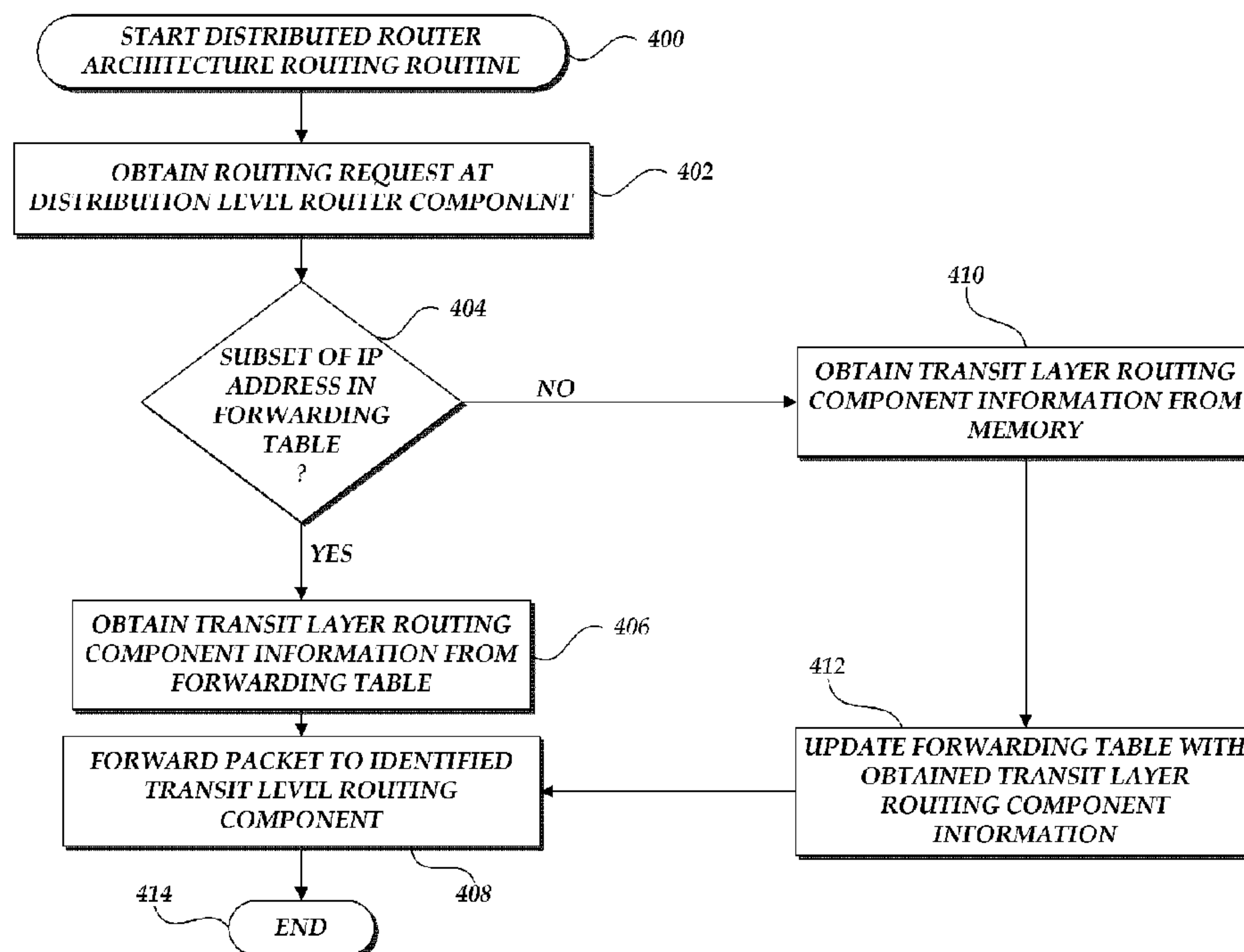




(86) Date de dépôt PCT/PCT Filing Date: 2010/12/15
(87) Date publication PCT/PCT Publication Date: 2011/07/14
(45) Date de délivrance/Issue Date: 2016/02/02
(85) Entrée phase nationale/National Entry: 2012/06/15
(86) N° demande PCT/PCT Application No.: US 2010/060567
(87) N° publication PCT/PCT Publication No.: 2011/084515
(30) Priorité/Priority: 2009/12/17 (US12/641,260)

(51) Cl.Int./Int.Cl. *H04L 12/715* (2013.01)
(72) Inventeurs/Inventors:
JUDGE, ALAN M., US;
MCGAUGH, DAVID J., US;
HAMILTON, JAMES R., US;
PIETSCH, JUSTIN O., US;
O'MEARA, DAVID J., US
(73) Propriétaire/Owner:
AMAZON TECHNOLOGIES, INC., US
(74) Agent: SMART & BIGGAR

(54) Titre : ARCHITECTURE DE ROUTAGE DISTRIBUE
(54) Title: DISTRIBUTED ROUTING ARCHITECTURE



(57) Abrégé/Abstract:

A hierarchical distributed routing architecture including at least three levels, or layers, for receiving, processing and forwarding data packets between network components is provided. The core level router components receive an incoming packet from a network component and identify a distribution level router component based on processing a subset of the destination address associated with the received packet. The distribution level router components that receiving a forwarded packet and identify a transit level router component based a second processing of at least a subset of the destination address associated with the received packet. The transit level router components receive the forwarded packet and forward the packet to a respective network. The mapping, or other assignment, of portions of the FIB associated with the distributed routing environment is managed by a router management component.

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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
14 July 2011 (14.07.2011)(10) International Publication Number
WO 2011/084515 A1(51) International Patent Classification:
H04L 12/56 (2006.01)(21) International Application Number:
PCT/US2010/060567(22) International Filing Date:
15 December 2010 (15.12.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/641,260 17 December 2009 (17.12.2009) US(71) Applicant (for all designated States except US): **AMAZON TECHNOLOGIES, INC.** [US/US]; P.o. Box 8102, Reno, NV 89507 (US).(72) Inventors: **JUDGE, Alan, M.**; 1200 12th Avenue South, Suite 1200, Seattle, WA 98144-2734 (US). **MCGAUGH, David, J.**; 1200 12th Avenue South, Suite 1200, Seattle, WA 98144-2734 (US). **HAMILTON, James, R.**; 1200 12th Avenue South, Suite 1200, Seattle, WA 98144-2734 (US). **PIETSCH, Justin, O.**; 1200 12th Avenue South, Suite 1200, Seattle, WA 98144-2734 (US). **O'MEARA, David, J.**; 1200 12th Avenue South, Suite 1200, Seattle, WA 98144-2734 (US).(74) Agent: **URIBE, Mauricio A.**; Knobbe Martens Olson & Bear, LLP, 2040 Main Street, Fourteenth Floor, Irvine, CA 92614 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: DISTRIBUTED ROUTING ARCHITECTURE

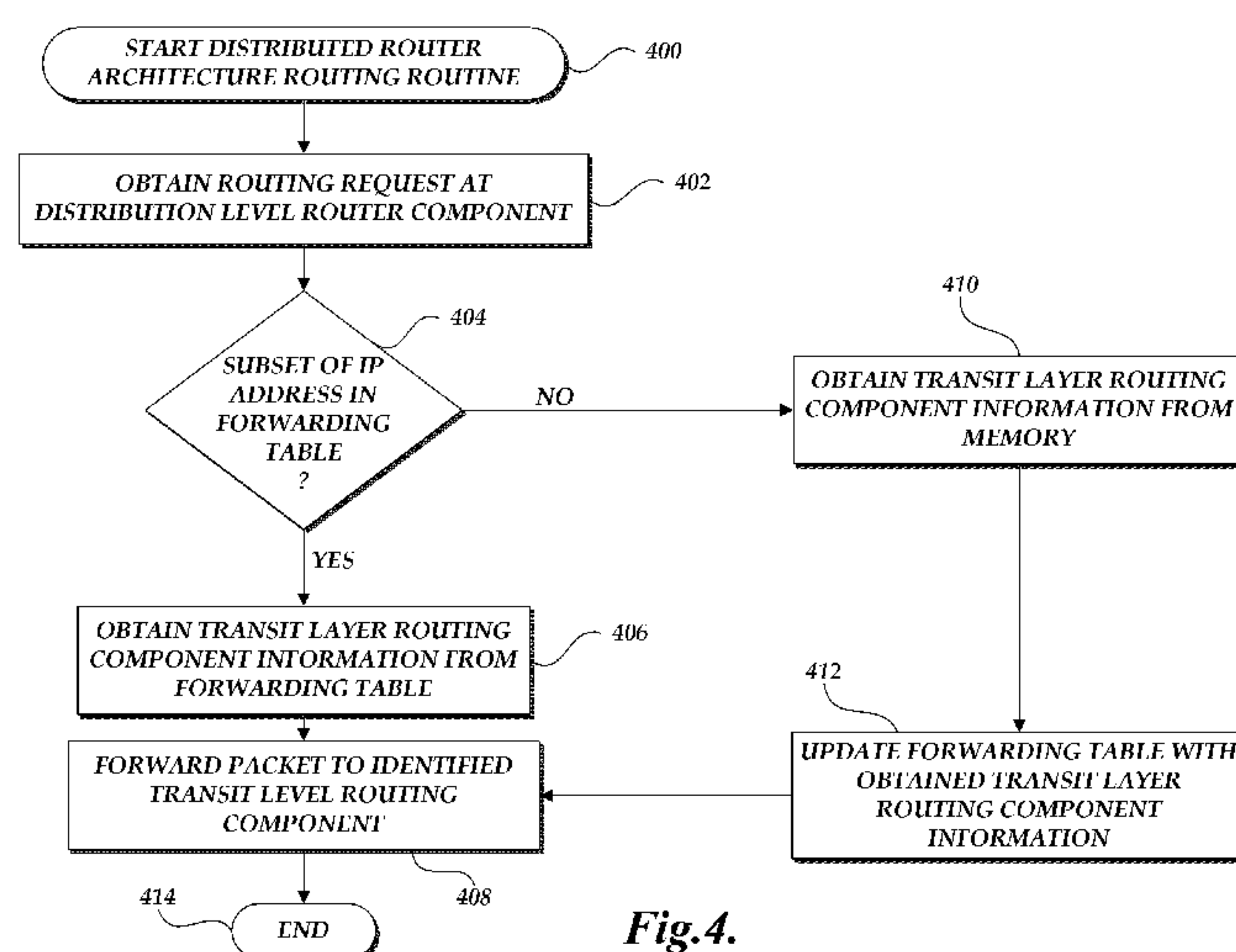


Fig. 4.

