



- (51) International Patent Classification:
H04L 12/70 (2013.01)
- (21) International Application Number:
PCT/US2015/036344
- (22) International Filing Date:
18 June 2015 (18.06.2015)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
2202/CHE/2015 29 April 2015 (29.04.2015) IN
- (71) Applicant: **HEWLETT PACKARD ENTERPRISE DEVELOPMENT LP** [US/US]; 11445 Compaq Center Drive West, Houston, TX 77070 (US).
- (72) Inventor: **DEVARAJAN, Venkatavaradhan**; Sy.no.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN).
- (74) Agents: **FERGUSON, Christopher, W.** et al.; Hewlett Packard Enterprise, 3404 E. Harmony Road, Mail Stop 79, Fort Collins, CO 80528 (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,

[Continued on next page]

(54) Title: CLIENT COMMUNICATIONS IN MULTI-TENANT DATA CENTER NETWORKS

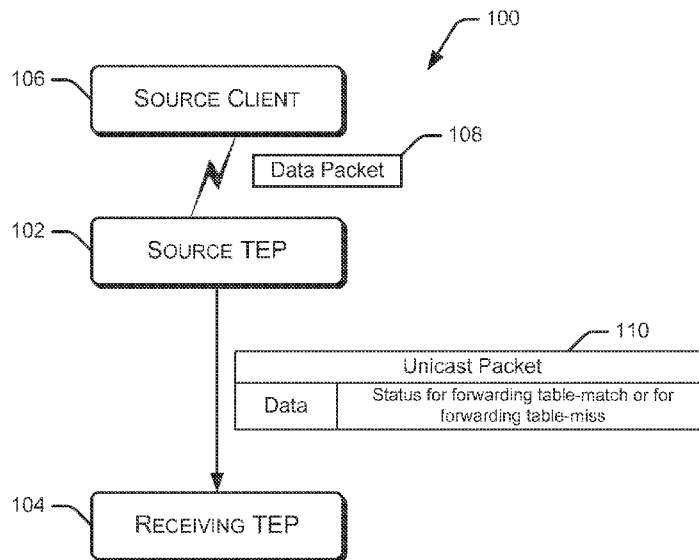


Fig. 1

(57) Abstract: The present subject matter relates to client communication in multi-tenant data center networks. In an example implementation, a unicast packet comprises a status indicative of one of a forwarding table-match and a forwarding table-miss at a source tunnel end-point (TEP) in a multi-tenant data center network.



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

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

Declarations under Rule 4.17:

- *as to the identity of the inventor (Rule 4.17(i))*
- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

CLIENT COMMUNICATIONS IN MULTI-TENANT DATA CENTER NETWORKS

BACKGROUND

[0001] A multi-tenant data center network generally involves a virtual internet protocol (IP) sub-network overlaid on an L3 fabric network. The L3 fabric network has multiple tunnel end-points (TEPs) connected in a meshed topology. The virtual IP sub-network has virtual network devices (VNETs) connected to the TEPs of the L3 fabric network. Clients, such as, personal computers, smartphones, tablets, servers, and the like, can connect to the multi-tenant data center network through any of the VNETs, and can communicate with each other through the respective VNETs and TEPs.

BRIEF DESCRIPTION OF DRAWINGS

[0002] The following detailed description references the drawings, wherein:

[0003] Fig. 1 illustrates a data center network environment, according to an example implementation of the present subject matter;

[0004] Fig. 2 illustrates a multi-tenant data center network environment for communication between a source client and a target client, according to an example implementation of the present subject matter;

[0005] Fig. 3 illustrates the multi-tenant data center network environment for communication between a source client and a target client, according to an example implementation of the present subject matter;

[0006] Fig. 4 illustrates the multi-tenant data center network environment for communication between a source client and a target client, according to an example implementation of the present subject matter;

[0007] Fig. 5 illustrates a method of client communication in a multi-tenant data center network, according to an example implementation of the present subject matter;

[0008] Fig. 6 illustrates a method of client communication in a multi-tenant data center network, according to an example implementation of the present subject matter;

[0009] Fig. 7 illustrates a method of client communication in a multi-tenant data center network, according to an example implementation of the present subject matter;

[0010] Fig. 8 illustrates a method of client communication in a multi-tenant data center network, according to an example implementation of the present subject matter; and

[0011] Fig. 9 illustrates a system environment, according to an example implementation of the present subject matter.

DETAILED DESCRIPTION

[0012] The present subject matter describes communication between clients in a multi-tenant data center network having a virtual IP sub-network overlaid on an L3 fabric network. Clients may be understood as devices, such as personal computers, smartphones, tablets, servers, and the like, that function as tenants to the multi-tenant data center network. Clients can connect to the multi-tenant data center network to communicate with each other.

[0013] A multi-tenant data center network, hereinafter referred to as the data center network, may deploy a tunneling scheme for communication across the L3 fabric network. The tunneling scheme may, for example, be based on a virtual extensible local area network (VxLAN) protocol, a network virtualization using generic routing encapsulation (NVGRE) protocol, a generic network virtualization encapsulation (GENEVE) protocol, or a stateless transport tunneling (STT) protocol. The tunneling scheme allows for the clients to roam across the data center network while retaining their IP address irrespective of the point of connection on the data center network.

[0014] The L3 fabric network and the virtual IP sub-network of the data center network are referred to as an underlay network and an overlay network, respectively. The underlay L3 fabric network includes tunnel end-points (TEPs) connected in a meshed topology, such that each TEP is one hop away from any other TEP in the mesh. The overlay virtual IP sub-network includes virtual

network devices (VNETs), with one VNET generally connected to one TEP of the underlay network. Clients roaming in the data center network can connect to any of the VNETs, and can communicate with each other through the connected VNETs and TEPs.

[0015] The connection between any two TEPs in the mesh is referred to as a tunneling path. According to the tunneling scheme, each TEP in the underlay L3 fabric network maintains a forwarding table with forwarding table (FT) entries for the clients. An FT entry for a client is indicative of a tunneling path applicable for the client, depending on the TEP associated with the client. A TEP is said to be associated with a client, when the client is locally connected to the TEP. A TEP learns about the location of a client and adds an FT entry for the client after receiving data from the client.

[0016] When data is to be sent from a source client to a target client, the TEP associated with the source client looks up its forwarding table for the FT entry for the target client. The TEP associated with the source client may hereinafter be referred to as a source TEP. If the FT entry for the target client is present, then the source TEP transmits the data to the TEP associated with the target client based on the tunneling path indicated by the FT entry. The TEP associated with the target client may hereinafter be referred to as a destination TEP. The destination TEP then forwards the data to the target client. If the FT entry for the target client is not present, then the source TEP transmits the data to all other TEPs in the mesh. The transmission of data to all the other TEPs is referred to as "data flooding". The TEPs that receive the flooded data look up their respective forwarding tables. The TEP having the target client, as a local client, forwards the data to the target client. For the purpose of description herein, the presence of an FT entry for a client in a forwarding table during a forwarding table lookup is referred to as a "forwarding table-match", and the absence of an FT entry for a client in a forwarding table during a forwarding table lookup is referred to as a "forwarding table-miss".

[0017] The data, whether sent to a destination TEP due to a forwarding table-match or transmitted to all the TEPs due to a forwarding table-miss, can be transmitted by the source TEP in a unicast packet. The unicast packet herein

may be understood as a packet configured for sending to a specific receiving TEP. The unicast packet includes the data of the source client for the target client. A TEP receiving the unicast packet cannot distinguish whether the packet is sent due to a forwarding table-match at the source TEP or is flooded due to a forwarding table-miss at the source TEP. In other words, with the unicast packet, the receiving TEP cannot determine whether the source TEP has an FT entry for the target client or not.

[0018] Further, when target client is roaming and has moved to a new TEP, the source TEP may not learn about the new TEP, i.e., the TEP currently associated with the target client, until the source TEP receives some data from the target client via the currently associated TEP. Till the source TEP learns about the currently associated TEP, the source TEP may either have a stale FT entry for the target client or have no FT entry at all for the target client. Further, the TEP previously associated with the target client may either have deleted the FT entry indicating the target client as a local client or have an updated FT entry indicating the tunneling path to the TEP currently associated with the target client. In a scenario where the source TEP has a stale FT entry and the previously associated TEP has deleted the FT entry for the target client, the source TEP sends a unicast packet to the previously associated TEP, but the data does not reach the target client. In a scenario where the source TEP has a stale FT entry and the previously associated TEP has the new FT entry for the target client, the source TEP sends a unicast packet to the previously associated TEP, but the previously associated TEP cannot forward the unicast packet to the currently associated TEP for further sending the data to the target client. In both the above scenarios, the data cannot reach the target client, which results in a “traffic black hole” from the source client to the target client. This traffic black hole continues till the source TEP learns about the TEP currently associated with the target client based on data sent by the target client via the currently associated TEP.

[0019] Further, in a scenario where the source TEP does not have the FT entry for the target client, the source TEP generates replicated unicast packets and floods the replicated unicast packets to all the other TEPs in the mesh.

