundancy. In such a redundant embodiment, multiple slicers can be introduced into the network, and, for non-intrusive embodiments, subsets thereof or all of

them can be treated as equal cost routes towards the SGW/PGW and GGSN. Removal or failure of one such slicer does not impact the operation of the network because the messages can be routed via other slicer instances. In some minimally intrusive embodiments, each network slicer can be given its own IP address 214. Accordingly, when one of the plural network slicers fails, the left side elements can detect the failure and use other network slicers instead.

[0083] FIGS. 6A and 6B are flow charts showing processes of routing uplink packets in a network using non-intrusive and minimally intrusive network slicers, respectively, according to some embodiments of the present disclosure.

[0084] Referring to step 602, network slicer 206a receives a message from a first node. In some embodiments, the first node can comprise one of an MME, SGW, SGSN, TWAN, or ePDG.

[0085] Referring to step 604, network slicer 206a determines a message type associated with the first node. A message type can be one of a Create Session Request, Create PDP Request, or other control message (e.g., a non-Create Session request such as Modify Bearer Request, Delete Session Request etc.).

[0086] Referring to step 606, network slicer 206a selects a network slice based on a policy of the slicer when the message type comprises a Create Session Request or a Create PDP Request. As described above, when the message type comprises a Create Session Request or a Create PDP Request, network slicer 206a searches for a policy in either an internal or external policy database or server based on information associated with the first node, such as one or more of IMSI, Location, Access Point Name (APN), Device ID, IMEI, Radio Access Technology type, MS-ISDN, and Time of Day. In some embodiments, the network slicer may record the content of the incoming request, the target slice intended by the originating network node, the target slice selected, and/or other identifying information about the request, which can be used to populate and/or forward new response messages based on response messages received from right side elements. The use of such recorded information is described, for example, with reference to FIGs. 5A and 7. In some non-intrusive embodiments, the network slicer 206a may also receive non-creation request type messages. In such embodiments, non-creation request type messages are routed by the routers of the

network slicer without selecting a new network slice as discussed, for example, with reference to FIG. 4A.

[0087] Referring to step 608, network slicer 206a creates a new Create Session Request or a Create PDP Request (or modifies the received request) with a different source and destination IP address different from the source and destination IP address associated with the received message. For example, the source IP address associated with the new message is an IP address associated with the network slicer 206a and the destination IP address is associated with the selected network slice.

[0088] Referring to step 610, network slicer 206a routes the new Create Session Request or a Create PDP Request to a second node of the selected network slice. As described above, in some embodiments, the first node can comprise one of an MME, SGW, SGSN, TWAN, or ePDG. In some embodiments, the second node can comprise one of SGW, PGW, and GGSN. The network slicer 206a can route traffic between at least MME and SGW, SGSN and SGW, SGSN and GGSN, SGW and PGW, TWAN and PGW, and ePDG and PGW.

[0089] FIG. 6B shows a process for network slicer 206b that is similar to the process shown in FIG. 6A for network slicer 206a.

[0090] Referring to step 612, network slicer 206b receives a message from a first left side node. In some embodiments, the first node can comprise one of an MME, SGW, SGSN, TWAN, or ePDG. Because the network slicer 206b expects to receive creation requests directly from such network elements at the network slicer 206b's host IP address 214 as the destination address, the network slicer 206b does not need to determine a message type associated with the message and further does not need to inform next hop routers that the network slicer 206b has lower routing costs relative to other routers.

[0091] Referring to step 616, network slicer 206b selects a network slice based on a policy of the slicer. As described above, network slicer 206b searches for a policy in either an internal or external policy database or server based on information associated with the first node, such as one or more of IMSI, Location, Access Point Name (APN), Device ID, IMEI, Radio Access Technology type, MS-ISDN, and Time of Day. In some embodiments, the network slicer 206b may record the content of the incoming request, the target slice intended by the originating network node, the target slice selected, and/or other identifying information

about the request, which can be used to populate and/or forward new response messages based on response messages received from right side elements. The use of such recorded information is described, for example, with reference to FIGs. 5B and 7.