(57) **Abstract:** A hierarchical distributed routing architecture including at least three levels, or layers, for receiving, processing and forwarding data packets between network components is provided. The core level router components receive an incoming packet from a network component and identify a distribution level router component based on processing a subset of the destination address associated with the received packet. The distribution level router components that receiving a forwarded packet and identify a transit level router component based a second processing of at least a subset of the destination address associated with the received packet. The transit level router components receive the forwarded packet and forward the packet to a respective network. The mapping, or other assignment, of portions of the FIB associated with the distributed routing environment is managed by a router management component.

WO 2011/084515 A1

DISTRIBUTED ROUTING ARCHITECTURE

BACKGROUND

[0001] Generally described, computing devices utilize a communication network, or a series of communication networks, to exchange data. In a common embodiment, data to be exchanged is divided into a series of packets that can be transmitted between a sending computing device and a recipient computing device. In general, each packet can be considered to include two primary components, namely, control information and payload data. The control information corresponds to information utilized by one or more communication networks to deliver the payload data. For example, control information can include source and destination network addresses, error detection codes, and packet sequencing identification, and the like. Typically, control information is found in packet headers and trailers included within the packet and adjacent to the payload data.

[0002] In practice, in a packet-switched communication network, packets are transmitted between multiple physical networks, or sub-networks. Generally, the physical networks include a number of hardware devices that receive packets from a source network component and forward the packet to a recipient network component. The packet routing hardware devices are typically referred to as routers. Generally described, routers can operate with two primary functions or planes. The first function corresponds to a control plane, in which the router learns the set of outgoing interfaces that are most appropriate for forwarding received packets to specific destinations. The second function is a forwarding plane, in which the router sends the received packet to an outbound interface.

[0003] To execute the control plane functionality, routers can maintain a forwarding information base ("FIB") that identifies, among other packet attribute information, destination information for at least a subset of possible network addresses, such as Internet Protocol ("IP") addresses. In a typical embodiment, the FIB corresponds to a table of values specifying network forwarding information for the router. In one aspect, commercial level routing hardware components can include customized chipsets, memory components, and software that allows a single router to support millions of entries in the FIB. However, such commercial level routing hardware components are typically very expensive and often require extensive customization. In

another aspect, commodity-based routing hardware components are made of more generic components and can be less expensive than commercial level routing hardware components by a significant order of magnitude. However, such commodity-based routing hardware components typically only support FIBs on the order of thousands of entries.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

[0005] FIGURE 1A is a block diagram illustrative of one embodiment of a distributed routing environment including a router management component and a hierarchical, distributed routing component architecture;

[0006] FIGURE 1B is a block diagram illustrative of components of a router component utilized in accordance with the distributed routing environment of FIGURE 1A;

[0007] FIGURES 2A-2C are block diagrams illustrative of the distributed routing environment of FIGURE 1A illustrating the routing of a received packet within the hierarchical distributed routing component architecture;

[0008] FIGURE 3 is a flow diagram illustrative of a distributed router architecture routing routine implemented within a distributed routing environment; and

[0009] FIGURE 4 is a flow diagram illustrative of a distributed router architecture routing routine implemented within a distributed routing environment

DETAILED DESCRIPTION

[0010] Generally described, the present disclosure corresponds to a distributed routing architecture. Specifically, the present disclosure corresponds to a hierarchical distributed routing architecture including at least three logical levels, or layers, for receiving, processing and forwarding data packets between network components. In one embodiment, the three logical levels can correspond to a core level, a distribution level and a transit level. Illustratively, the core level corresponds to one or more router components that receive an incoming packet from a network component and processes the destination address information associated with the received packet. The core level router component then identifies a distribution level router

component based on a subset of the destination address associated with the received packet. The distribution level corresponds to one or more router components that receive a forwarded packet from a core level router component and further processes the destination address information associated with the received packet. The distribution level router component identifies a transit level router component based on at least a subset of the destination address associated with the received packet. Each distribution level router component is associated with, or otherwise corresponds to, a subset of the FIB associated with the distributed routing architecture. Finally, the transit level router components correspond to one or more router components that receive the forwarded packet from a distribution level router component and forward the packet “upstream” to a respective network, or network node. The mapping, or other assignment, of portions of the FIB associated with the distributed routing environment is managed by a router management component.

[0011] In one embodiment, each of the router components associated with the core level, distribution level and transit level can correspond more closely to commodity based router components/hardware. In another embodiment, the core level, distribution level and transit level router components correspond to logical router components that do not necessarily have a corresponding hardware router component. For example, one or more logical router components within each level may be implemented in the same hardware router component. Likewise, the logical router components associated with different levels of the distributed routing architecture may be implemented in the same hardware router component. In both embodiments, however, because responsibility for maintaining the FIB associated with the distributed routing environment is divided among several router components, the processing and memory restraints associated with commodity based router components/hardware can be mitigated. Various implementations, combination, and applications for dividing the FIB associated with the distributed routing environment will be described in accordance with the distributed routing environment. However, one skilled in the relevant art will appreciate that such embodiment and examples are illustrative in nature and should not be construed as limiting.

[0011a] In accordance with one disclosed aspect there is provided a system for routing packets. The system includes a router management component, executed on a computing device, for associating destination address information to a router hierarchy includes a plurality of levels, the router management component is configured to, for individual levels of the router hierarchy, dynamically allocate responsibility for portions of destination addresses of incoming packets to one or more router components of the individual level of the router hierarchy/ The system also includes one or more router components corresponding to a first level of the router hierarchy, the first level of the router hierarchy for receiving the incoming packets for routing and transmitting the incoming packets to router components of a second level of the router hierarchy. The system further includes one or more router components corresponding to the second level of the hierarchy, the second level of the router hierarchy for routing the incoming packets received from the first level of a router hierarchy to router components of a third level of the router hierarchy. At least one router component corresponding to the second level is dynamically allocated responsibility for a portion of a first subset of each destination address of each incoming packet by the router management component based at least in part on the incoming packets. The system further includes one or more router components corresponding to the third level of the router hierarchy, the third level of the router hierarchy for routing the incoming packets received from the second level of the router hierarchy, at least one router component corresponding to the third level is dynamically allocated responsibility for a portion of a second subset of each destination address of each incoming packet by the router management component based at least in part on the incoming packets. The one or more router components corresponding to the first level of the router hierarchy identify one or more router components from the second level of the router hierarchy to which to route each packet of the incoming packets based at least in part on the first subset of each destination address associated with each packet, the first subset is designated by the router management component. The one or more router components corresponding to the second level of the router hierarchy identify one or more router components of the third level of the router hierarchy to which to route each packet of the incoming packets based at least in part on the second subset of each destination address associated with each incoming packet, the second subset is designated

by the router management component and the second subset is greater than the first subset.

[0011b] In accordance with another disclosed aspect there is provided a system for routing packets. The system includes a first set of logical router components for receiving an incoming set of packets for routing, a second set of logical router components for routing the set of packets received from the first set of router components, a third set of logical router components for routing the set of packets received from the second set of router components. The first set of logical router components identify a router component from the second set of logical router components to which to route a first packet of the set of packets based at least in part on a dynamic assignment by a router management component of portions of a first subset of destination addresses associated with the incoming set of packets to individual router components from the second set of logical router components. Each of the one or more router components corresponding to the second set of logical router components is dynamically assigned by the router management component to the portions of the first subset of the destination addresses associated with the incoming set of packets. The dynamic assignment is based at least in part on the set of packets. The second set of logical router components identify a router component from the third set of logical router components to which to route the first packet based at least in part on a dynamic assignment by the router management component of portions of a second subset of destination addresses associated with the incoming set of packets to individual router components from the third set of logical router components.

[0011c] In accordance with another disclosed aspect there is provided a method for routing packets. The method involves obtaining a routing request corresponding to a data packet received from a first communication network, identifying a first router corresponding to a first level of a router hierarchy, the first level of the router hierarchy corresponding to one or more router components, forwarding the received data packet to the identified first router, identifying a second router corresponding to a second level of the router hierarchy, the second level of the router hierarchy corresponding to one or more router components. The method also involves forwarding the received data packet to the identified second router, and identifying a third router corresponding to a third

level of the router hierarchy, the third level of the router hierarchy corresponding to one or more router components. Identifying the second router is based at least in part on a dynamic assignment of the second router to a portion of a first subset of a destination address associated with the received data packet, and the dynamic assignment is based at least in part on a set of previously received data packets, and identifying a third router is based at least in part on a dynamic assignment of the third router to a portion of a second subset of the destination address associated with the received data packet, and the dynamic assignment is based at least in part on a set of previously received data packet.