Each of the TEPs, on receiving a replicated unicast packet, looks up its forwarding table for the FT entry for target client as a local client. The TEP currently associated with the target client then forwards the data to the target client. In this scenario, the source TEP continues to flood replicated unicast packets to all the other TEPs, till the source TEP learns about the TEP currently associated with the target client based on data sent by the target client via the currently associated TEP. This results in increase in network utilization in the underlay network.

[0020] The present subject matter describes methods and systems for communication between clients in a multi-tenant data center network. The methods and the systems of the present subject matter facilitate in reducing the time duration of traffic black hole from a source client to a target client and in reducing the network utilization in the undelay network when the target client is roaming and has moved to a new TEP.

[0021] In accordance with an example implementation of the present subject matter, a source TEP transmits unicast packets with a status indicative of a forwarding table-match or a forwarding table-miss at the source TEP. With this status in the unicast packet, any receiving TEP can determine whether the source TEP has an FT entry for the target client in its forwarding table or not. With this determination, the receiving TEP can initiate actions, in accordance with the present subject matter, for reducing the time duration of traffic black hole and for reducing the network utilization of the underlay network when the target client has moved to a new TEP.

[0022] In a scenario where the source TEP has a stale FT entry and the previously associated TEP has deleted the FT entry for the target client, the source TEP transmits a unicast packet to the previously associated TEP with a status indicative of the forwarding table-match. Since the previously associated TEP does not have the FT entry for the target client, the previously associated TEP determines that the source TEP has a stale FT entry for the target client. At this, the previously associated TEP may send a response message to the source TEP with an indication to delete the stale FT entry for the target client. On receiving the response message, the source TEP may delete the stale FT

entry and transmit subsequent unicast packets for the target client to all other TEPs with a status indicative of the forwarding table-miss. With this, the TEP currently associated with the target client, on receiving a unicast packet, sends the data to the target client. The actions initiated by the previously associated TEP and the actions performed by the source TEP, as described above, enable a reduction in the time duration of traffic black hole from the source client to the target client.

[0023] In a scenario where the source TEP has a stale FT entry and the previously associated TEP has the new FT entry for the target client indicating the currently associated TEP, the source TEP transmits a unicast packet to the previously associated TEP with a status indicative of the forwarding table-match. The previously associated TEP may forward the unicast packet to the currently associated TEP with a status indicating relaying of the packet by the previously associated TEP. The relaying of the unicast packet to the currently associated TEP helps in substantially eliminating the traffic black hole for the target client. Also, with the status indicating the relaying of the packet, the current associated TEP determines that the previous associated TEP is not the original source of the unicast packet, but a packet forwarder for data from the source client.

[0024] Further, since the previously associated TEP has the new FT entry indicating the currently associated TEP, the previously associated TEP determines that the source TEP has a stale FT entry for the target client. At this, the previously associated TEP may send a response message to the source TEP with an indication to update the stale FT entry for the target client. On receiving the response message, the source TEP may update the stale FT entry and transmit subsequent unicast packets for the target client to the currently associated TEP with a status indicative of the forwarding table-match. The actions initiated by the previously associated TEP and the actions performed by the source TEP, as described above, enable a reduction in the time duration of transmission of packets from the source client to the target client.

[0025] Further, in a scenario where the source TEP does not have the FT entry for the target client, the source TEP transmits replicated unicast packets to

all the other TEPs with a status indicative of the forwarding table-miss. The TEP currently associated with the target client accordingly then sends the data to the target client. Since the unicast packet has the status of the forwarding table-miss, the currently associated TEP determines that the source TEP does not have an FT entry for the target client. At this, the currently associated TEP may send a response message to the source TEP with an indication to add the FT entry for the target client. On receiving the response message, the source TEP may add the FT entry and transmit subsequent unicast packets for the target client to the currently associated TEP with a status indicative of the forwarding table-match. The actions initiated by the currently associated TEP and the actions performed by the source TEP, as described above, enable a reduction in the network utilization of the underlay network.

[0026] The following detailed description refers to the accompanying drawings. Wherever possible, the same reference numbers are used in the drawings and the following description to refer to the same or similar parts. While several examples are described in the description, modifications, adaptations, and other implementations are possible. Accordingly, the following detailed description does not limit the disclosed examples. Instead, the proper scope of the disclosed examples may be defined by the appended claims.

[0027] Fig. 1 illustrates a data center network environment 100, according to an example implementation of the present subject matter. The data center network environment 100 may be a multi-tenant network environment having a virtual IP sub-network overlaid on an L3 fabric network for client communication. A tunneling scheme based on, for example, a VxLAN protocol, an NVGRE protocol, a GENEVE protocol, or an STT protocol, may be deployed for communication across the L3 fabric network. Clients, such as personal computers, smartphones, tablets, servers, and the like, can roam across the data center network while retaining their IP address irrespective of the point of connection on the data center network.

[0028] The L3 fabric network of the data center network environment 100 includes a source TEP 102 and a receiving TEP 104, as shown. The source TEP 102 is a TEP to which a source client 106 is connected locally. The

receiving TEP 104 can be any TEP of the underlay network, connected to the source TEP 102 for receiving packets from and transmitting packets to the source TEP 102. Each of the source TEP 102 and the receiving TEP 104 maintains a forwarding table with FT entries for the clients depending on the TEP associated with the respective client. It may be noted that, for the sake of simplicity, only two TEPs, the source TEP 102 and the receiving TEP 104 are shown in the data center network environment 100; however, the data center network environment 100 may include more than two TEPs.

[0029] When the source client 106 wishes to send data to a target client, the source client 106 sends data in the form of a data packet 108 to the source TEP 102. The source TEP 102 on receiving the data packet 108 looks up its forwarding table and determines whether the forwarding table has the FT entry for the target client. The source TEP 102 generates and transmits a unicast packet 110 with the data and a status of a forwarding table-match when the FT entry for the target client is present in the forwarding table. The source TEP 102 generates and transmits the unicast packet 110 with the data and a status of a forwarding table-miss when the FT entry for the target client is not present in the forwarding table.

[0030] In an example implementation, the source TEP 102 may set a reserved bit in a packet header of the unicast packet 110 to '0' to indicate the forwarding table-match, and set the reserved bit in the packet header to '1' to indicate the forwarding table-miss. The receiving TEP 104, on receiving the unicast packet 110, decodes the packet header to determine the status of the reserved bit, and accordingly determine whether the source TEP 102 has the FT entry for the target client or not. The determination of whether the source TEP has the FT entry for the target client or not, based on the status in the unicast packet, enables initiation of actions to reduce the time duration of traffic black hole and to reduce utilization of the underlay network. The actions initiated based on the determination are described in detail through the description with reference to Fig. 2, Fig. 3, and Fig.4.

[0031] In an example implementation, a TEP of the underlay network may include one or more processors. The processor(s) may be implemented as

microprocessors, microcomputers, microcontrollers, digital signal processors, central processing units, state machines, logic circuitries, and/or any devices that manipulate signals based on operational instructions. Among other capabilities, the processor(s) may fetch and execute computer-readable instructions stored in a memory coupled to the processor(s). The memory can be internal or external to the TEP. The memory may include any non-transitory computer-readable storage medium including, for example, volatile memory (e.g., RAM), and/or non-volatile memory (e.g., EPROM, flash memory, NVRAM, memristor, etc.). The functions of the various elements shown in Fig. 1, including any functional blocks referring to a TEP may be provided through the use of dedicated hardware as well as hardware capable of executing computer-readable instructions.

[0032] Fig. 2, Fig. 3, and Fig. 4 illustrate a multi-tenant data center network environment 200 for communication between a source client and a target client, according to an example implementation of the present subject matter. As shown, the multi-tenant data center environment 200 includes an underlay network 202 having four TEPs, namely TEP 1 204-1, TEP 2 204-2, TEP 3 204-3, and TEP 4 204-4. The TEPs as shown are connected in a meshed topology, such that each TEP is just one hop away from the other TEP in the mesh. The multi-tenant data center network environment 200 also includes an overlay network 206 having VNETs 206-1, 206-2, 206-3, 206-4, with one VNET connected to one TEP. It may be noted that the VNETs 206-1, ... , 206-4 may share a same sub-net. Further, as shown, four clients, namely client 1 210-1, client 2 210-2, client 3 210-3, and client 4 210-4, are connected to the VNETs of the overlay network 206. Client 2 is roaming and has moved from TEP 2 to TEP 3.

[0033] Each of the TEPs 204-1, ... , 204-4 maintains a forwarding table with FT entries indicative of tunneling paths for the clients 210-1, ... 210-4. Table 1 shows a forwarding table that may be maintained by TEP 1 before client 2 has moved from TEP 2 to TEP 3. Since client 1 is local to TEP 1, the FT entry for client 1 in the forwarding table of TEP 1 does not indicate a tunneling path. Since client 2 is connected to TEP 2, the FT entry for client 2 indicates a

tunneling path TEP 1 to TEP 2. Since client 3 is connected to TEP 3, the FT entry for client 3 indicates a tunneling path TEP 1 to TEP 3. Similarly, since client 4 is connected to TEP 4, the FT entry for client 4 indicates a tunneling path TEP 1 to TEP 4.