[0092] Referring to step 618, network slicer 206b creates a new Create Session Request or a Create PDP Request (or modifies the received request) with a different source and destination IP address different from the source and destination IP address associated with the received message. For example, the source IP address associated with the new message is an IP address associated with the network slicer 206b and the destination IP address is associated with the selected network slice.

[0093] Referring to step 620, network slicer 206b routes the new Create Session Request or a Create PDP Request to a second node of the selected network slice. As described above, in some embodiments, the first node can comprise one of an MME, SGW, SGSN, TWAN, or ePDG. In some embodiments, the second node can comprise one of SGW, PGW, and GGSN. The network slicer 206a can route traffic between at least MME and SGW, SGSN and SGW, SGSN and GGSN, SGW and PGW, TWAN and PGW, and ePDG and PGW.

[0094] FIG. 7 is a flow chart showing a process of routing downlink packets in a network using a non-intrusive or minimally intrusive network slicer, according to some embodiments of the present disclosure.

[0095] Referring to step 702, network slicer 206a/206b receives a Create Session Response or a Create PDP Response. The network slicer 206a/206b receives a Create Session Response or a Create PDP Response only for responses to requests that were routed through the network slicer 206a/206b.

[0096] Referring to step 704, network slicer 206a/206b creates a new response (or modifies the received response) based on information recorded when the original request was processed as well as the content of the incoming response. The information recorded includes the IP address of the MME, SGSN, TWAN, and/or ePDG, which sent the original Create Session Request or Create PDP Context Request. The information also includes the destination IP address of the original Create Session Request or Create PDP Context Request. In some embodiments, the recorded MME, SGSN, TWAN, and/or ePDG address becomes the destination IP address of the new response message. For network slicer 206a, the recorded

destination IP address from the original creation message can become the source IP of the new message. For the network slicer 206b, the host IP address 214 of the network slicer 206b can become the source IP address for the new message. The content of the incoming response message can be copied over to the new response message, but with a modification indicating that the selected slice (e.g., slice 2 in the example above) is to be used for future communications as discussed above. The IP address of the selected slice can be included in the message content in compliance with, for example, 3GPP standards 29.274 for GTPv2 messages, or 29.060 for GTPv1 messages.

[0097] Referring to step 706, network slicer 206a/206b transmits the new response to the first node. In some embodiments, the response includes content indicating an IP address of the selected slice and node as described above with reference to step 706. By providing this information, further messages sent by the first node to the network can be directed to the selected second node. In some embodiments, at step 706, the network slicer 206a/206b deletes any information recorded related to the creation messages discussed with reference to FIGs. 6A/6B and the response messages discussed with reference to FIG. 7. Thus the network no longer maintains any information related to the established session after forwarding a response message to the left side element, making the network slicer a “stateless” solution.

[0098] In addition to the network slicing capabilities described above, the network slicer 206a/206b can also be used for the following purposes discussed below.

[0099] According to some embodiments, the network slicers 206a and/or 206b can be used for gateway maintenance. The network slicer can be introduced to a running network to help migrate users from one gateway to others without impacting the service. This allows a network operator to perform upgrades, load balancing of a running network, or other tasks that can be facilitated by adaptive or controlled network slicing. Once the procedure is complete, the Network Slicer can be removed seamlessly.

[0100] According to some embodiments, the network slicers 206a and/or 206b can be used for trouble shooting. The network slicer can be used to force a PDN or PDP Context to a particular set of gateways, for example, for debugging purposes. The same function can also be used to test new capability of a gateway or users of a gateway. For example, particular users can be unloaded or offloaded from a new or updated gateway in a controlled

manner in order to test the new capabilities in a controlled manner. Because the network slicer has access to many different types of information, including, for example, IMSI, Location, Access Point Name (APN), Device id IMEI, Radio Access Technology type and Time of Day, a network operator can control the conditions of the test scenario for particular characteristics of users or other network elements.

[0101] In some embodiments, advantages of using network slicer 206a and/or 206b can include those described below.