[0012] Turning now to FIGURE 1A, a distributed routing environment 100 for implemented a hierarchical distributed routing architecture will be described. The distributed routing environment 100 includes a router management component 102 for controlling the

routing information utilized by the distributed routing environment 100. Specifically, the router managed component 102 can receive all upstream routing information to be used by the distributed routing environment 100 and allocate the assignment of the upstream routing information among the components of the distributed routing environment 100 as will be described. In one embodiment, the router management component 102 can correspond to a computing device in communication with one or more components of the distributed routing environment 100. Illustrative computing devices can include server computing devices, personal computing devices or other computing devices that include a processor, memory and other components for executing instructions associated with the function of the router management component 102. In another embodiment, the router management component 102 may be implemented as a software component that is executed on one or more of the router components described below. Illustratively, the router management component 102 maintains and updates the FIB associated with the distributed routing environment 100. Additionally, the router management component 102 can allocate responsibility for portions of the FIB entries to the various layers of the distributed routing environment 100, as will be described below. In one embodiment, the router management component 102 can partition the FIB according to the distribution to the various router components of the distributed routing environment 100 and distribute respective portions of the FIB to be maintained in a memory associated with the various router components.

[0013] With continued reference to FIGURE 1A, the distributed routing environment 100 includes a first communication network 104 that transmits data packets to the distributed routing environment 100. The first communication network 104 may encompass any suitable combination of networking hardware and protocols necessary to establish packet-based communications to the distributed routing environment 100. For example, the communication network 104 may include private networks such as local area networks (LANs) or wide area networks (WANs) as well as public or private wireless networks. In such an embodiment, the communication network 104 may include the hardware (e.g., modems, routers, switches, load balancers, proxy servers, etc.) and software (e.g., protocol stacks, accounting software, firewall/security software, etc.) necessary to establish a networking link with the distributed routing environment 100. Additionally, the communication network 104 may implement one of

various communication protocols for transmitting data between computing devices. As will be explained in greater detail below, the communication protocols can include protocols that define packet flow information, such as network address information corresponding to the Internet Protocol version 4 (IPv4) and the Internet Protocol version 6 (IPv6) Internet Layer communication network protocols. One skilled in the relevant art will appreciate, however, that present disclosure may be applicable with additional or alternative protocols and that the illustrated examples should not be construed as limiting.

[0014] In communication with the first communication network 104 is a first level of the distributed routing environment 100, generally referred to as the core layer or core level. In one embodiment, the core level corresponds to one or more logical router components, generally referred to as core level routers 106A, 106B, and 106C. As previously described, within the distributed routing environment 100, the core level routers 106A, 106B, 106C receive an incoming packet from a component from the network 104 and process the destination address by identifying a distribution level router component based on a subset of the destination address associated with the received packet. Illustratively, the subset of the destination address can correspond to less than the entire destination IP address, such as the highest most values of the IP address. As previously described, the core level routers 106A, 106B, 106C can correspond to logical router components implemented on one or more hardware components. In one embodiment, each logical router component can correspond with a dedicated physical router component. In another embodiment, each logical router component can correspond to a physical router component shared by at least one other logical router component in the distributed router environment 100. In an alternative embodiment, at least some portion of the core layer may be implemented by components outside the distributed routing environment 100. In such an embodiment, such external components would directly address a distribution level router component (described below) of the distributed routing environment 100.

[0015] The distributed routing environment 100 can further include a second level of logical router components, generally referred to as the distribution layer or distribution level. In one embodiment, the distribution level corresponds to one or more router components, generally referred to as distribution level routers 108A, 108B, and 108C. As previously described, within the distributed routing environment 100 the distribution level routers 108A, 108B and 108C

receiving an incoming packet from a core routing component 102 and process the destination address by identifying a transit level router component based on at least a subset of the destination address associated with the received packet. Illustratively, the subset of the destination address can correspond to a larger subset of the destination IP address used by the core level routers 106A, 106B, 106C. In this embodiment, the routing performed by the distribution level can correspond to a more refined routing of the received packet relative to the core level routing. As described above with the core level routers 106A, 106B, 106C, the distribution level routers 108A, 108B, and 108C can correspond to logical router components implemented on one or more hardware components. In one embodiment, each logical router component can correspond with a dedicated physical router component. In another embodiment, each logical router component can correspond to a physical router component shared by at least one other logical router component in the distributed router environment 100.

[0016] In communication with the distribution level router components is a third level of router components, generally referred to as the transit layer or transit level. In one embodiment, the transit level corresponds to one or more router components, generally referred to as transit level routers 110A, 110B, and 110C. As previously described, the transit level routers 110A, 110B, 110C receive the forwarded packet from a distribution level router component 108A, 108B, 108C and forward the packet “upstream” to another communication network 112 node. Illustratively, each transit level router 110A, 110B, 110C can be configured to communicate with one or more upstream peers such that all packets destined for an associated peer network component will be transmitted through the assigned transit level router 110A, 110B, 110C (or a redundant router). As described above with the core level routers 106A, 106B, 106C and the distribution level routers 108A, 108B and 108C, the transit level routers 110A, 110B, and 110C can correspond to logical router components implemented on one or more hardware components. In one embodiment, each logical router component can correspond with a dedicated physical router component. In another embodiment, each logical router component can correspond to a physical router component shared by at least one other logical router component in the distributed router environment 100.

[0017] Similar to communication network 102, communication network 112 may encompass any suitable combination of networking hardware and protocols necessary to establish

packet-based communications to the distributed routing environment 100. For example, the communication network 112 may include private networks such as local area networks (LANs) or wide area networks (WANs) as well as public or private wireless networks. In such an embodiment, the communication network 112 may include the hardware (e.g., modems, routers, switches, load balancers, proxy servers, etc.) and software (e.g., protocol stacks, accounting software, firewall/security software, etc.) necessary to establish a networking link with the distributed routing environment 100. As described above with regard to the communication network 104, the communication network 112 may implement one of various communication protocols for transmitting data between computing devices. One skilled in the relevant art will appreciate, however, that present disclosure may be applicable with additional or alternative protocols and that the illustrated examples should not be construed as limiting.

[0018] In an illustrative embodiment, the logical router components (106, 108, 110) in FIGURE 1A may correspond to a computing device having processing resources, memory resources, networking interfaces, and other hardware/software for carrying the described functionality for each of the logical router components. With reference now to FIGURE 1B, a block diagram illustrative of components of a router component 150 utilized in accordance with the distributed routing environment 100 of FIGURE 1A will be described. The general architecture of the router component 150 depicted in FIGURE 1B includes an arrangement of computer hardware and software components that may be used to implement one or more logical router components 106, 108, 110. Those skilled in the art will appreciate that the router component 150 may include many more (or fewer) components than those shown in FIGURE 1B. It is not necessary, however, that all of these generally conventional components be shown in order to provide an enabling disclosure.

[0019] As illustrated in FIGURE 1B, the router component 150 includes a processing unit 152, at least one network interface 156, and at least one computer readable medium drive 158, all of which may communicate with one another by way of a communication bus. The processing unit 152 may thus receive information and instructions from other computing systems or services via a network. The processing unit 152 may also be associated with a first memory component 154 for recalling information utilized in the processing of destination address information, such as at least a portion of a FIB associated with the distributed routing

environment 100. The memory 154 generally includes RAM, ROM and/or other persistent memory. The processing unit 152 may also communicate to and from memory 160. The network interface 156 may provide connectivity to one or more networks or computing systems. The at least one computer readable medium drive 158 can also correspond to RAM, ROM, optical memory, and/or other persistent memory that may persists at least a portion of the FIB associated with the distributed routing environment 100. In an illustrative embodiment, the access time associated with the memory component 154 may be faster than the access time associated with the computer readable medium driver 158. Still further, the computer readable medium drive 158 may be implemented in a networked environment in which multiple router components 150 share access to the information persisted on the computer readable medium drive 158.

[0020] The memory 160 contains computer program instructions that the processing unit 152 executes in order to operate the dynamic classifier. The memory 160 generally includes RAM, ROM and/or other persistent memory. The memory 160 may store an operating system 162 that provides computer program instructions for use by the processing unit 152 in the general administration and operation of the router component 150. The memory 160 may further include computer program instructions and other information for implementing one or more of the logical router components in the distributed routing environment 100. For example, in one embodiment, the memory 160 includes a router module 164 that implements the functionality associated with any of the routers 106, 108, 110. In the event that multiple logical routers are implemented by the same router component 150, memory 160 may have each instance of a router module 164.

[0021] In an illustrative embodiment, each router component 150 may be embodied as an individual hardware component for implementing one or more logical routers 106, 108, 110. Alternatively, multiple router components 150 may be grouped and implemented together. For example, each router component 150 may correspond to an application-specific integrated circuit (ASIC) having a processing unit 152, memory 154 and memory 160 (or other components with similar functionality). The router components 150 may share one or more components, such as the network interface 156 and computer readable medium 158, via a common communication bus.

[0022] With reference now to FIGURES 2A-2C, the processing of receiving packets by the distributed routing environment 100 will be described. With reference first to FIGURE 2A, an incoming packet is received from the communication network 104 to a core level router 106. The core level router 106 that receives the incoming packet may be selected according to a variety of techniques including, but not limited to, load balancing, random selection, round robin, hashing, and other packet distribution techniques. Upon receipt, the core level router 106 processes destination IP address and utilizes a subset of the destination IP address to identify a second level destination router component that will perform a second level of routing. In an illustrative embodiment, the core level router 106 utilizes the most significant bits of the IP address, such as the eight most significant bits of the destination address. The selection of the subset of IP addresses corresponding to a selection of the most significant bits is generally referred to as prefix. For example, selection of the eight most significant bits corresponds to a prefix length of "8." Selection of the sixteen most significant bits corresponds to a prefix length of "16." One skilled in the relevant art will appreciate that the number of bits utilized by the core level router 106 may vary. Additionally, in an alternative embodiment, the core level router 106 may use different methodologies to allocate, or otherwise subdivide, the address space serviced by the distributed routing environment 100.