Table 1

Forwarding Table at TEP 1	
Client 1	Local
Client 2	Tunneling path TEP 1 to TEP 2
Client 3	Tunneling path TEP 1 to TEP 3
Client 4	Tunneling path TEP 1 to TEP 4

[0034] For the purpose of description herein, consider a case where client 1 wishes to send data to client 2, where client 2 is roaming and has moved from TEP 2 to TEP 3 as shown in Fig. 2 to Fig. 4. Client 1 may be referred to as a source client, and client 2 may be referred to as a target client. Further, TEP 1 may be referred to as a source TEP, and TEP 3 may be referred to as a currently associated TEP of client 2.

[0035] Fig. 2 illustrates a scenario where client 2 has moved from TEP 2 to TEP 3, TEP 1 has a stale FT entry for client 2, and TEP 2 has deleted the FT entry for client 2. As shown, the forwarding table 212 of TEP 1 has the FT entry for client 2 indicating a tunneling path TEP 1 to TEP 2, and the forwarding table 214 of TEP 2 has no FT entry for client 2. On receiving the data in the form of a data packet 216 from client 1, TEP 1 looks up its forwarding table 212 and determines whether the forwarding table 212 has the FT entry for client 2. Since TEP 1 has the stale FT entry for client 2, TEP 1 transmits a unicast packet 218 to TEP 2 with the data and a status indicative of a forwarding table-match. In an example, a reserved bit in a packet header of the unicast packet 218 may be set as '0' to indicate the forwarding table-match. TEP 2 on receiving the unicast packet 218 decodes the packet header and determines based on the status that TEP 1 has the FT entry for client 2. Since TEP 2 does not have the FT entry for client 2, TEP 2 determines that TEP 1 has a stale FT entry for client 2. After determining this, TEP 2 sends a response message 220 to TEP 1 with an

indication that the forwarding table 214 of TEP 2 does not include the FT entry for client 2 and TEP 1 can delete the stale FT entry for client 2. In an example implementation, the response message 220 may include the following information: client # = 2 and TEP ID = Null. On receiving the response message 220, TEP 1 deletes the stale FT entry from its forwarding table 212. After deleting the stale FT entry for client 2, TEP 1 transmits subsequent unicast packets for client 2 to all the TEPs with a status indicative of a forwarding table-miss. In an example implementation, the reserved bit in the packet header of the unicast packet may be set to '1' to indicate the forwarding table-miss. TEP 3 on receiving the unicast packet from TEP 1 sends the data to client 2.

[0036] Fig. 3 illustrates a scenario where client 2 has moved from TEP 2 to TEP 3, TEP 1 does not have the FT entry for client 2. As shown, the forwarding table 302 of TEP 1 has no FT entry for client 2. Since client 2 is moved to TEP 3, the forwarding table 304 of TEP 3 has an FT entry for client 2 indicating that client 2 is local to TEP 3. On receiving the data in the form of a data packet 306 from client 1, TEP 1 looks up its forwarding table 302 and determines whether the forwarding table 302 has the FT entry for client 2. Since TEP 1 does not have the FT entry for client 2, TEP 1 transmits a replicated unicast packet 308-1, 308-2, 308-3 to each of TEP 2, TEP 3, and TEP 4 with the data and a status indicative of a forwarding table-miss. In an example, a reserved bit in a packet header of the unicast packet 218 may be set as '1' to indicate the forwarding table-miss. Each of TEP 1, TEP 2, and TEP 3 on receiving the unicast packet from TEP 1 looks up its forwarding table and determines whether the respective forwarding table has the FT entry for client 2. Since TEP 3 determines from its forwarding table 304 that client 2 is local to it, TEP 3 sends the data to client 2.

[0037] Further, since the unicast packet 308-2 has the status of forwarding table-miss, TEP 3 determines based on the status that TEP 1 does not have the FT entry for client 2. After determining this, TEP 3 sends a response message 310 to TEP 1 with an indication that TEP 1 can add an FT entry for client 2. On receiving the response message 310, TEP 1 adds an FT entry for client 2 in its forwarding table 302. After adding the FT entry for client 2, TEP 1 transmits

subsequent unicast packets for client 2 to TEP 3 with a status indicative of a forwarding table-match.

[0038] In an example implementation, the response message 310 may include the following information: client # = 2 and TEP ID = TEP 3. In an example implementation, the response message 310 may also include a timestamp information indicative of a time at which TEP 3 learnt about client 2. The timestamp information allows TEP 1 to add the FT entry for client 2 based on the latest location of client 2.

[0039] Fig. 4 illustrates a scenario where client 2 has moved from TEP 2 to TEP 3, TEP 1 has a stale FT entry for client 2, and TEP 2 has an FT entry for client 2 that indicates a tunneling path TEP 2 to TEP 3. As shown, the forwarding table 402 of TEP 1 has the FT entry for client 2 indicating a tunneling path TEP 1 to TEP 2, and the forwarding table 404 of TEP 2 has the FT entry for client 2 indicating a tunneling path TEP 2 to TEP 3. On receiving the data in the form of a data packet 406 from client 1, TEP 1 looks up its forwarding table 402 and determines whether the forwarding table 402 has the FT entry for client 2. Since TEP 1 has the stale FT entry for client 2, TEP 1 transmits a unicast packet 408 to TEP 2 with the data and a status indicative of a forwarding table-match. In an example, a reserved bit in a packet header of the unicast packet 408 may be set as '0' to indicate the forwarding table-match. Since TEP 2 has the FT entry for client 2, TEP 2 forwards a unicast packet 410 to TEP 3 with a status indicating relaying of the unicast packet 410 by TEP 2. In an example implementation, a reserved bit in the packet header of the unicast packet may be set as '1' to indicate that the unicast packet is relayed by TEP 2, and may be set as '0' to indicate that the unicast packet is originally transmitted by the TEP 2. It may be noted that the reserved bit used for indicating the relaying of packets is different from the reserved bit used for indicating the forwarding table-match and the forwarding table-miss. TEP 3 on receiving the unicast packet 410 from TEP 2 sends the data to client 2.

[0040] Further, TEP 2 on receiving the unicast packet 408 decodes the packet header and determines based on the status that TEP 1 has the FT entry for client 2. Since TEP 2 has the FT entry for client 2 that indicates the client 2 is

local to TEP 3, TEP 2 determines that TEP 1 has a stale FT entry for client 2. After determining this, TEP 2 sends a response message 412 to TEP 1 with an indication that TEP 1 can update the stale FT entry for client 2. On receiving the response message 412, TEP 1 updates the stale FT entry in its forwarding table 402. After updating the stale FT entry for client 2, TEP 1 transmits subsequent unicast packets for client 2 to TEP 3 with a status indicative of a forwarding table-match.

[0041] In an example implementation, the response message 412 may include the following information: client # = 2 and TEP ID = TEP 3. In an example implementation, the response message 412 may also include a timestamp information indicative of a time at which TEP 2 learnt about TEP 3 having client 2 as a local client. The timestamp information allows TEP 1 to update the FT entry for client 2 based on the latest location of client 2.

[0042] Fig. 5 illustrates a method 500 of client communication in a multi-tenant data center network, according to an example implementation of the present subject matter. The order in which the method 500 is described is not intended to be construed as a limitation, and any number of the described method steps can be combined in any order to implement the method 500. Furthermore, the method 500 can be implemented by processor(s) or computing device(s) through any suitable hardware, a non-transitory machine readable medium, or combination thereof. Further, although the method 500 is described in context of a source TEP in a multi-tenant data center network, other suitable computing devices or systems may be used for execution of at least one step of method 500. It may be understood that steps of method 500 can be executed based on instructions stored in a non-transitory computer readable medium, as will be readily understood. The non-transitory computer readable medium may include, for example, digital memories, magnetic storage media, such as a magnetic disks and magnetic tapes, hard drives, or optically readable digital data storage media.

[0043] Referring to Fig. 5, at block 502, a data packet of a source client for a target client is received at a source TEP. At block 504, it is determined whether a forwarding table of the source TEP includes an FT entry for the target

client. Based on the determination, if the FT entry for the target client is present in the forwarding table of the source TEP, then a unicast packet is transmitted to a destination TEP with a status indicative of a forwarding table-match, at block 506. Based on the determination, if the FT entry for the target client is not present in the forwarding table of the source TEP, a unicast packet is transmitted to all other TEPs with a status indicative of a forwarding table-miss, at block 508.

[0044] In an example implementation, a reserved bit in a packet header of the unicast packet may be set to a first status, for example '0', to indicate the forwarding table-match, and may be set to a second status, for example '1', to indicate the forwarding table-miss.

[0045] Fig. 6 illustrates a method 600 of client communication in a multi-tenant data center network, according to an example implementation of the present subject matter. The order in which the method 600 is described is not intended to be construed as a limitation, and any number of the described method steps can be combined in any order to implement the method 600. Furthermore, the method 600 can be implemented by processor(s) or computing device(s) through any suitable hardware, a non-transitory machine readable medium, or combination thereof. Further, although the method 600 is described in context of a source TEP in a multi-tenant data center network, other suitable computing devices or systems may be used for execution of at least one step of method 600. It may be understood that steps of method 600 can be executed based on instructions stored in a non-transitory computer readable medium, as will be readily understood. The non-transitory computer readable medium may include, for example, digital memories, magnetic storage media, such as magnetic disks and magnetic tapes, hard drives, or optically readable digital data storage media.