[0102] In some embodiments, the network slicer can be transparent from the MME or SGSN's perspective. No changes to existing network function such as MME, SGSN, TWAN, ePDG, DNS server, SGW, PGW and GGSN are required. It is much easier to introduce the network slicer function as compared to traditional proxy approach. Furthermore, even for a minimally intrusive slicer, modification only needs to occur to the DNS server.

[0103] In some embodiments, the network slicers 206a, 206b can be stateless. The network slicer 206a/206b does not need to maintain the session state, rather the session state is still maintained by the MME/SGSN/TWAN/ePDG and the SGW/PGW/GGSN. Thus, the network slicer 206a/206b may not add another failure domain to the network. Failure of a network slicer may not impact established sessions.

[0104] In some embodiments, the network slicers disclosed herein can provide more comprehensive slice selection functions as compared to what can be achieved using DNS based solution on the existing MME/SGSN/TWAN/ePDG. The network slicer allows the operator to use any UE related information to determine the right slice to use.

[0105] The subject matter described herein can be implemented in digital electronic circuitry, or in computer software, firmware, or hardware, including the structural means disclosed in this specification and structural equivalents thereof, or in combinations of them. The subject matter described herein can be implemented as one or more computer program products, such as one or more computer programs tangibly embodied in an information carrier (e.g., in a machine readable storage device), or embodied in a propagated signal, for execution by, or to control the operation of, data processing apparatus (e.g., a programmable processor, a computer, or multiple computers). A computer program (also known as a program, software, software application, or code) can be written in any form of programming

language, including compiled or interpreted languages, and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment. A computer program does not necessarily correspond to a file. A program can be stored in a portion of a file that holds other programs or data, in a single file dedicated to the program in question, or in multiple coordinated files (e.g., files that store one or more modules, sub programs, or portions of code). A computer program can be deployed to be executed on one computer or on multiple computers at one site or distributed across multiple sites and interconnected by a communication network.

[0106] The processes and logic flows described in this specification, including the method steps of the subject matter described herein, can be performed by one or more programmable processors executing one or more computer programs to perform functions of the subject matter described herein by operating on input data and generating output. The processes and logic flows can also be performed by, and apparatus of the subject matter described herein can be implemented as, special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application specific integrated circuit).

[0107] Processors suitable for the execution of a computer program include, by way of example, both general and special purpose microprocessors, and any one or more processor of any kind of digital computer. Generally, a processor will receive instructions and data from a read only memory or a random access memory or both. The essential elements of a computer are a processor for executing instructions and one or more memory devices for storing instructions and data. Generally, a computer will also include, or be operatively coupled to receive data from or transfer data to, or both, one or more mass storage devices for storing data, e.g., magnetic, magneto optical disks, or optical disks. Information carriers suitable for embodying computer program instructions and data include all forms of nonvolatile memory, including by way of example semiconductor memory devices, (e.g., EPROM, EEPROM, and flash memory devices); magnetic disks, (e.g., internal hard disks or removable disks); magneto optical disks; and optical disks (e.g., CD and DVD disks). The processor and the memory can be supplemented by, or incorporated in, special purpose logic circuitry.

[0108] To provide for interaction with a user, the subject matter described herein can be implemented on a computer having a display device, e.g., a CRT (cathode ray tube) or LCD (liquid crystal display) monitor, for displaying information to the user and a keyboard and a

pointing device, (e.g., a mouse or a trackball), by which the user can provide input to the computer. Other kinds of devices can be used to provide for interaction with a user as well. For example, feedback provided to the user can be any form of sensory feedback, (e.g., visual feedback, auditory feedback, or tactile feedback), and input from the user can be received in any form, including acoustic, speech, or tactile input.

[0109] The subject matter described herein can be implemented in a computing system that includes a back end component (e.g., a data server), a middleware component (e.g., an application server), or a front end component (e.g., a client computer having a graphical user interface or a web browser through which a user can interact with an implementation of the subject matter described herein), or any combination of such back end, middleware, and front end components. The components of the system can be interconnected by any form or medium of digital data communication, e.g., a communication network. Examples of communication networks include a local area network (“LAN”) and a wide area network (“WAN”), e.g., the Internet.