[0023] Based on the processing of the first subset of the destination address, the core level router 106 forwards the packet to a distribution level router, in this case illustratively 108A. As previously described, the receiving distribution level router 108A processes the destination address of the received packet and also utilizes a subset of the destination IP address to identify a third level router component that will forward the packet to a next network destination (outside of the distributed routing environment 100). Similar to the core level router 106, the receiving distribution level router can be configured to utilize a selection of the most significant bits of the IP address (e.g., the prefix) to route the packet. In an illustrative embodiment, the prefix used by the distribution level router 108A is greater than the prefix used by the core level router 106. Based on the processing by the distribution level router 106A, the transit level router 110B receives the forwarded packet and forwards the packet to a designated designation associated with the communication network 112.

[0024] Turning now to FIGURES 2B and 2C, the allocation of IP addresses or subsets of IP addresses within the distributed routing environment 100 will be described. With reference to FIGURE 2B, the core level router 106 distributes some portion of the subset of destination IP addresses to distribution level router 108A (illustrated at 202). Distribution level router 108A in turn further distributes the portions of the IP addresses to transit level routers 110A, 110B, and 110C (illustrated at 204, 206, and 208). With reference to FIGURE 2C, the core level router 106 distributes a different portion of the subset of destination IP addresses to distribution level router 108B (illustrated at 210). Distribution level router 108B in turn further distributes the portions of the IP addresses to transit level routers 110A and 110B (illustrated at 212 and 214).

[0025] In an illustrative embodiment, the router management component 102 (FIGURE 1) can allocate responsibility of subsets of IP addresses to the distribution level routers in a variety of manners. In one embodiment, the router management component 102 can allocate responsibility for the entire set of IP addresses in accordance with assignment of IP addresses equally, or substantially equally, among available routers. In this embodiment, each distribution level router 108 becomes responsible for an equal subset of IP addresses or substantially equal if the IP addresses cannot be divided equally. In another embodiment, the router management component 102 can specify specific distribution level router 108 to handle high traffic IP addresses or prefixes. In this example, the entire subset of IP addresses may be custom selected by the router management component 102. Alternatively, only the subset of IP addresses meeting a traffic threshold may be custom selected with the remaining portions of IP address automatically distributed.

[0026] In still a further embodiment, multiple distribution level routers 108 may be selected for a subset of IP addresses. In this embodiment, each core level router 106 can select from multiple distribution level routers 108 based on an equal cost multi-path routing (ECMP) technique in which a specific distribution level router 108 is selected based on a standard load sharing technique. Other factors that can be utilized to select from multiple assigned distribution level router 108 include carrier preference, Internet weather, resource utilization/health reports, an allocated or determine routing cost, service level agreements (SLAs), or other criteria.

[0027] In one embodiment, each distribution router 108 can maintain the portion of the FIB that is associated with the subset of IP addresses assigned the respective distribution level router 108. In another embodiment, each distribution level router 108 can maintain the entire FIB associated with the distributed routing environment 100 in a memory component, such as computer readable medium 158 (FIGURE 1B). Once a subset of IP addresses are assigned to each respective distribution level router 108 (or otherwise updated), the applicable portions of the FIB are loaded in a different memory components, such as memory component 154 (FIGURE 1B) utilized by the router (e.g., a routing chip level content addressable memory or a processor level cache memory). The maintenance of the applicable portions of the FIB in a memory component facilitates better router performance by faster memory access times for the applicable portion of the FIB. However, in this embodiment, the allocation of FIBs to each distribution level router 108 can be modified by loading different portions of the stored FIB from a first memory component storing the entire FIB (e.g., the computer readable medium 158) to the memory component maintaining the portion of the FIB allocated to the distribution level router 108 (e.g., memory component 154). Accordingly, this embodiment facilitates the dynamic allocation of distribution level routers 108, the creation of redundant distribution level routers, and additional failover for distribution level routers. Additionally, one or more core level routers 106 can utilize a similar technique in performing the functions associated with the core level of the distributed routing environment 100.

[0028] In still a further embodiment, as a variation to the above embodiment, each distribution level router can be allocated a larger portion of the FIB associated with the distributed routing environment 100 than is capable of being maintained in a first memory component of the router, such as memory component 154 (e.g., a processor level cache memory). If a core level router 106 routes to a distribution level router 108 and the corresponding prefixes of the destination IP address do not correspond to the FIB maintained in the first memory component of the distribution level router, the distribution level router can recall the necessary information from the larger subset of the FIB maintained in a different memory component (e.g., computer readable medium 158 (FIGURE 1B)). The FIB maintained in the first memory component (e.g., memory component 152) may be updated to store the prefix in the primary memory component. Alternatively, the FIB in the first memory component may not be

automatically updated based on a single request, but based on increases in traffic for a given prefix.

[0029] In yet another embodiment, lower traffic prefixes may be assigned to multiple distribution level routers 108. In one example, each assigned distribution level router 108 does not maintain the lower traffic routing portion of the assigned FIB in the primary memory component. Rather, routing requests for the lower traffic prefixes can be directed to a specific distribution level router based on selection techniques, such as ECMP, and can be processed by a selected distribution level router 108 based on the larger FIB maintained in a different memory component within the selected distribution level router.

[0030] With reference now to FIGURE 3, a routine 300 for routing packets and implemented in a distributed routing environment 100 will be described. At block 302, the distributed routing environment 100 obtains a routing request. As previously described, the routing request is received from a first network 102 (FIGURE 1) and includes information identifying a destination IP address. At block 304, a core level router 106 corresponding to a first level of the distributed routing environment 100 is selected and receives the routing request. In an illustrative embodiment, each core level router 106 can perform the same function and can be selected in accordance with standard selection techniques, including, but not limited to, random selection, round robin selection, load balancing selection and the like.

[0031] At block 306, the selected core level router 106 identifies a distribution level router 108 corresponding to a second level of the distributed routing environment 100. The core level router 106 selects the distribution level router 108 based on processing the destination IP address and utilizing a subset of the destination IP addresses (e.g., the prefix) to determine the appropriate distribution level router 108. Illustratively, in accordance with an embodiment corresponding to the IPv4 communication protocol, the core level router 106 processing can be based on consideration of a prefix of the eight most significant bits. At block 308, the selected distribution level router 108 identifies a transit level router 110 based on processing the destination IP address and utilizing a subset of the destination IP address to determine the appropriate transit level router 110. Illustratively, in accordance with an embodiment corresponding to the IPv4 communication protocol, the distribution level router 108 processing can be based on a larger subset of IP address (e.g., a longer prefix such as 16 or 24 bits, as needed

to select the appropriate transit level router 110). One skilled in the relevant art will appreciate, however, the blocks 306 and 308 may be implemented in a manner such the core level router 106 and distribution level router 108 may utilize additional or alternative attributes (including different portions of a destination IP address) of received packets in identifying the next router component to forward the received packet.

[0032] At block 310, the selected transit level router 110 transmits the receive packet to the destination recipient associated, or otherwise configured, with the transit level router 110. At block 312, the routine 300 terminates.

[0033] With reference now to FIGURE 4, another routine 400 for routing packets and implemented in a distributed routing environment 100 will be described. In an illustrative embodiment, routine 400 may be implemented in embodiments in which less than all the FIB associated with a particular distribution router 108 is maintained in a primary memory component. At block 402, a routing request is received at a distribution level router 108. The selection and routing to a distribution level router 108 was previously described above. Although routine 400 will be described with regard to implementation by a distribution level router 108, one skilled in the relevant art will appreciate that at least portions of routine 400 may be implemented by other components of the distributed routing environment 100, such as core level routers 106 or transit level router 110. At decision block 404, a test is conducted to determine whether the subset of the destination IP address associated with the routing request is in the portion of the FIB table maintained in the primary memory of the selected distribution level router 108. If so, at block 406, the distribution level router 108 obtains the transit layer routing information from the FIB maintained in the first memory component (e.g., memory component 152 (FIGURE 1B)). At block 408, the distribution level router 108 forwards the packet to the selected transit level router 110.

[0034] Alternatively, if at decision block 404 the subset of the destination IP address associated with the routing request is not maintained in the portion of the FIB table maintained in the primary memory of the selected distribution level router 108, at block 410, distribution level router 108 attempts to obtain additional transit routing information from a separate memory component associated with the distribution level router. At block 410, the distribution level router 108 can update the forwarding table information maintained in the primary memory

component with the information obtained from the other memory component. Alternatively, block 410 can be omitted or is otherwise optional. At block 412, the routine terminates.

[0035] While illustrative embodiments have been disclosed and discussed, one skilled in the relevant art will appreciate that additional or alternative embodiments may be implemented within the spirit and scope of the present disclosure. Additionally, although many embodiments have been indicated as illustrative, one skilled in the relevant art will appreciate that the illustrative embodiments do not need to be combined or implemented together. As such, some illustrative embodiments do not need to be utilized or implemented in accordance with the scope of variations to the present disclosure.

[0036] Conditional language, such as, among others, “can,” “could,” “might,” or “may,” unless specifically stated otherwise, or otherwise understood within the context as used, is generally intended to convey that certain embodiments include, while other embodiments do not include, certain features, elements, or steps. Thus, such conditional language is not generally intended to imply that features, elements or steps are in any way required for one or more embodiments or that one or more embodiments necessarily include logic for deciding, with or without user input or prompting, whether these features, elements or steps are included or are to be performed in any particular embodiment. Moreover, unless specifically stated otherwise, or otherwise understood within the context as used, is generally intended to convey utilization of the conjunction “or” in enumerating a list of elements does not limit the selection of only a single element and can include the combination of two or more elements.