[0046] Referring to Fig. 6, at block 602, a unicast packet is transmitted to a destination TEP with a status indicative of a forwarding table-match when an FT entry for a target client is present in a forwarding table of a source TEP. At block 604, a response message is received from the destination TEP indicating that a forwarding table of the destination TEP does not include an FT entry for the

target client. Based on the response message, the FT entry for the target client is deleted from the forwarding table of the source TEP, at block 606. In an example implementation, the response message may include the information of the target client and a TEP ID as "Null".

[0047] Fig. 7 illustrates a method 700 of client communication in a multi-tenant data center network, according to an example implementation of the present subject matter. The order in which the method 700 is described is not intended to be construed as a limitation, and any number of the described method steps can be combined in any order to implement the method 700. Furthermore, the method 700 can be implemented by processor(s) or computing device(s) through any suitable hardware, a non-transitory machine readable medium, or combination thereof. Further, although the method 700 is described in context of a source TEP in a multi-tenant data center network, other suitable computing devices or systems may be used for execution of at least one step of method 700. It may be understood that steps of method 700 can be executed based on instructions stored in a non-transitory computer readable medium, as will be readily understood. The non-transitory computer readable medium may include, for example, digital memories, magnetic storage media, such as a magnetic disks and magnetic tapes, hard drives, or optically readable digital data storage media.

[0048] Referring to Fig. 7, at block 702, a unicast packet is transmitted to all other TEPs with a status indicative of a forwarding table-miss when an FT entry for a target client is not present in a forwarding table of a source TEP. At block 704, a response message is received from a TEP currently associated with the target client indicating that a forwarding table of the currently associated TEP includes an FT entry for the target client. The currently associated TEP is from the other TEPs that received the unicast packet with the status indicative of the forwarding table-miss. Based on the response message, the FT entry for the target client is added in the forwarding table of the source TEP, at block 706. In an example implementation, the response message may include the information of the target client and a TEP ID as TEP 3. In an example implementation, the response message may also include a timestamp information indicative of a

time at which the currently associated TEP learnt about the target client. The timestamp information allows the source TEP to add the FT entry for the target client based on the latest location of the target client.

[0049] Fig. 8 illustrates a method 800 of client communication in a multi-tenant data center network, according to an example implementation of the present subject matter. The order in which the method 800 is described is not intended to be construed as a limitation, and any number of the described method steps can be combined in any order to implement the method 800. Furthermore, the method 800 can be implemented by processor(s) or computing device(s) through any suitable hardware, a non-transitory machine readable medium, or combination thereof. Further, although the method 800 is described in context of a source TEP in a multi-tenant data center network, other suitable computing devices or systems may be used for execution of at least one step of method 800. It may be understood that steps of method 800 can be executed based on instructions stored in a non-transitory computer readable medium, as will be readily understood. The non-transitory computer readable medium may include, for example, digital memories, magnetic storage media, such as a magnetic disks and magnetic tapes, hard drives, or optically readable digital data storage media.

[0050] Referring to Fig. 8, at block 802, a unicast packet is transmitted to a destination TEP with a status indicative of a forwarding table-match when an FT entry for a target client is present in a forwarding table of a source TEP. At block 804, a response message is received from the destination TEP comprising information of a TEP currently associated with the target client. Based on the information of the currently associated TEP in the response message, the FT entry for the target client is updated in the forwarding table of the source TEP, at block 806. In an example implementation, the response message may include the following information of the target client and a TEP ID as TEP 3. In an example implementation, the response message may also include a timestamp information indicative of a time at which the destination TEP learnt about the currently associated TEP having the target client as a local client. The

timestamp information allows the source client to update the FT entry for the target client based on the latest location of the target client.

[0051] Further, in an example implementation, the destination TEP may forward the unicast packet to the currently associated TEP with a status indicative of relaying of the unicast packet by the destination TEP to the currently associated TEP. In an example implementation, a reserved bit in a packet header of the unicast packet may be set to a first status, for example '1', to indicate the relaying of the unicast packet, and may be set to a second status, for example '0', to indicate the original transmission of the unicast packet.

[0052] Fig. 9 illustrates a system environment 900, according to an example implementation of the present subject matter. In an example implementation, the system environment 900 includes a processor 902 communicatively coupled to a non-transitory computer readable medium 904 through a communication link 906. The processor 902 functions to fetch and execute computer-readable instructions from the non-transitory computer readable medium 904. The processor 902 may be a processing resource of a TEP in an underlay network of a multi-tenant data center network, in accordance with the present subject matter. The processor 902 and the non-transitory computer readable medium 904 are also communicatively coupled to other TEP(s) 908.

[0053] The non-transitory computer readable medium 904 can be, for example, an internal memory device or an external memory device. In an example implementation, the communication link 906 may be a direct communication link, such as any memory read/write interface. In another example implementation, the communication link 906 may be an indirect communication link, such as a network interface. In such a case, the processor 902 can access the non-transitory computer readable medium 904 through a network (not shown). The network may be a single network or a combination of multiple networks and may use a variety of different communication protocols.

[0054] In an example implementation, the non-transitory computer readable medium 904 includes a set of computer readable instructions that

facilitate in reducing the time duration of traffic black hole and utilization of underlay network when a target client is roaming and has moved from one TEP to another TEP. The set of computer readable instructions can be accessed by the processor 902 through the communication link 906 and subsequently executed to perform acts for communication between clients in a multi-tenant data center network.

[0055] Referring to Fig. 9, in an example, the non-transitory computer readable medium 904 includes instructions 910 that cause the processor 902 to receive at a TEP a data packet of a source client for a target client. The non-transitory computer readable medium 904 includes instructions 912 that cause the processor 902 to transmit a unicast packet to a destination TEP with a status indicative of a forwarding table-match when an FT entry for the target client is present in a forwarding table of the TEP. The non-transitory computer readable medium 904 includes instructions 914 that cause the processor 902 to transmit a unicast packet to all other TEPs with a status indicative of a forwarding table-miss when the FT entry for the target client is not present in the forwarding table of the TEP.

[0056] In an example implementation, the non-transitory computer readable medium 904 may further include instructions that cause the processor 902 to: receive from the destination TEP a response message indicating to delete the FT entry for the target client from the forwarding table of the TEP; and delete the FT entry of the target client from the forwarding table of the TEP.

[0057] In an example implementation, the non-transitory computer readable medium 904 may further include instructions that cause the processor 902 to: receive from a TEP currently associated with the target client, a response message indicating that a forwarding table of the currently associated TEP includes an FT entry for the target client, wherein the currently associated TEP is from the other TEPs that received the unicast packet with the status indicative of the forwarding table-miss; and add the FT entry for the target client in the forwarding table of the TEP.

[0058] In an example implementation, the non-transitory computer readable medium 904 may further include instructions that cause the processor

902 to: receive from the destination TEP a response message comprising information of a TEP currently associated with the target client; and update the FT entry for the target client in the forwarding table of the TEP based on the currently associated TEP.

[0059] Although implementations for client communication in multi-tenant data center networks have been described in language specific to structural features and/or methods, it is to be understood that the present subject matter is not limited to the specific features or methods described. Rather, the specific features and methods are disclosed and explained as example implementations for client communication in multi-tenant data center networks.

We claim:

1. A method of communication between clients in a multi-tenant data center network, the method comprising:

receiving at a source tunnel end-point (TEP) a data packet of a source client for a target client;

determining whether a forwarding table of the source TEP includes a forwarding table (FT) entry for the target client;

transmitting a unicast packet with a status indicative of a forwarding table-match to a destination TEP when the FT entry for the target client is present in the forwarding table of the source TEP; and

transmitting a unicast packet with a status indicative of a forwarding table-miss to all other TEPs when the FT entry for the target client is not present in the forwarding table of the source TEP.

2. The method as claimed in claim 1, comprising:

receiving from the destination TEP a response message indicating that a forwarding table of the destination TEP does not include an FT entry for the target client; and

deleting the FT entry for the target client from the forwarding table of the source TEP.

3. The method as claimed in claim 1, comprising:

receiving from a TEP currently associated with the target client, a response message indicating that a forwarding table of the currently associated TEP includes an FT entry for the target client, wherein the currently associated TEP is from the other TEPs that received the unicast packet with the status indicative of the forwarding table-miss; and

adding the FT entry for the target client in the forwarding table of the source TEP.

4. The method as claimed in claim 1, comprising:

receiving from the destination TEP a response message comprising information of a TEP currently associated with the target client; and

updating the FT entry for the target client in the forwarding table of the source TEP based on the currently associated TEP.

5. The method as claimed in claim 1, comprising:

setting a reserved bit in a packet header of the unicast packet to a first status to indicate the forwarding table-match; and

setting the reserved bit in the packet header of the unicast packet to a second status to indicate the forwarding table-miss.

6. The method as claimed in claim 1, comprising:

forwarding the unicast packet by the destination TEP to a TEP currently associated with the target client, wherein the unicast packet comprises a status indicative of relaying of the unicast packet by the destination TEP to the currently associated TEP.

