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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 February 2009 (12.02.2009)

PCT

(10) International Publication Number
WO 2009/020980 A1

(51) International Patent Classification:
H04L 12/28 (2006.01)

(21) International Application Number:
PCT/US2008/072250

(22) International Filing Date: 5 August 2008 (05.08.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/954,201 6 August 2007 (06.08.2007) US
12/173,195 15 July 2008 (15.07.2008) US

(71) Applicant *(for all designated States except US)*: **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **MARUCHECK, Michael, J.**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **LOVERING, Bradford, H.**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **FEINGOLD, Max, Attar**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **HASHA, Richard L.**; One Microsoft Way, Redmond, Washington 98052-6399 (US). **ABBOT, Michael**; One Microsoft Way, Redmond, Washington 98052-6399 (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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(54) Title: FITNESS BASED ROUTING

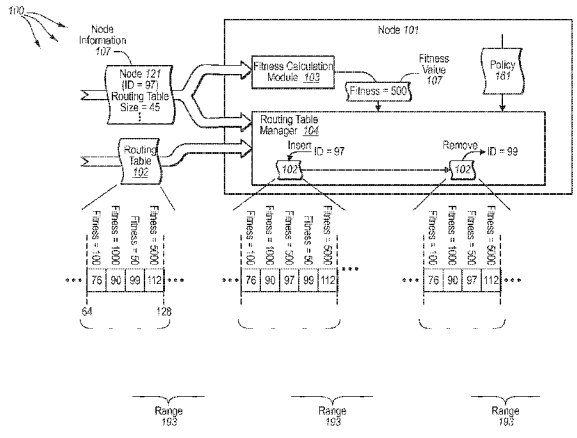


FIG. 1B

(57) Abstract: The present invention extends to methods, systems, and computer program products for fitness based routing. Embodiments of the invention significantly improve the likelihood that routing nodes contained in routing table have adequate (or even relatively increased) ability to transfer and process messages in an overlay network. Thus, when the node is to make a routing decision for a message, the node has some assurances that any selected routing node is adequate (or is at least the best currently available). Further, a sending node can take preference to routing nodes with higher fitness values when sending a message. Preference to higher fitness metric values further insures that messages are adequately transferred and processed. Accordingly, embodiments of the invention can be used to route messages in a manner that optimizes bandwidth and provides efficient routing capability.

WO 2009/020980 A1

FITNESS BASED ROUTING

BACKGROUND

1. Background and Relevant Art

[0001] Computer systems and related technology affect many aspects of society. Indeed, the computer system's ability to process information has transformed the way we live and work. Computer systems now commonly perform a host of tasks (e.g., word processing, scheduling, accounting, etc.) that prior to the advent of the computer system were performed manually. More recently, computer systems have been coupled to one another and to other electronic devices to form both wired and wireless computer networks over which the computer systems and other electronic devices can transfer electronic content. Accordingly, the performance of many computing tasks are distributed across a number of different computer systems and/or a number of different computing components.

[0002] An overlay network is a fabric that extends over today's traditional networks (both private networks and the internet) and provides a uniform view that masks the specifics of the underlying networks. Communication between nodes on such an overlay network consists of a message being routed by the nodes (computer systems) using specific algorithms until the intended end point is reached. Routing over overlay networks today is primarily static, that it is not inherently adaptive.

[0003] The nodes that participate in an overlay network are often treated as being homogeneous in nature as far as routing is concerned. Thus, there is a general assumption that all nodes have similar computational power, network connectivity and perceived load. This assumption is valid at least in some environments, such as, for example, where the nodes are in a controlled datacenter type environment with consistent loads. However, in many other environments this assumption is not valid. For example, nodes on the Internet or other distributed computing environments have heterogeneous varying configurations load, and computation capacity.

[0004] Further, network link capacities between nodes on the Internet and other distributed networks vary from broadband to dialup speeds. For example, assuming there are two potential routes to get from point A to point B. One route involves

going through a node that is connected by a dialup and another route involves going through a node that is connected by a high speed broadband connection. If all nodes were to be treated alike, probabilistically, messages can be routed through the node connected by dial-up, which is a sub-optimal route and affects the efficiency of message transfer. These, as well as other, inefficiencies can be magnified using overlay networks that are large scale (e.g., millions of nodes) and have high capacity machines and links that can be effectively utilized.

[0005] At least one other problem with overlay networks is the mechanism that is used to update presence information between nodes (i.e., when one node knows of the presence of another node). Typical mechanisms for propagating presence information include flooding the overlay or, at least a large portion of the overlay, with the node's presence information. The presence information is picked up by members of the overlay and stored as part of their routing table and used for subsequent routing decisions. However, flooding presence information onto the overlay is very expensive and consumes significant network capacity. The consumed network capacity is then unavailable for actual application messages. The problem is potentially significantly exacerbated as the number of nodes on an overlay network increases, such as, for example, when there are thousands or even millions of nodes publishing presence information.

[0006] There are also other problems associated with overlay networks. For example, using an overlay network it is often beneficial to communicate between two arbitrary nodes on the overlay. In order to do that, the source node needs the end point address (IP address/DNS name) of the destination. Overlay networks employ a decentralized mechanism for routing messages across the overlay, wherein each node has a partial knowledge of the location of other nodes in the overlay. Messages are passed across the overlay from the source to a node that is "closer" to the destination and with each hop gets numerically closer to the destination, until the destination node is reached. The use of "closeness" as a sole factor in considering a route between nodes can often cause less efficient routes between nodes.