[0037] Any process descriptions, elements, or blocks in the flow diagrams described herein and/or depicted in the attached figures should be understood as potentially representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process. Alternate implementations are included within the scope of the embodiments described herein in which elements or functions may be deleted, executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those skilled in the art. It will further be appreciated that the data and/or components described above may be stored on a computer-readable medium and loaded into memory of the computing device using a drive mechanism associated with a computer-readable medium storing

the computer executable components, such as a CD-ROM, DVD-ROM, or network interface. Further, the component and/or data can be included in a single device or distributed in any manner. Accordingly, general purpose computing devices may be configured to implement the processes, algorithms, and methodology of the present disclosure with the processing and/or execution of the various data and/or components described above. Alternatively, some or all of the methods described herein may alternatively be embodied in specialized computer hardware. In addition, the components referred to herein may be implemented in hardware, software, firmware or a combination thereof.

[0038] It should be emphasized that many variations and modifications may be made to the above-described embodiments, the elements of which are to be understood as being among other acceptable examples. All such modifications and variations are intended to be included herein within the scope of this disclosure and protected by the following claims.

**THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE
PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:**

1. A system for routing packets comprising:

5 a router management component, executed on a computing device, for
associating destination address information to a router hierarchy comprising a
plurality of levels, wherein the router management component is configured to,
for individual levels of the router hierarchy, dynamically allocate responsibility
for portions of destination addresses of incoming packets to one or more router
10 components of the individual level of the router hierarchy;

 one or more router components corresponding to a first level of the router
hierarchy, the first level of the router hierarchy for receiving the incoming packets
for routing and transmitting the incoming packets to router components of a
second level of the router hierarchy;

15 one or more router components corresponding to the second level of the
hierarchy, the second level of the router hierarchy for routing the incoming
packets received from the first level of a router hierarchy to router components of
a third level of the router hierarchy, and wherein at least one router component
corresponding to the second level is dynamically allocated responsibility for a
20 portion of a first subset of each destination address of each incoming packet by
the router management component based at least in part on the incoming packets;

 one or more router components corresponding to the third level of the
router hierarchy, the third level of the router hierarchy for routing the incoming
packets received from the second level of the router hierarchy, wherein at least
25 one router component corresponding to the third level is dynamically allocated
responsibility for a portion of a second subset of each destination address of each
incoming packet by the router management component based at least in part on
the incoming packets;

 wherein the one or more router components corresponding to the first
30 level of the router hierarchy identifies one or more router components from the

second level of the router hierarchy to which to route each packet of the incoming packets based at least in part on the first subset of each destination address associated with each packet, wherein the first subset is designated by the router management component; and

5 wherein the one or more router components corresponding to the second level of the router hierarchy identifies one or more router components of the third level of the router hierarchy to which to route each packet of the incoming packets based at least in part on the second subset of each destination address associated with each incoming packet, wherein the second subset is designated by the router management component and wherein the second subset is greater than the first subset.

10 2. The system as recited in Claim 1, wherein the one or more router components corresponding to the second level of the router hierarchy are dynamically allocated responsibility for approximately equal portions of the first subset of each destination address of each incoming packet.

15 3. The system as recited in Claim 1, wherein the one or more router components corresponding to the second level of the router hierarchy are dynamically allocated responsibility for portions of the first subset of each destination address of each incoming packet based on traffic volumes attributed to the portions of the first subset.

20 4. The system as recited in Claim 1, wherein the one or more router components corresponding to the first level of a router hierarchy are selected in accordance with one of random selection, round robin selection, hashing and load balancing.

 5. The system as recited in Claim 1, wherein each destination address corresponds to an IP address.

25 6. The system as recited in Claim 5, wherein first subset of each IP address corresponds to the eight most significant bits of each IP address.

 7. The system as recited in Claim 5, wherein the second subset of each IP address corresponds to at least one of the sixteen or twenty four most significant bits of each IP address.

8. The system as recited in Claim 1, wherein at least two of the one or more router components corresponding to the first level of the router hierarchy are implemented in a common physical router component.

5 9. The system as recited in Claim 1, wherein at least two of the one or more router components corresponding to the second level of the router hierarchy are implemented in a common physical router component.

10 10. The system as recited in Claim 1, wherein at least one physical router component implements at least two of a router component corresponding to the first level of the router hierarchy, a router component corresponding to the second level of the router hierarchy and a router component corresponding to the second level of the router hierarchy.

11. A system for routing packets comprising:

a first set of logical router components for receiving an incoming set of packets for routing;

15 a second set of logical router components for routing the set of packets received from the first set of router components;

a third set of logical router components for routing the set of packets received from the second set of router components;

20 wherein the first set of logical router components identify a router component from the second set of logical router components to which to route a first packet of the set of packets based at least in part on a dynamic assignment by a router management component of portions of a first subset of destination addresses associated with the incoming set of packets to individual router components from the second set of logical router components;

25 wherein each of the one or more router components corresponding to the second set of logical router components is dynamically assigned by the router management component to the portions of the first subset of the destination addresses associated with the incoming set of packets, wherein the dynamic assignment is based at least in part on the set of packets; and

30 wherein the second set of logical router components identify a router component from the third set of logical router components to which to route the

first packet based at least in part on a dynamic assignment by the router management component of portions of a second subset of destination addresses associated with the incoming set of packets to individual router components from the third set of logical router components.

5 **12.** The system as recited in Claim 11, wherein the dynamic assignment of router components from the second set of router components to portions of the first subset of the destination addresses is based at least in part on a substantially equal allocation of the destination addresses.

10 **13.** The system as recited in Claim 11, wherein dynamic assignment of router components from the second set of logical router components to portions of the first subset of the destination addresses is based at least in part on traffic volumes associated with the portions of the first subset of the destination addresses.

15 **14.** The system as recited in Claim 11, wherein the dynamic assignment of router components from the second set of logical router components to portions of the first subset of the destination addresses is based at least in part on a combination of traffic volumes associated with some portions of the first subset of the destination addresses and an equal allocation of remaining portions of the first subset of the destination addresses.

20 **15.** The system as recited in Claim 11, wherein each of the router components from the second set of logical router components is associated with a threshold number of destination addresses and wherein the dynamic assignment of router components from the second set of logical router components to portions of the first subset of the destination addresses is based on allocation of portions of the first subset of the destination addresses including a number of destination addresses greater than the threshold number of destination addresses .

25 **16.** The system as recited in Claim 11, wherein the dynamic assignment of router components from the second set of logical router components to portions of the first subset of the destination addresses is based on low traffic volumes associated with one or more of the portions of the first subset of the destination addresses.

30 **17.** The system as recited in Claim 11, wherein the dynamic assignment of router components from the second set of logical router components to portions of the

first subset of the destination addresses includes an assignment of a plurality of router components for the same portion of the first subset of destination addresses.

18. The system as recited in Claim 11, wherein the first set of router components are selected in accordance with one of random selection, round robin selection, hash selection and load balancing.

19. The system as recited in Claim 11, wherein the destination addresses correspond to IP addresses.

20. The system as recited in Claim 19, wherein the first subset of the destination addresses corresponds to the eight most significant bits of the IP addresses.

10 21. The system as recited in Claim 20, wherein the second subset of the destination addresses corresponds to at least one of the sixteen or twenty four most significant bits of the IP addresses.

22. The system as recited in Claim 11, wherein each of the first set of logical router components correspond to a physical router component.

15 23. The system as recited in Claim 11, wherein two or more of the first set of logical router components correspond to a single physical router component.

24. The system as recited in Claim 11, wherein each of the second set of logical router components correspond to a physical router component.

20 25. The system as recited in Claim 11, wherein two or more of the second set of logical router components correspond to a single physical router component.

26. The system as recited in Claim 11, wherein each of the third set of logical router components correspond to a physical router component.

27. The system as recited in Claim 11, wherein two or more of the third set of logical router components correspond to a single physical router component.

25 28. The system as recited in Claim 11, wherein at least one of the first set of logical router components, at least one of the second set of logical router components, and at least one of the third set of logical router components correspond to a single physical router component.

29. A method for routing packets comprising:
30 obtaining a routing request corresponding to a data packet received from a first communication network;

identifying a first router corresponding to a first level of a router hierarchy, the first level of the router hierarchy corresponding to one or more router components;

forwarding the received data packet to the identified first router;

5 identifying a second router corresponding to a second level of the router hierarchy, the second level of the router hierarchy corresponding to one or more router components;

forwarding the received data packet to the identified second router; and

10 identifying a third router corresponding to a third level of the router hierarchy, the third level of the router hierarchy corresponding to one or more router components;

wherein identifying the second router is based at least in part on a dynamic assignment of the second router to a portion of a first subset of a destination address associated with the received data packet, and wherein the dynamic assignment is based at least in part on a set of previously received data packets; and

15

wherein identifying a third router is based at least in part on a dynamic assignment of the third router to a portion of a second subset of the destination address associated with the received data packet, and wherein the dynamic assignment is based at least in part on a set of previously received data packet.

20

30. The method as recited in Claim **29**, wherein each of the one or more router components corresponding to the second level of the router hierarchy is dynamically assigned to a portion of the first subset of the destination address associated with the received packet.

25 **31.** The method as recited in Claim **30**, wherein the dynamic assignment of routers from the second level of the router hierarchy to portions of the first subset of the destination address is based at least in part on at least a substantially equal allocation of portions of the first subset of the destination address associated with the received packet.

32. The method as recited in Claim **30**, wherein the dynamic assignment of routers from the second level of the router hierarchy to portions of the first subset of the destination address is based at least in part on traffic volumes associated with the

30

dynamically assigned portion of the first subset of the destination address associated with the received packet.

33. The method as recited in Claim 30, wherein the dynamic assignment of routers from the second level of the router hierarchy to portions of the first subset of the destination address is based on a combination of traffic volumes associated with some portions of the first subset of the destination addresses and an equal allocation of remaining portions of the first subset of the destination address.

34. The method as recited in Claim 30, wherein each of the routers from the second level of the router hierarchy is associated with a threshold number of destination addresses and wherein the dynamic assignment of routers from the second level of the router hierarchy to portions of the first subset of the destination address is at least in part based on allocation of portions of the first subset of the destination addresses including a number of destination addresses greater than the threshold number of destination addresses .