7. A system for communication between clients in a multi-tenant data center network, the system comprising a tunnel end-point (TEP) to:

receive, from a source TEP, a unicast packet with data of a source client for a target client, wherein the unicast packet comprises a status indicative of one of a forwarding table-match and a forwarding table-miss at the source TEP; and

determine, based on the status, whether the unicast packet is for the forwarding table-match or the forwarding table-miss.

8. The system as claimed in claim 7, wherein, when the unicast packet is determined to be for the forwarding table-match, the TEP is to:

determine whether a forwarding table of the TEP includes a forwarding table (FT) entry for the target client; and

when the FT entry for the target client is not present in the forwarding table of the TEP, send a response message to the source TEP indicating to

delete an FT entry for the target client from a forwarding table of the source TEP.

9. The system as claimed in claim 7, wherein, when the unicast packet is determined to be for the forwarding table-match, the TEP is to:

determine whether a forwarding table of the TEP includes a forwarding table (FT) entry indicating a TEP currently associated with the target client;

forward the unicast packet to the currently associated TEP with a status indicative of relaying of the unicast packet to the currently associated TEP; and

send a response message to the source TEP indicating to update an FT entry for the target client in a forwarding table of the source TEP based on the currently associated TEP.

10. The system as claimed in claim 9, wherein the TEP is to set a reserved bit in a packet header of the unicast packet to a first status to indicate the relaying of the unicast packet.

11. The system as claimed in claim 9, wherein the TEP is to add in the response message a timestamp information indicative of a time at which the TEP learnt about the currently associated TEP.

12. The system as claimed in claim 7, wherein, when the unicast packet is determined to be for the forwarding table-miss, the TEP is to:

determine whether a forwarding table of the TEP includes a forwarding table (FT) entry for the target client; and

when the FT entry for the target client is present in the forwarding table of the TEP, send a response message to the source TEP indicating to add an FT entry for the target client in a forwarding table of the source TEP.

13. The system as claimed in claim 12, wherein the TEP is to add in the response message a timestamp information indicative of a time at which the TEP learnt about the target client.

14. A non-transitory computer-readable medium comprising computer-readable instructions for communication between clients in a multi-tenant data center network, the computer-readable instructions are executable by a processor:

- receive at a tunnel end-point (TEP) a data packet of a source client for a target client;

- transmit a unicast packet to a destination TEP with a status indicative of a forwarding table-match when a forwarding table (FT) entry for the target client is present in a forwarding table of the TEP; and

- transmit a unicast packet to all other TEPs with a status indicative of a forwarding table-miss when the FT entry for the target client is not present in the forwarding table of the TEP.

15. The non-transitory computer-readable medium as claimed in claim 14, comprising computer-readable instructions executable by the processor to:

- receive from the destination TEP a response message indicating to delete the FT entry for the target client from the forwarding table of the TEP; and

- delete the FT entry of the target client from the forwarding table of the TEP.

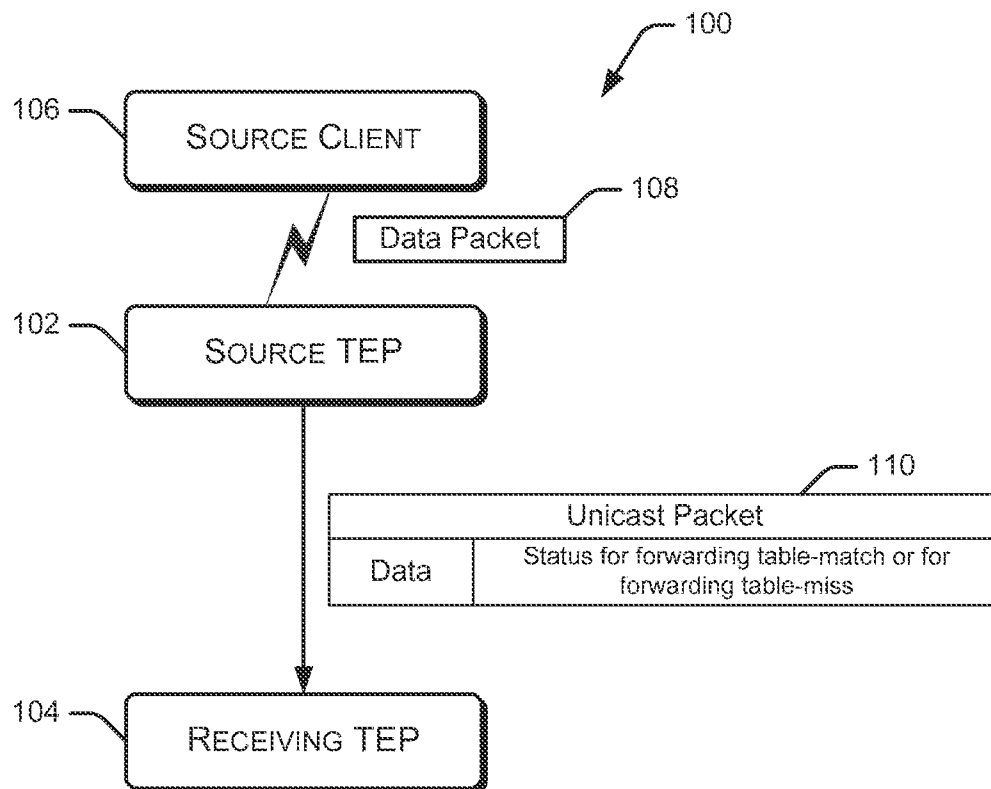
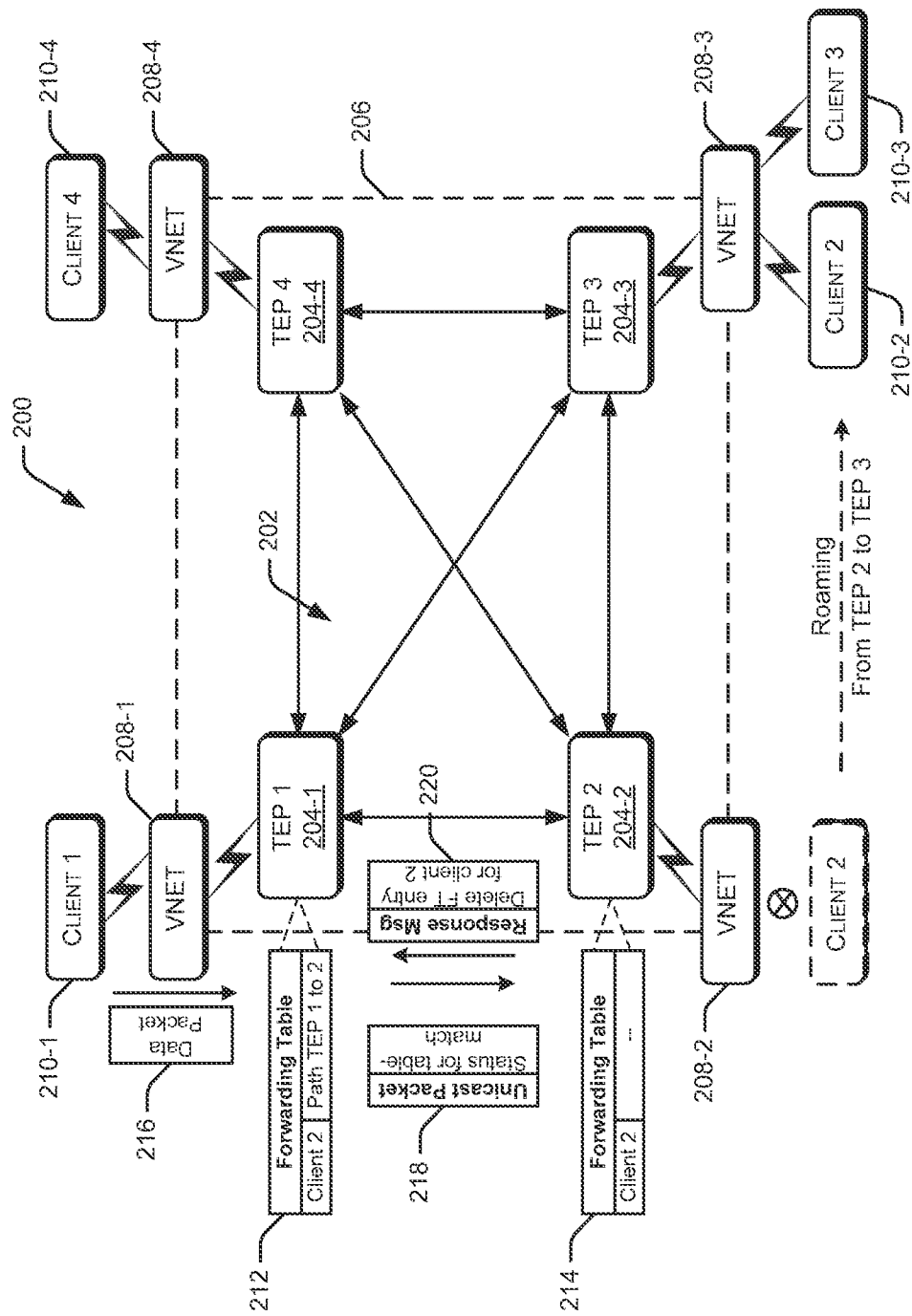