[0110] It is to be understood that the disclosed subject matter is not limited in its application to the details of construction and to the arrangements of the components set forth in the following description or illustrated in the drawings. The disclosed subject matter is capable of other embodiments and of being practiced and carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein are for the purpose of description and should not be regarded as limiting.

[0111] As such, those skilled in the art will appreciate that the conception, upon which this disclosure is based, may readily be utilized as a basis for the designing of other structures, methods, and systems for carrying out the several purposes of the disclosed subject matter. It is important, therefore, that the claims be regarded as including such equivalent constructions insofar as they do not depart from the spirit and scope of the disclosed subject matter.

[0112] Although the disclosed subject matter has been described and illustrated in the foregoing exemplary embodiments, it is understood that the present disclosure has been made only by way of example, and that numerous changes in the details of implementation of the disclosed subject matter may be made without departing from the spirit and scope of the disclosed subject matter, which is limited only by the claims which follow.

CLAIMS

What is claimed is:

1. A method comprising:

receiving, at a network slicer in a networked system, a first message from a first network node of the networked system;

determining, by the network slicer, a destination address from the first message;

determining, by the network slicer, if a message type of the first message is a predetermined message type;

if the determined message type of the first message is not the predetermined message type, routing the first message to a second network node of the networked system associated with the destination address; and

if the determined message type of the first message is the predetermined message type:

determining a destination type from the first message,

selecting, by the network slicer, a second destination address for the first message based on one or more of a policy of the network slicer or an external policy, the second destination being associated with a third network node of the networked system having a destination type that is the same as the determined destination type,

creating, by the network slicer, a new message of the predetermined type including the second destination address, and

routing the new message of the predetermined type to the third network node of the networked system associated with the second destination address.

2. The method of claim 1, wherein the predetermined message type is a creation request message.

3. The method of claim 1, wherein the first node is one or more of a mobility management entity (MME), a serving general packet radio services (GPRS) support node (SGSN), serving gateway (SGW), enhanced packet data gateway (ePDG), or a trusted WLAN.

4. The method of claim 1, wherein the determined destination type comprises a gateway general packet radio services (GPRS) support node (GGSN), a serving gateway (SGW), or a packet data network gateway (PGW).

5. The method of claim 1, wherein the selecting the second destination address for the first message based on the one or more of the policy of the network slicer or the external policy comprises:

determining, by the network slicer, one or more of a throughput capability level of a user equipment (UE) associated with the first message, an enterprise associated with the UE associated with the first message, a traffic speed of the UE associated with the first message, a security level or requirement of the UE associated with the first message, or an employee status of the owner of the UE associated with the first message, and

selecting, by the network slicer, the second destination address associated with one or more of the capability level of the UE, the enterprise of the UE, the traffic speed of the UE, the security level or requirement of the UE, or the employee status of the owner of the UE from a plurality of second destination addresses.

6. The method of claim 1, wherein the selecting the second destination address for the first message based on the one or more of the policy of the network slicer or the external policy comprises determining one or more of an IMSI of a user equipment associated with the first message, location of the user equipment associated with the first message, access point name (APN) associated with the first message, an IMEI identification number of the user equipment associated with the first message, a Radio Access Technology type associated with the first message, or a timestamp of the first message.

7. The method of claim 1, further comprising:

receiving, by the network slicer, a response message from the third network node; and

creating, by the network slicer, a new response message including:

the destination address associated with the second network node as a source address of the new response message, and

an indication of the second destination address associated with the third network node in the message content of the new response message.

8. The method of claim 7, further comprising:

storing, at the network slicer, one or more of an address associated with the first network node or the destination address prior to the receiving the response message from the third network node; and

deleting, by the network slicer, the stored one or more of the address associated with the first network node or the destination address after the creating the new response message.

9. The method of claim 1, further comprising indicating to a next hop router by the network slicer that a router associated with the network slicer has a lower routing cost to the second network node than an expected routing cost from other routers in the networked system to the second network node prior to the receiving the first message, wherein the lower routing cost to the second network node is lower than an actual routing cost to the second network node.