35. The method as recited in Claim 30, wherein the dynamic assignment of routers from the second level of the router hierarchy to portions of the first subset of the destination address is based on low traffic volumes associated with one or more of the portions of the first subset of the destination address.

36. The method as recited in Claim 30, wherein the dynamic assignment of routers from the second set of router components to portions of the first subset of the destination address includes an assignment of a plurality of routers for the same portion of the first subset of the destination address.

37. The method as recited in Claim 29, wherein first set of router components are selected in accordance with one of random selection, round robin selection, hash selection and load balancing.

38. The method as recited in Claim 29, wherein the destination address corresponds to an IP address.

39. The system of Claim 1, wherein the router management component is further configured, for individual levels of the router hierarchy, to distribute portions of a routing table containing address information for the incoming packets to router components of the individual level of the router hierarchy.

40. The system of Claim 39, wherein the portions of the routing table distributed to the router components of the first level of the router hierarchy are selected based at least in part on the portion of the first subset of each destination address for which the router components of the first level of the router hierarchy are assigned responsibility.

5 41. The system of Claim 39, wherein the portions of the routing table distributed to the router components of the second level of the router hierarchy are selected based at least in part on the portion of the second subset of each destination address for which the router components of the second level of the router hierarchy are assigned responsibility.

10 42. The system of Claim 11, wherein the router management component is further configured to distribute portions of a routing table to individual router components of the second set of logical router components based at least in part on the portion of the first subset of destination addresses for which the individual router components have been dynamically assigned.

43. The method of Claim 29 further comprising:

15 dividing a routing table including address information for forwarding the received data packet between levels of the routing hierarchy based at least in part on the portion of the first subset of the destination addresses dynamically assigned to the identified second router and the portion of the second subset of the destination addresses dynamically assigned to the identified third router;

20 distributing a first part of the divided routing table to the identified second router, wherein the first part includes address information corresponding to the portion of the first subset of the destination addresses dynamically assigned to the identified second router; and

25 distributing a second part of the divided routing table to the identified third router, wherein the second part includes address information corresponding to the portion of the second subset of the destination addresses dynamically assigned to the identified third router.

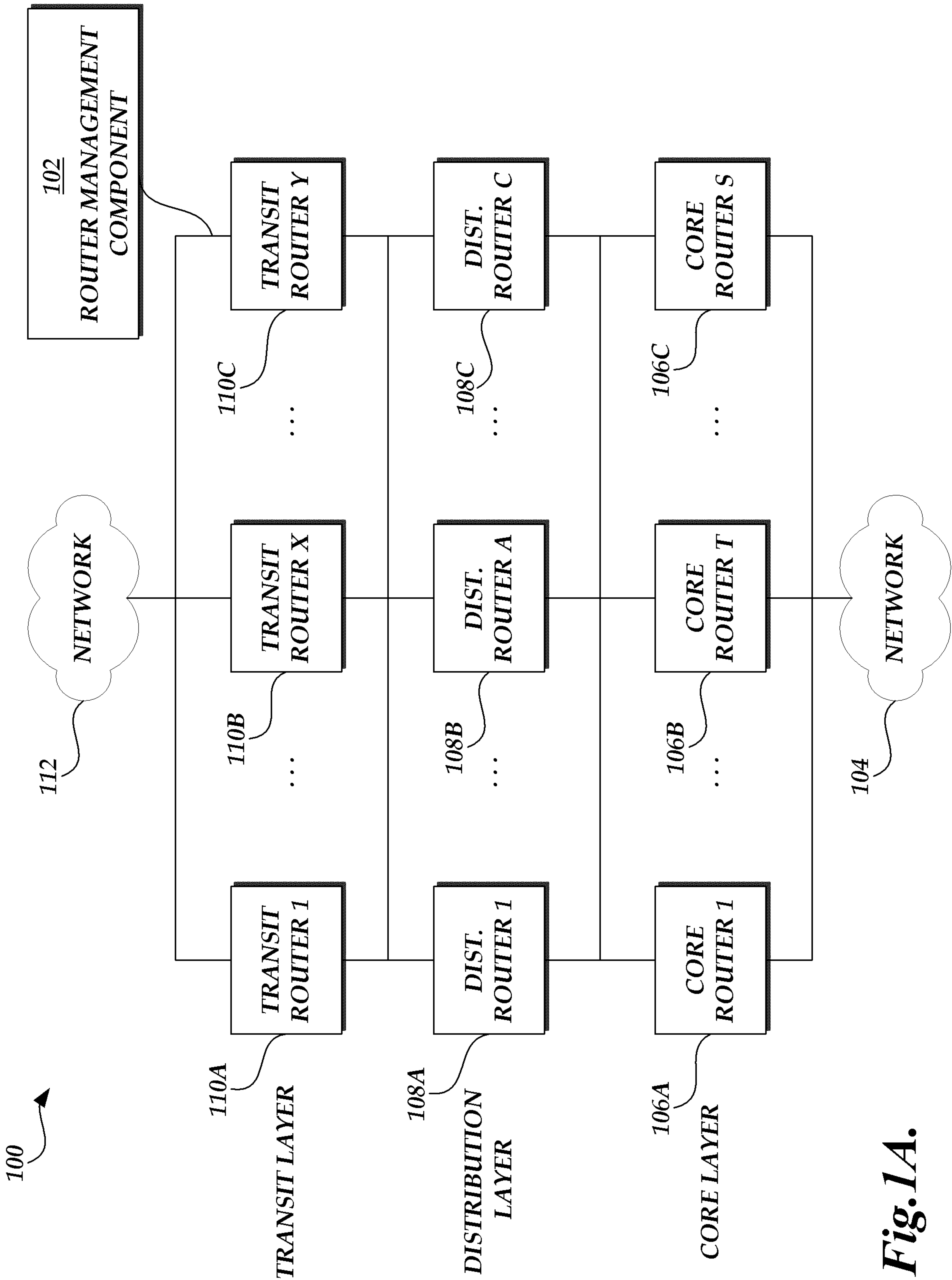


Fig. 1A.

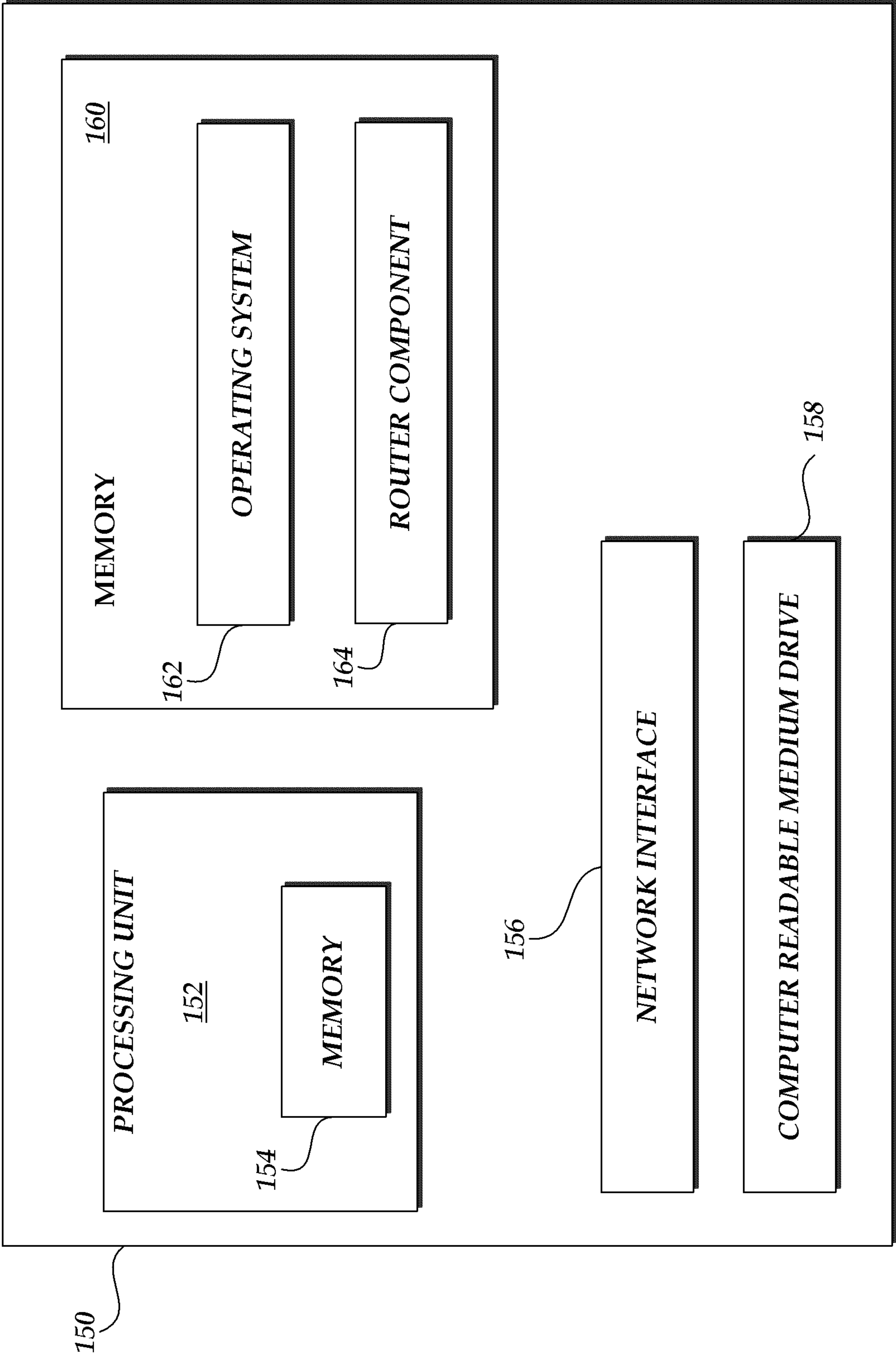


Fig. 1B.