Fig. 1



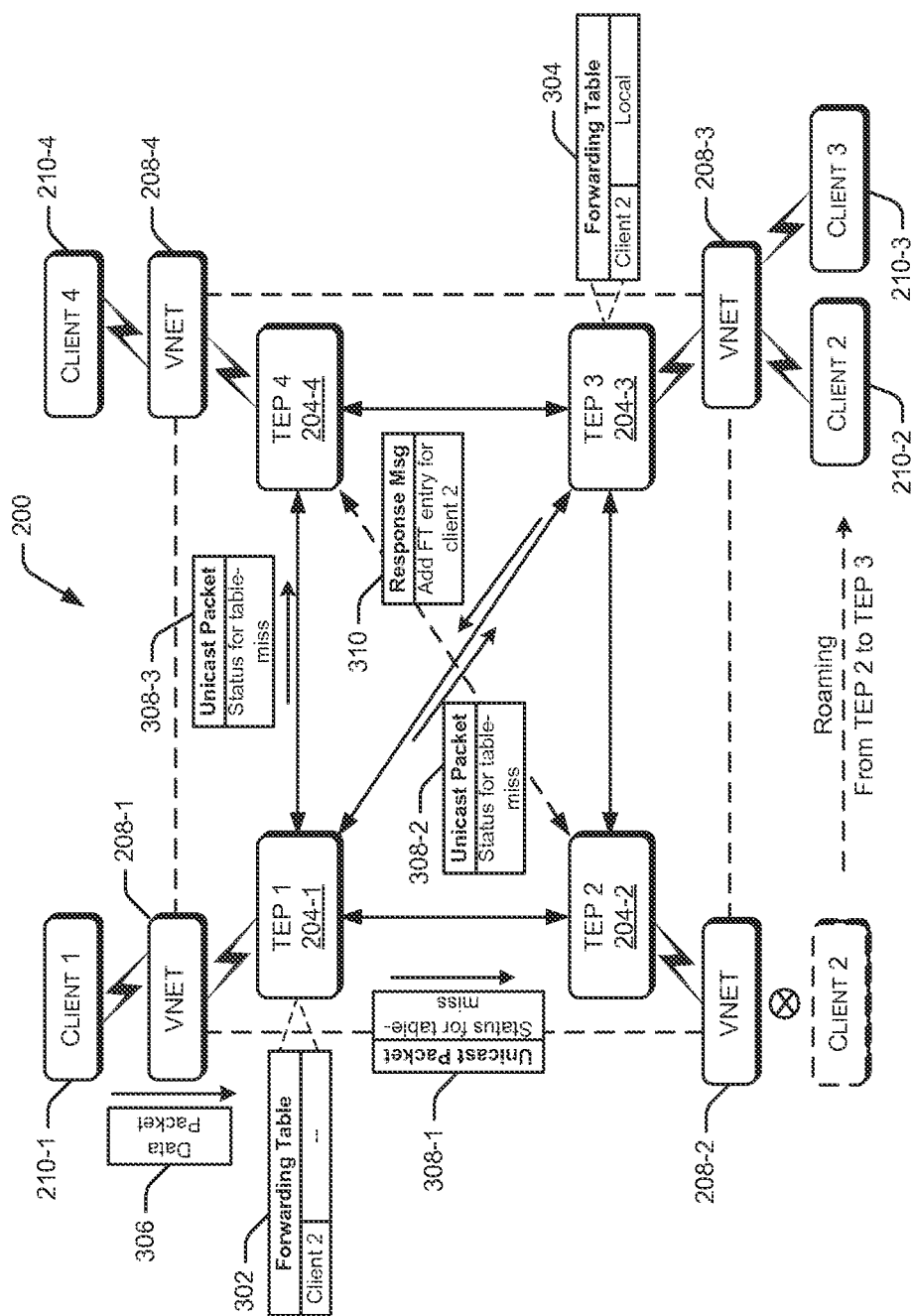


Fig. 3

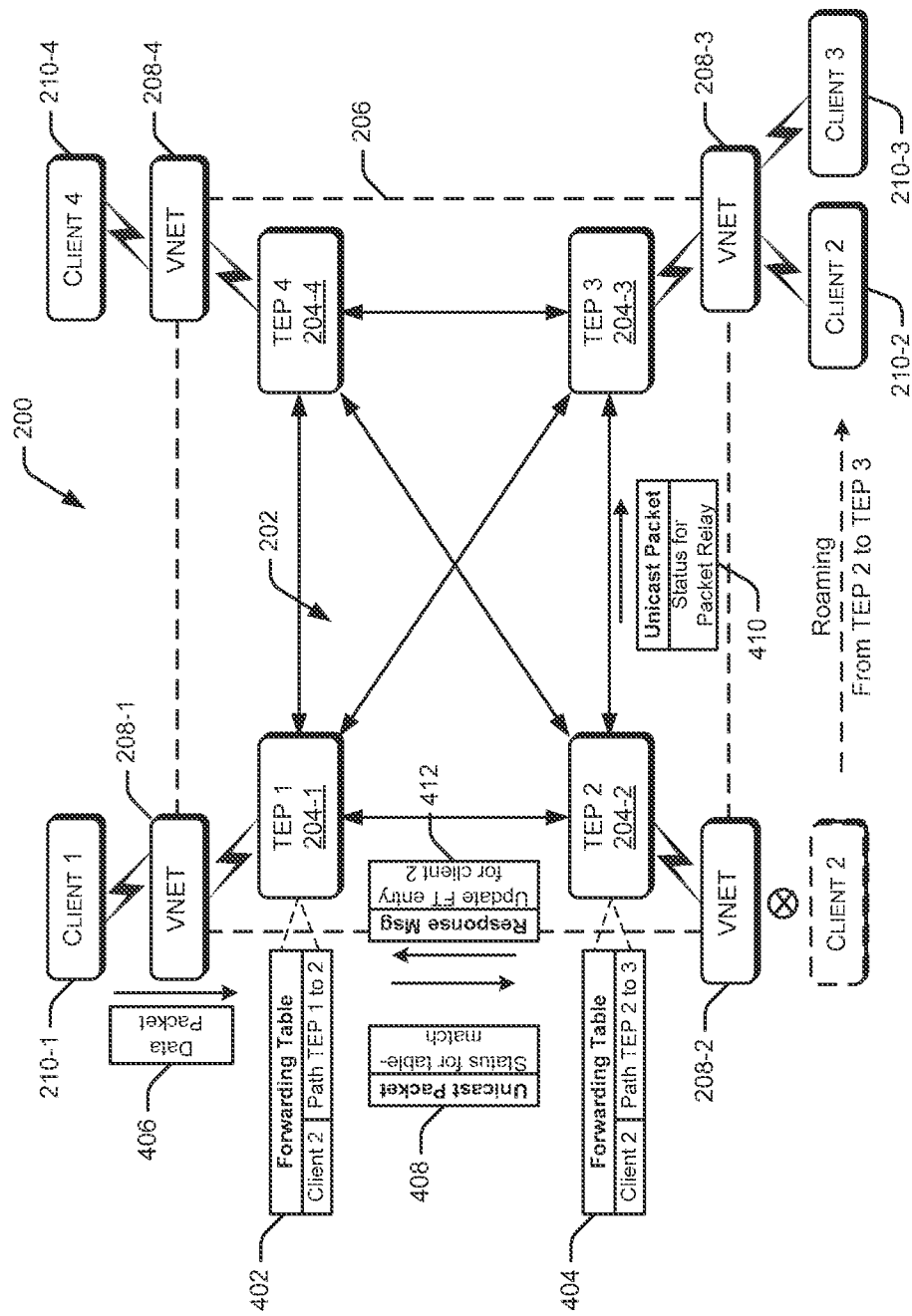
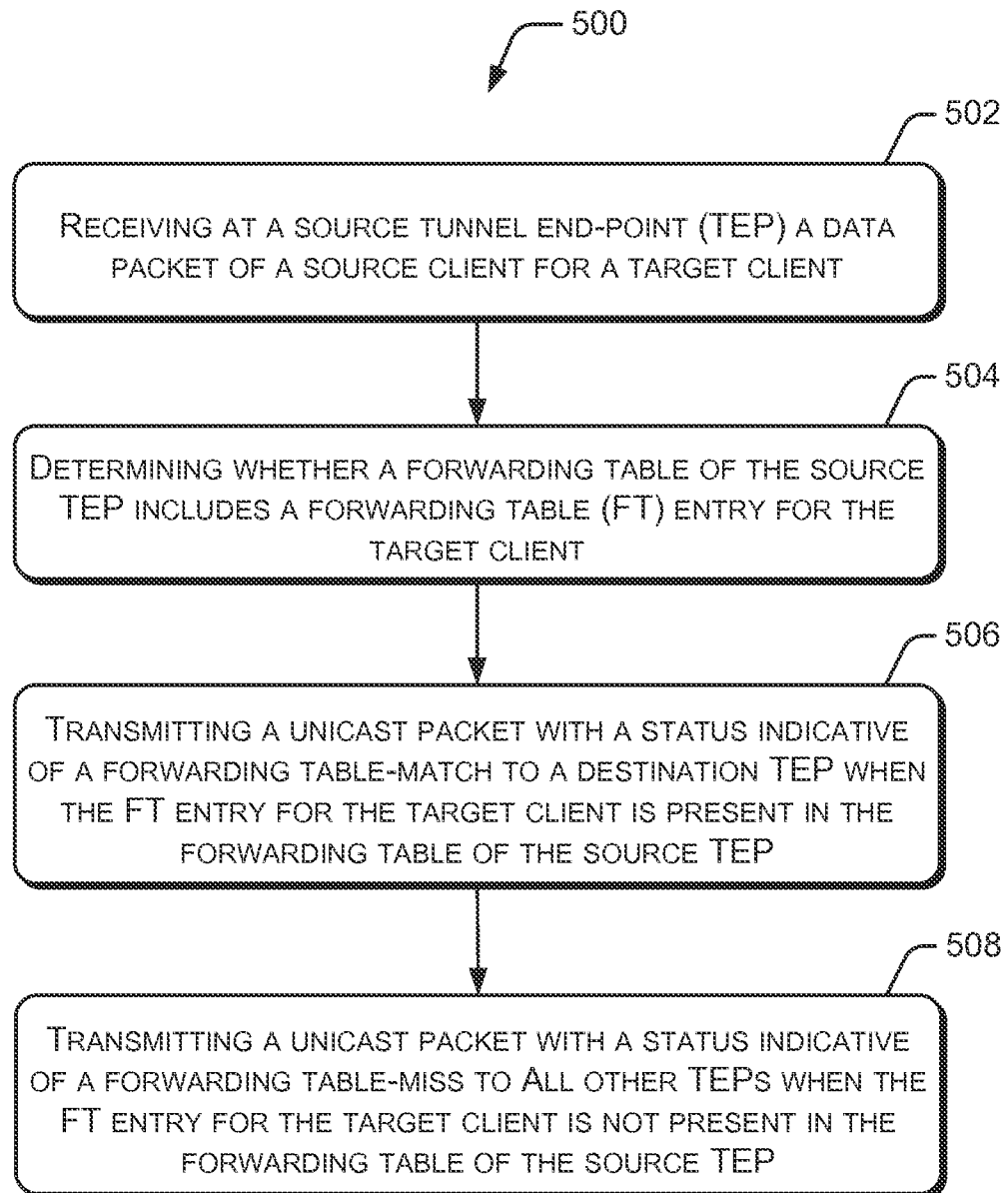