10. A method comprising:

receiving, at a network slicer in a networked system, a creation request message from a first network node, the creation request message having as a first destination address an address associated with the network slicer;

determining a destination type from the creation request message;

selecting, by the network slicer, a second destination address for the creation request message based on one or more of a policy of the network slicer or an external policy, the destination being associated with a second network node of the networked system having a destination type that is the same as the determined destination type;

creating, by the network slicer, a new creation request message including the second destination address and the message content from the creation request message; and

routing the new creation request message to the second network node of the networked system associated with the second destination address.

11. The method of claim 10, further comprising:

receiving, by the network slicer, a response message from the second network node;
and

creating, by the network slicer, a new response message including the address associated with the second network node in the message content of the new response message and the address associated with the network slicer as the source address.

12. The method of claim 11, further comprising:

storing, at the network slicer, one or more of the address associated with the first network node or the second destination address prior to the receiving the response message from the second network node; and

deleting, by the network slicer, the stored one or more of the address associated with the first network node or the second destination address after the creating the new response message.

13. The method of claim 10, wherein the first node is one or more of a mobility management entity (MME), a serving general packet radio services (GPRS) support node (SGSN), serving gateway (SGW), enhanced packet data gateway (ePDG), or a trusted WLAN.

14. The method of claim 10, wherein the determined destination type comprises a gateway general packet radio services (GPRS) support node (GGSN), a serving gateway (SGW), or a packet data network gateway (PGW).

15. The method of claim 10, wherein the selecting the second destination address for the creation request message based on the one or more of the policy of the network slicer or the external policy comprises:

determining, by the network slicer, one or more of a throughput capability level of a user equipment (UE) associated with the first message, an enterprise associated with the UE associated with the first message, a traffic speed of the UE associated with the first message,

a security level or requirement of the UE associated with the first message, or an employee status of the owner of the UE associated with the first message, and

selecting, by the network slicer, the second destination address associated with one or more of the capability level of the UE, the enterprise of the UE, the traffic speed of the UE, the security level or requirement of the UE, or the employee status of the owner of the UE from a plurality of second destination addresses.

16. The method of claim 10, further comprising indicating to a domain name system of the networked system that the address associated with the network slicer has a higher priority than at least the second network node prior to the receiving the creation request message.

17. A network slicer in a networked system, the network slicer comprising:

a processor; and

a storage medium having instructions contained thereon that cause the processor to perform the following steps:

receive, at the network slicer in the networked system, a first message from a first network node of the networked system;

determine, by the network slicer, a destination address from the first message,

determine, by the network slicer, if a message type of the first message is a predetermined message type;

if the determined message type of the first message is not the predetermined message type, route the first message to a second network node of the networked system associated with the destination address; and

if the determined message type of the first message is the predetermined message type:

determine a destination type from the first message;

select, by the network slicer, a second destination address for the first message based on one or more of a policy of the network slicer or an external policy, the second destination being associated with a third network node of

the networked system having a destination type that is the same as the determined destination type,

create, by the network slicer, a new message of the predetermined type including the second destination address, and

route the new message of the predetermined type to the third network node of the networked system associated with the second destination address.

18. The network slicer of claim 17, wherein the predetermined message type is a creation request message.

19. The network slicer of claim 17, wherein the first node is one or more of a mobility management entity (MME), a serving general packet radio services (GPRS) support node (SGSN), serving gateway (SGW), enhanced packet data gateway (ePDG), or a trusted WLAN.

20. The network slicer of claim 17, wherein the determined destination type comprises a gateway general packet radio services (GPRS) support node (GGSN), a serving gateway (SGW), or a packet data network gateway (PGW).

21. The network slicer of claim 17, wherein the selecting the second destination address for the first message based on the one or more of the policy of the network slicer or the external policy comprises causing the processor to:

determine, by the network slicer, one or more of a throughput capability level of a user equipment (UE) associated with the first message, an enterprise associated with the UE associated with the first message, a traffic speed of the UE associated with the first message, a security level or requirement of the UE associated with the first message, or an employee status of the owner of the UE associated with the first message, and

select, by the network slicer, the second destination address associated with one or more of the capability level of the UE, the enterprise of the UE, the traffic speed of the UE, the security level or requirement of the UE, or the employee status of the owner of the UE from a plurality of second destination addresses.