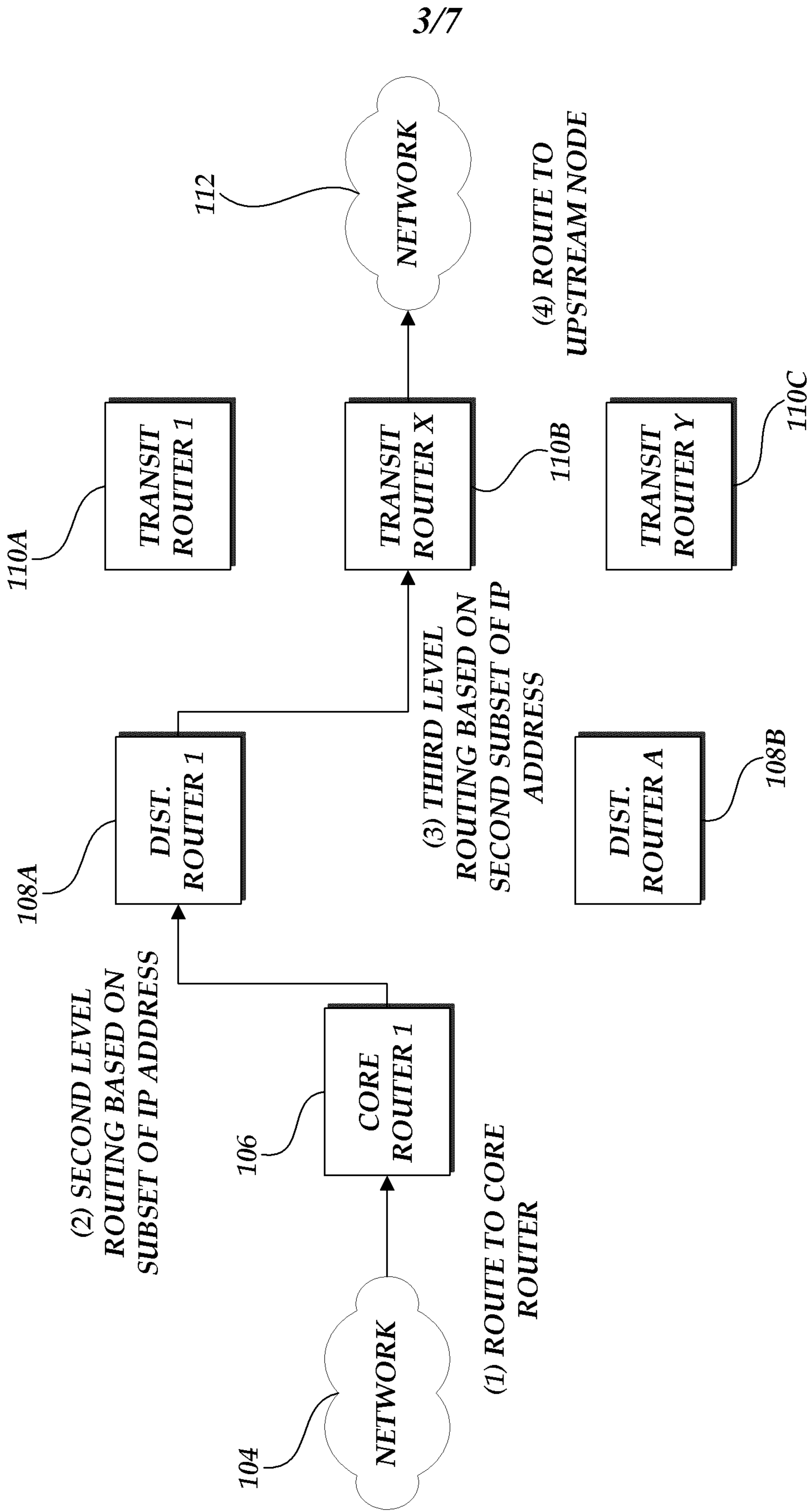


Fig.2A.

4/7

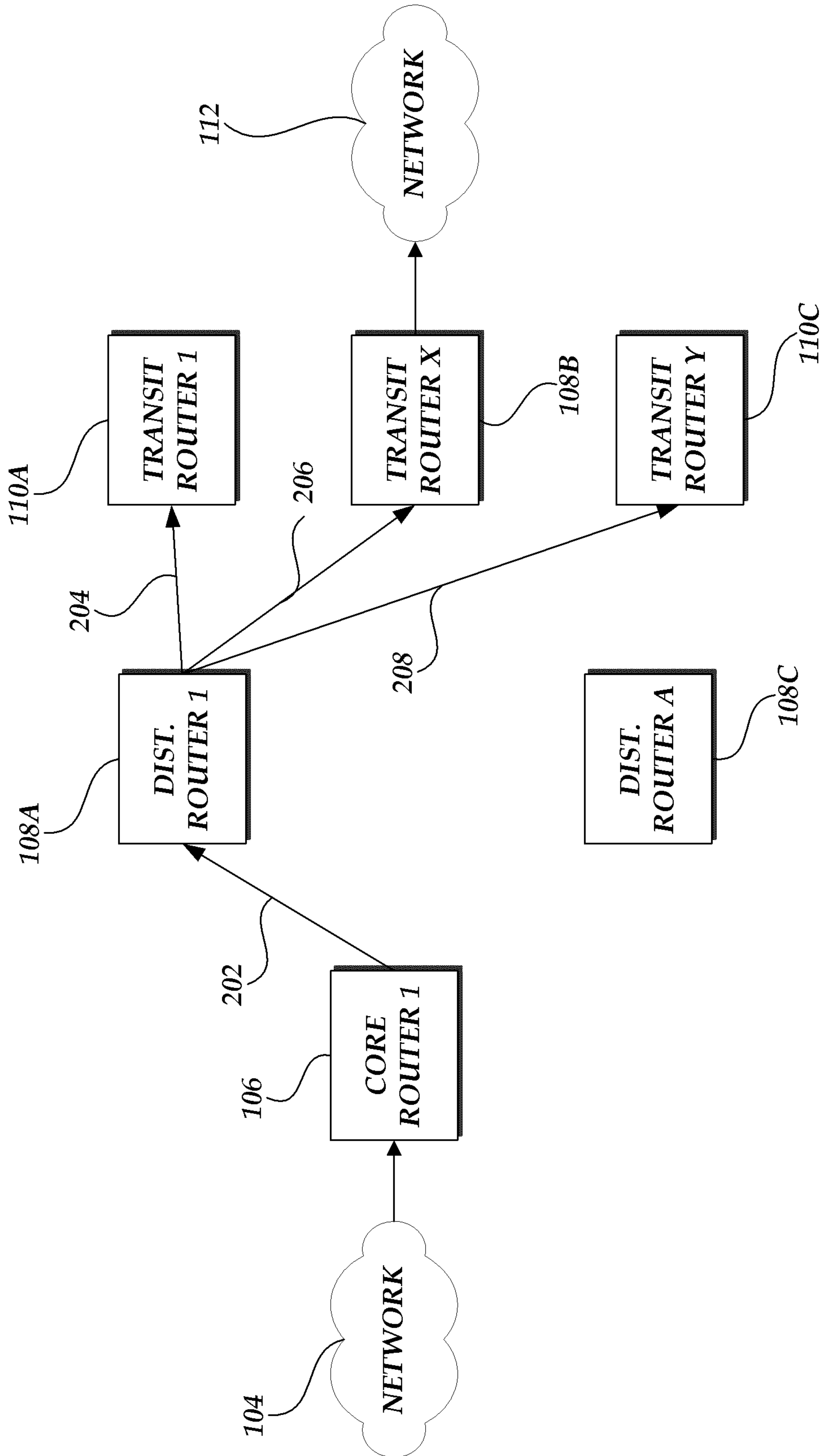


Fig. 2B.