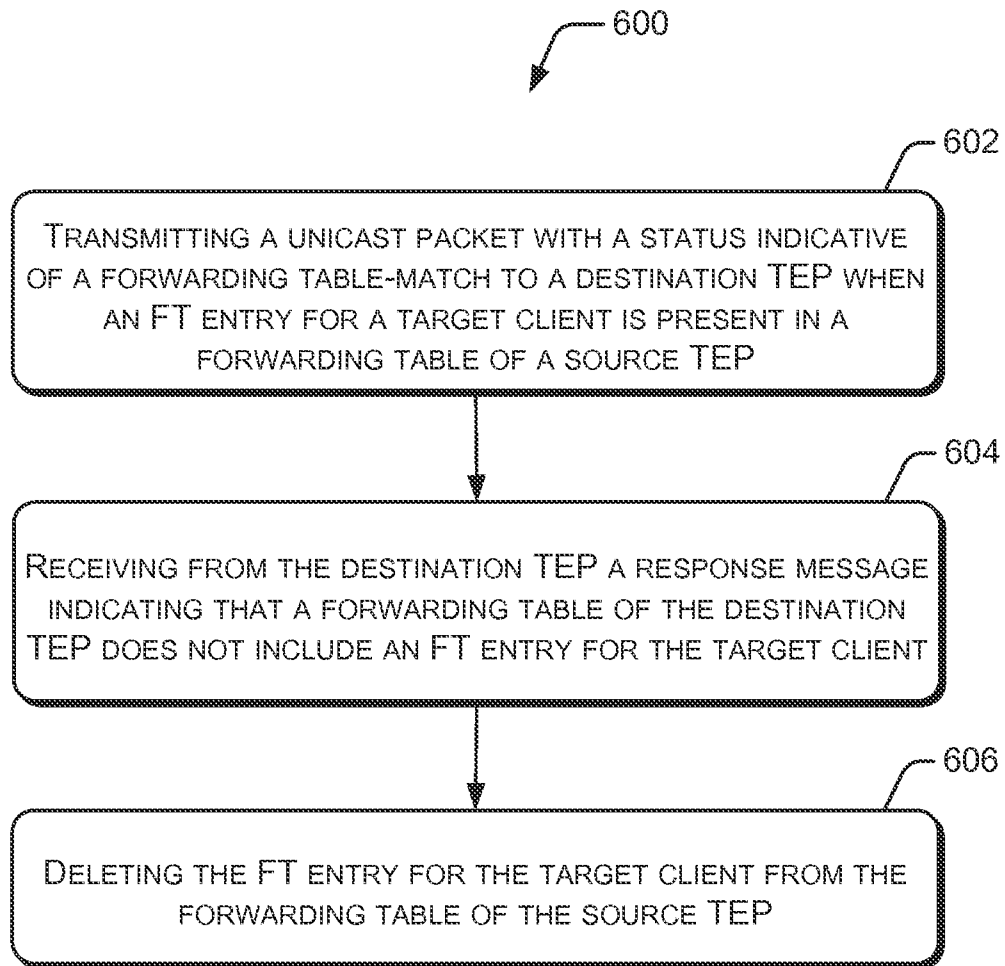
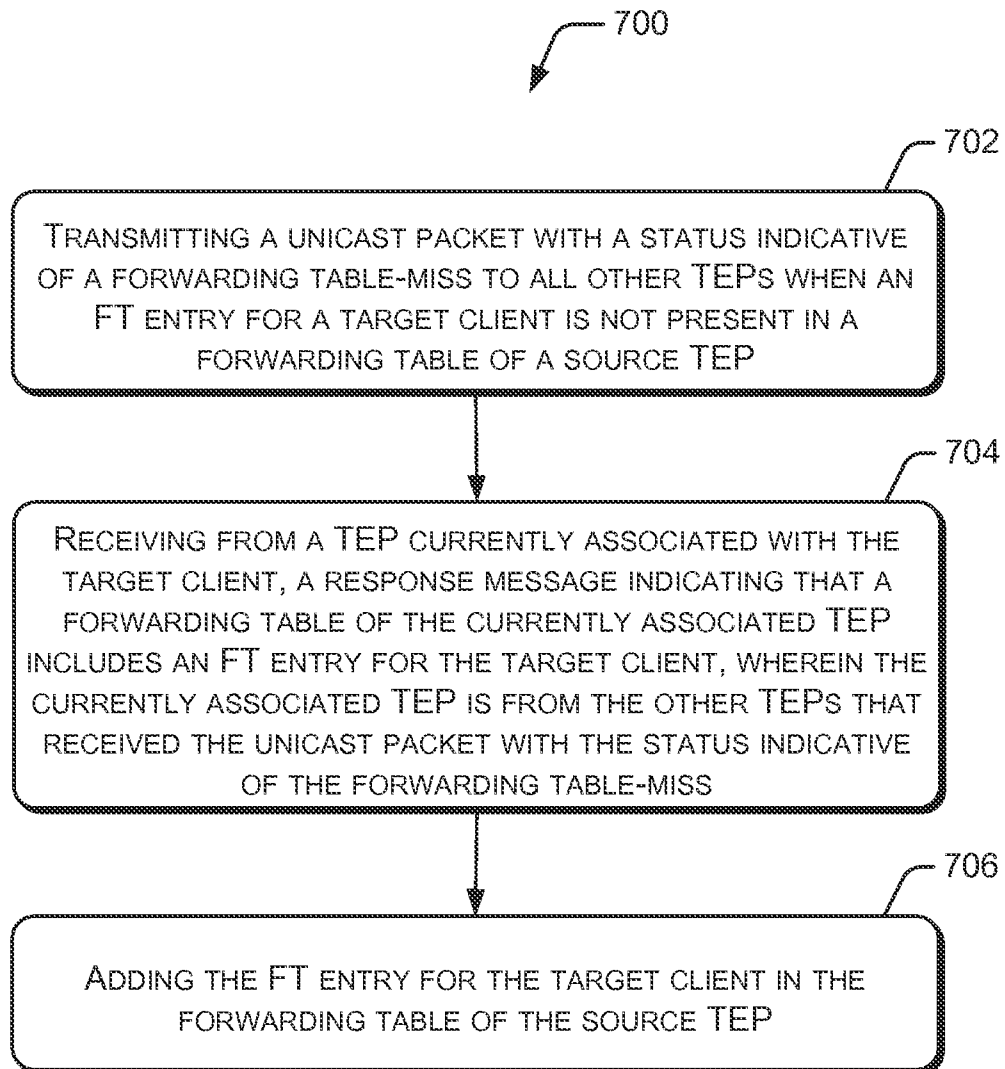
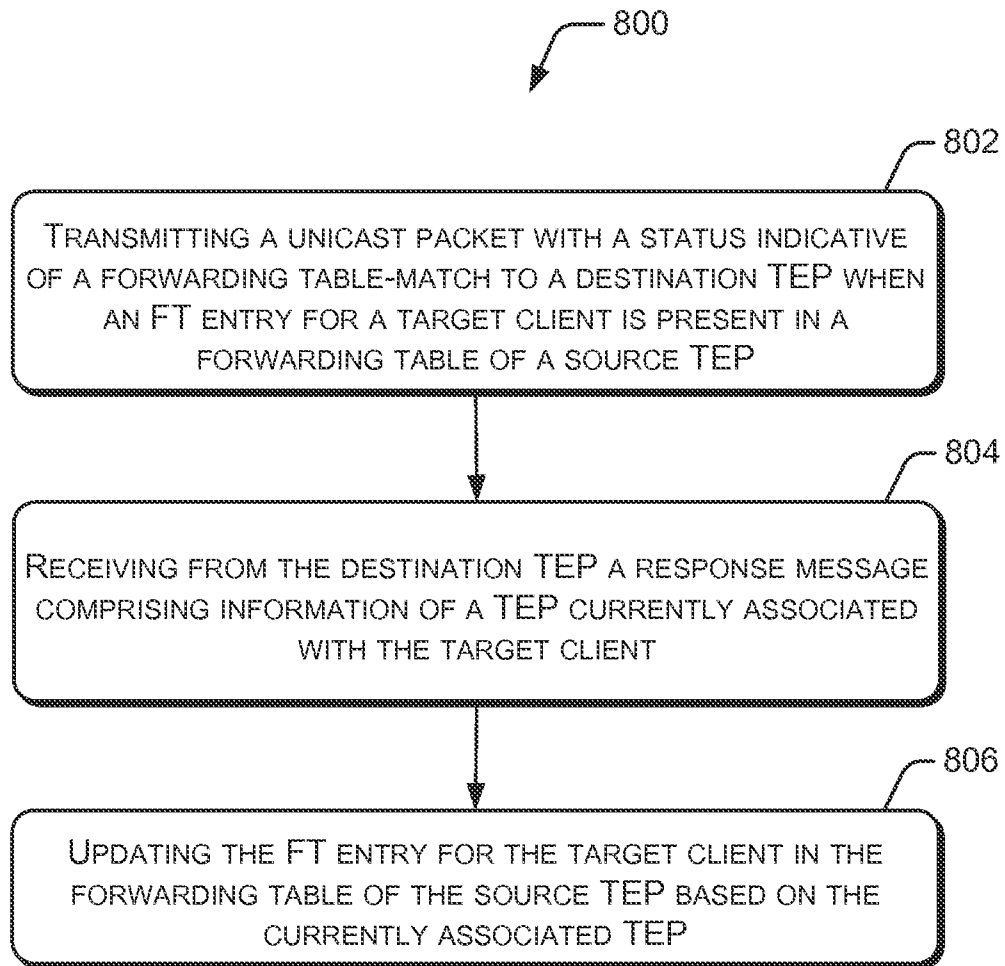