22. The network slicer of claim 17, wherein the selecting the second destination address for the first message based on the one or more of the policy of the network slicer or the external policy comprises determining one or more of an IMSI of a user equipment associated with the first message, location of the user equipment associated with the first message, access point name (APN) associated with the first message, an IMEI identification number of the user equipment associated with the first message, a Radio Access Technology type associated with the first message, or a timestamp of the first message.

23. The network slicer of claim 17, wherein the processor is further caused to:

receive, by the network slicer, a response message from the third network node; and

create, by the network slicer, a new response message including:

the destination address associated with the second network node as a source address of the new response message, and

an indication of the second destination address associated with the third network node in the message content of the new response message.

24. The network slicer of claim 23, wherein the processor is further caused to:

store, at the network slicer, one or more of an address associated with the first network node or the destination address prior to the receiving the response message from the third network node; and

delete, by the network slicer, the stored one or more of the address associated with the first network node or the destination address after the creating the new response message.

25. The network slicer of claim 17, wherein the processor is further caused to indicate to a next hop router by the network slicer that a router associated with the network slicer has a lower routing cost to the second network node than an expected routing cost from other routers in the networked system to the second network node prior to the receiving the first message, wherein the lower routing cost to the second network node is lower than an actual routing cost to the second network node.

26. A network slicer in a networked system, the network slicer comprising:

a processor; and

a storage medium having instructions contained thereon that cause the processor to perform the following steps:

receive, at the network slicer in the networked system, a creation request message from a first network node, the creation request message having as a first destination address an address associated with the network slicer;

determine a destination type from the creation request message;

select, by the network slicer, a second destination address for the creation request message based on one or more of a policy of the network slicer or an external policy, the destination being associated with a second network node of the networked system having a destination type that is the same as the determined destination type;

create, by the network slicer, a new creation request message including the second destination address and the message content from the creation request message; and

route the new creation request message to the second network node of the networked system associated with the second destination address.

27. The network slicer of claim 26, wherein the processor is further caused to:

receive, by the network slicer, a response message from the second network node; and

create, by the network slicer, a new response message including the address associated with the second network node in the message content of the new response message and the address associated with the network slicer as the source address.

28. The network slicer of claim 27, wherein the processor is further caused to:

store, at the network slicer, one or more of the address associated with the first network node or the second destination address prior to the receiving the response message from the second network node; and

delete, by the network slicer, the stored one or more of the address associated with the first network node or the second destination address after the creating the new response message.

29. The network slicer of claim 26, wherein the first node is one or more of a mobility management entity (MME), a serving general packet radio services (GPRS) support node (SGSN), serving gateway (SGW), enhanced packet data gateway (ePDG), or a trusted WLAN.

30. The network slicer of claim 26, wherein the determined destination type comprises a gateway general packet radio services (GPRS) support node (GGSN), a serving gateway (SGW), or a packet data network gateway (PGW).

31. The network slicer of claim 26, wherein the selecting the second destination address for the creation request message based on the one or more of the policy of the network slicer or the external policy comprises causing the processor to:

determine, by the network slicer, one or more of a throughput capability level of a user equipment (UE) associated with the first message, an enterprise associated with the UE associated with the first message, a traffic speed of the UE associated with the first message, a security level or requirement of the UE associated with the first message, or an employee status of the owner of the UE associated with the first message, and

select, by the network slicer, the second destination address associated with one or more of the capability level of the UE, the enterprise of the UE, the traffic speed of the UE, the security level or requirement of the UE, or the employee status of the owner of the UE from a plurality of second destination addresses.

32. The network slicer of claim 26, wherein the processor is further caused to indicate to a domain name system of the networked system that the address associated with the network slicer has a higher priority than at least the second network node prior to the receiving the creation request message.