5/7

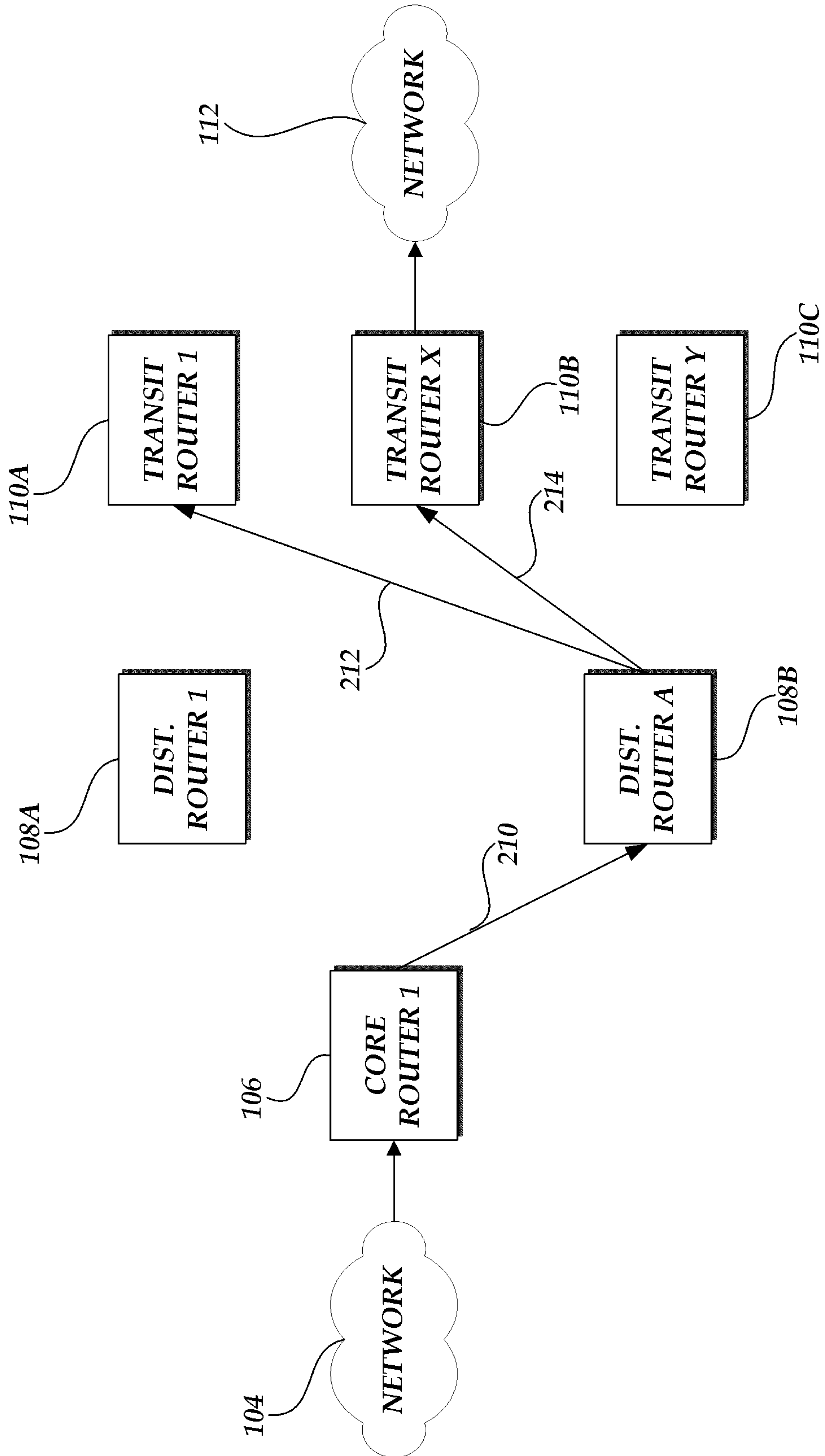
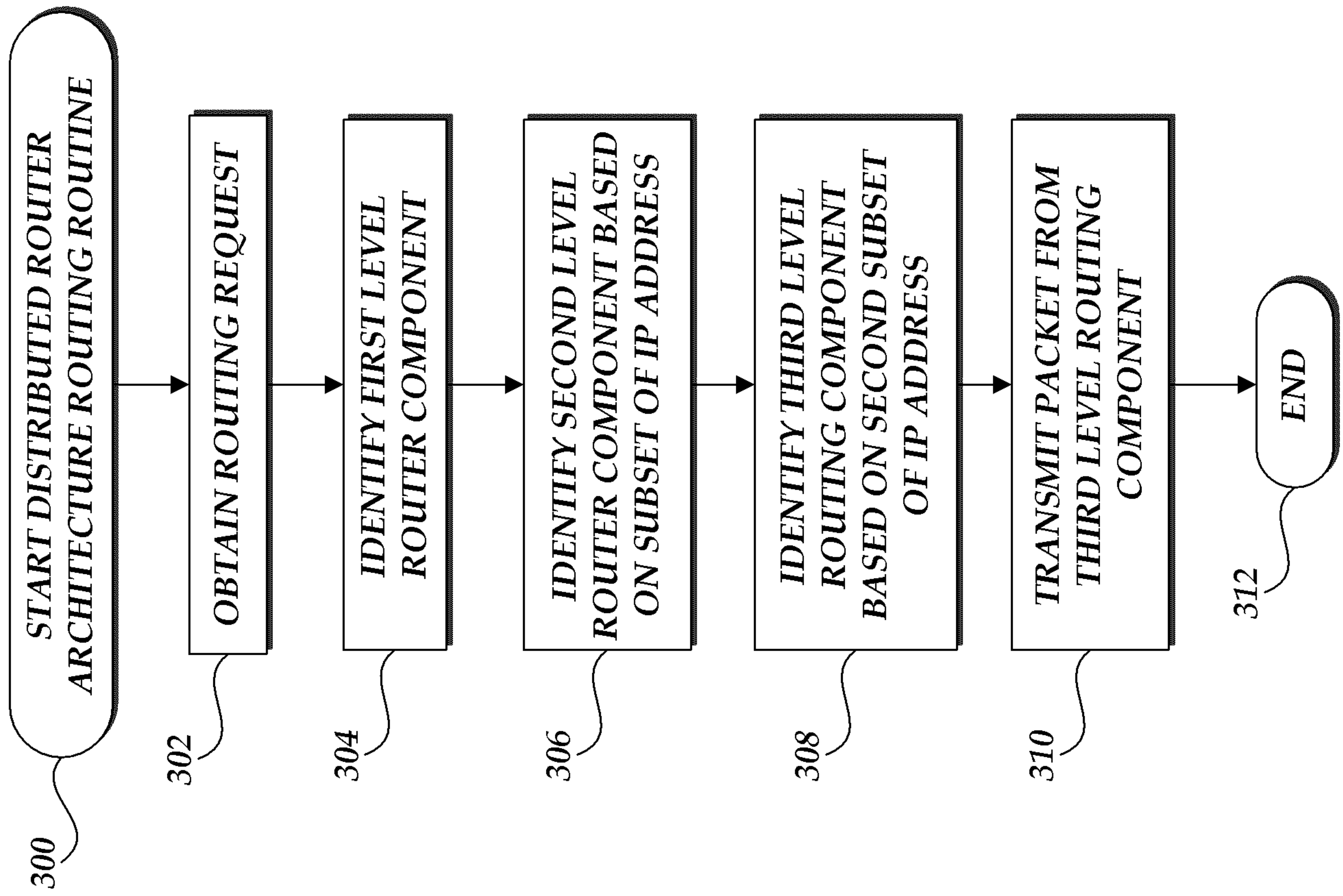
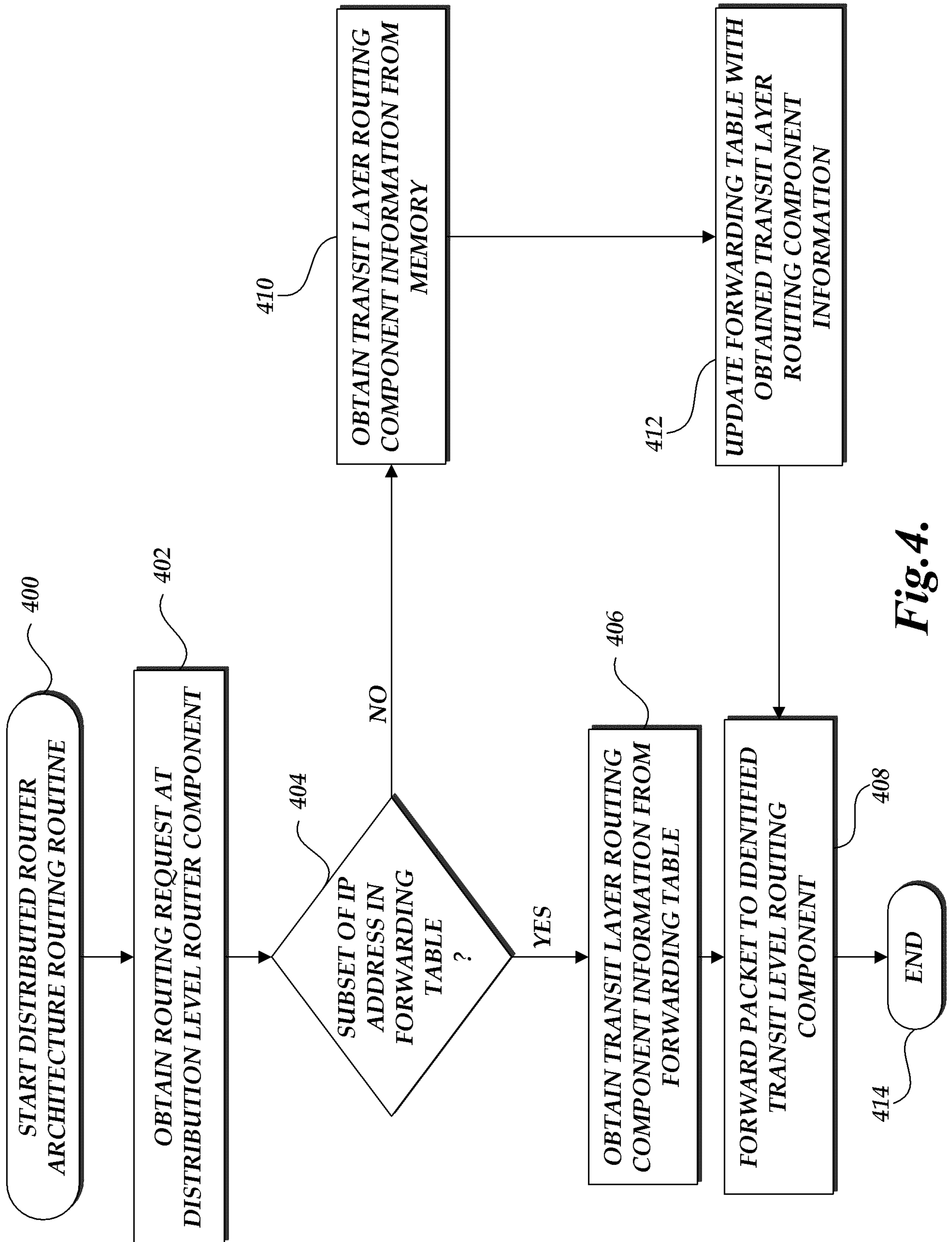


Fig.2C.

6/7

**Fig.3.**

7/7

**Fig. 4.**