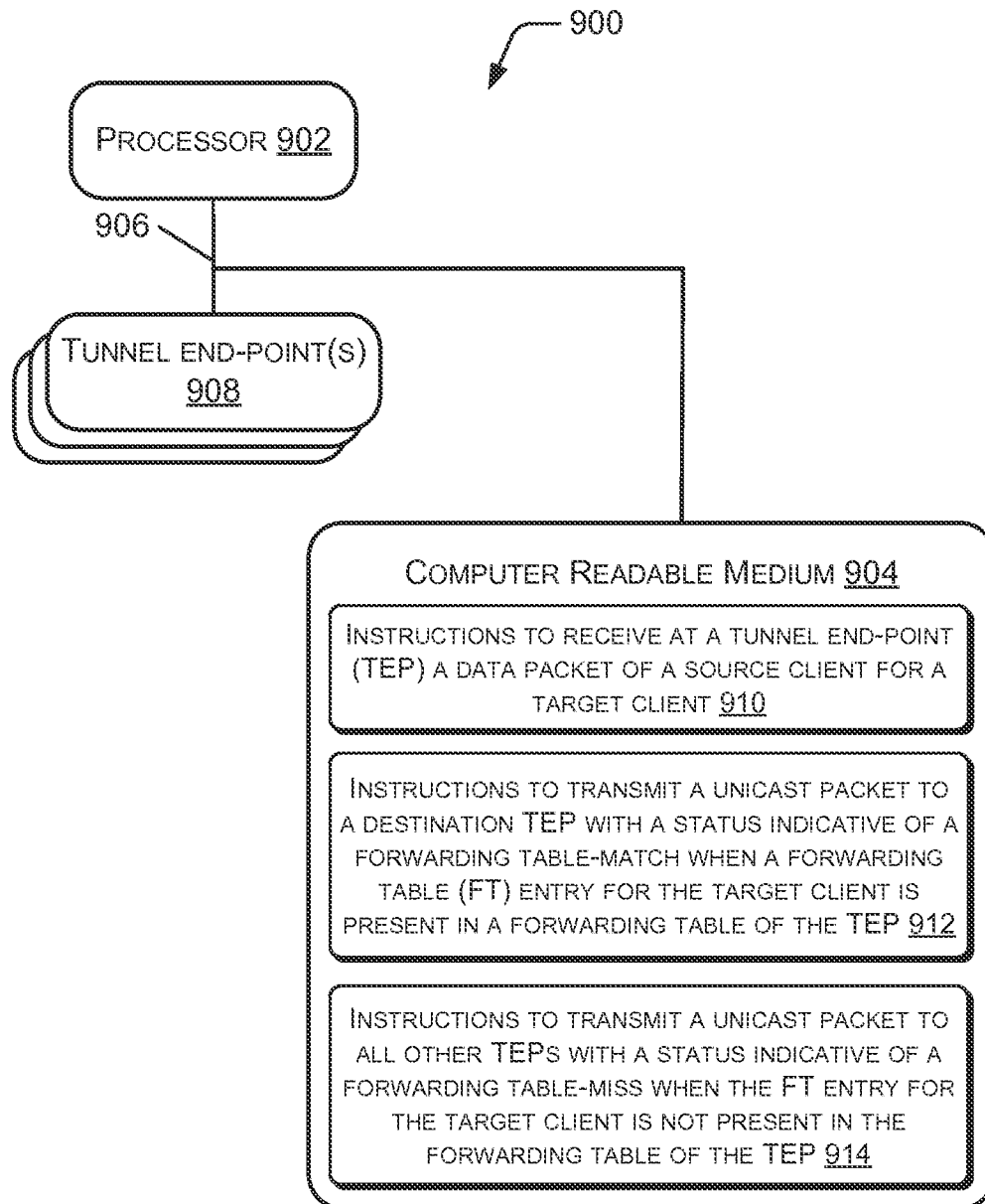
Fig. 4

**Fig. 5**

**Fig. 6**

**Fig. 7**

**Fig. 8**

**Fig. 9**

A. CLASSIFICATION OF SUBJECT MATTER**H04L 12/70(2013.01)i**

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 12/70; H04L 12/26; H04L 29/06; G06F 9/455; H04L 12/56; H04L 29/12

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords: forwarding table, match indication, tenant network

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014-0115584 A1 (JAYARAM MUDIGONDA et al.) 24 April 2014 See paragraphs [0041]-[0044], [0052]-[0058] and figures 5A-5B.	1-15
Y	US 2014-0173104 A1 (HUAWEI TECHNOLOGIES CO., LTD.) 19 June 2014 See paragraphs [0116]-[0118], [0122]-[0124], [0135]-[0136], claim 1 and figures 1-3.	1-15
A	WO 2014-040569 A1 (HANGZHOU H3C TECHNOLOGIES CO., LTD.) 20 March 2014 See paragraphs [0079]-[0083], claim 1 and figure 11.	1-15
A	US 2014-0064278 A1 (JOSE RENATO G. SANTOS et al.) 06 March 2014 See paragraphs [0049]-[0051] and figure 4.	1-15
A	US 2015-0033321 A1 (HANGZHOU H3C TECHNOLOGIES CO., LTD.) 29 January 2015 See paragraphs [0037]-[0058] and figure 3.	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

"E" earlier application or patent but published on or after the international filing date

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

23 January 2016 (23.01.2016)

Date of mailing of the international search report

18 February 2016 (18.02.2016)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

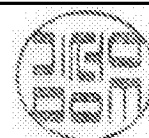
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

KIM, Seong Woo

Telephone No. +82-42-481-3348



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2015/036344

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014-0115584 A1	24/04/2014	CN 103563329 A EP 2719124 A1 EP 2719124 A4 WO 2012-170016 A1	05/02/2014 16/04/2014 03/12/2014 13/12/2012
US 2014-0173104 A1	19/06/2014	CA 2838215 A1 CN 103067199 A EP 2768197 A1 WO 2014-094432 A1	19/06/2014 24/04/2013 20/08/2014 26/06/2014
WO 2014-040569 A1	20/03/2014	CN 103685022 A US 2015-0146727 A1	26/03/2014 28/05/2015
US 2014-0064278 A1	06/03/2014	None	
US 2015-0033321 A1	29/01/2015	CN 102594678 A CN 102594678 B EP 2815546 A1 EP 2815546 A4 WO 2013-120406 A1	18/07/2012 14/01/2015 24/12/2014 07/10/2015 22/08/2013