BRIEF SUMMARY

[0007] The present invention extends to methods, systems, and computer program products for fitness based routing. Embodiments of the invention include maintaining a routing table at a computer system, such as, for example, a node in an overlay network. The computer system receives node information for another node that exists at a specified location within the overlay network. The node information includes fitness information for the other node.

[0008] The computer system accesses a routing table that includes one or more nodes, each node in the routing table being a node that the computer system can send a message to to delivery the message to a destination node within the overlay network. Each node in the routing table having a fitness metric value representing an ability of the node to transfer and process messages within the overlay network. The computer system calculates a fitness metric value for the other node. The calculated fitness metric value represents the other node's ability to transfer and process messages within the overlay network. The fitness metric value based at least in part on the fitness information for the other node.

[0009] The computer system inserts the other node is inserted into the routing table. The routing table is divided into a plurality of ranges. Each range corresponds to a portion of the overlay network. The computer system assigns each node in the routing table a specified range based on the location of the node in the overlay network.

[0010] The computer identifies the range that includes the most nodes. The computer system identifies the node within the identified range that is least able to transfer and process messages within the overly network based on fitness metric values of the nodes in the identified range. The computer system removes the identified node form the routing table.

[0011] In some embodiments, node information for a plurality of nodes in an overlay network is received in a message from another node in the overlay network. A fitness metric is calculated for each of the plurality of nodes and each of the other nodes is inserted in the computer system's routing table. It is determined that the number of nodes in the routing table exceeds a specified number.

[0012] The routing table is divided into a plurality of ranges, each range corresponding to a portion of the overlay network. Each of the nodes in the routing table is assigned to one of the plurality of ranges. The range with the most nodes is identified. From among the nodes in the identified range, the node least fit to transfer and process messages, based on fitness metric values, is removed from the routing table. Identification of the range with the most nodes and removal of least fit nodes from within that range can continue until the number of nodes in the routing table no longer exceeds the specified number.

[0013] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used as an aid in determining the scope of the claimed subject matter.

[0014] Additional features and advantages of the invention will be set forth in the description which follows, and in part will be obvious from the description, or may be learned by the practice of the invention. The features and advantages of the invention may be realized and obtained by means of the instruments and combinations particularly pointed out in the appended claims. These and other features of the present invention will become more fully apparent from the following description and appended claims, or may be learned by the practice of the invention as set forth hereinafter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] In order to describe the manner in which the above-recited and other advantages and features of the invention can be obtained, a more particular description of the invention briefly described above will be rendered by reference to specific embodiments thereof which are illustrated in the appended drawings. Understanding that these drawings depict only typical embodiments of the invention and are not therefore to be considered to be limiting of its scope, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

[0016] Figure 1A illustrates a view of an example overlay ring network architecture that facilitates fitness based routing.

[0017] Figure 1B illustrates a view of an example node of the overlay ring network architecture maintaining a routing table.

[0018] Figures 1C and 1D illustrates another view of the example node of the overlay ring network architecture maintaining a routing table.

[0019] Figure 2 illustrates a flow chart of an example method for maintaining a routing table.

[0020] Figure 3 illustrates a flow chart of another example method for maintaining a routing table.

DETAILED DESCRIPTION

[0021] The present invention extends to methods, systems, and computer program products for fitness based routing. Embodiments of the invention include maintaining a routing table at a computer system, such as, for example, a node in an overlay network. The computer system receives node information for another node that exists at a specified location within the overlay network. The node information includes fitness information for the other node.

[0022] The computer system accesses a routing table that includes one or more nodes, each node in the routing table being a node that the computer system can send a message to to delivery the message to a destination node within the overlay network. Each node in the routing table having a fitness metric value representing an ability of the node to transfer and process messages within the overlay network. The computer system calculates a fitness metric value for the other node. The calculated fitness metric value represents the other node's ability to transfer and process messages within the overlay network. The fitness metric value based at least in part on the fitness information for the other node.

[0023] The computer system inserts the other node is inserted into the routing table. The routing table is divided into a plurality of ranges. Each range corresponds to a portion of the overlay network. The computer system assigns each node in the routing table a specified range based on the location of the node in the overlay network.

[0024] The computer identifies the range that includes the most nodes. The computer system identifies the node within the identified range that is least able to transfer and process messages within the overly network based on fitness metric values of the nodes in the identified range. The computer system removes the identified node form the routing table.

[0025] In some embodiments, node information for a plurality of nodes in an overlay network is received in a message from another node in the overlay network. A fitness metric is calculated for each of the plurality of nodes and each of the other nodes is inserted in the computer system's routing table. It is determined that the number of nodes in the routing table exceeds a specified number.

[0026] The routing table is divided into a plurality of ranges, each range corresponding to a portion of the overlay network. Each of the nodes in the routing table is assigned to one of the plurality of ranges. The range with the most nodes is identified. From among the nodes in the identified range, the node least fit to transfer and process messages, based on fitness metric values, is removed from the routing table. Identification of the range with the most nodes and removal of least fit nodes from within that range can continue until the number of nodes in the routing table no longer exceeds the specified number.

[0027] Embodiments of the present invention may comprise or utilize a special purpose or general-purpose computer including computer hardware, as discussed in greater detail below. Embodiments within the scope of the present invention also include physical and other computer-readable media for carrying or storing computer-executable instructions and/or data structures. Such computer-readable media can be any available media that can be accessed by a general purpose or special purpose computer system. Computer-readable media that store computer-executable instructions are physical storage media. Computer-readable media that carry computer-executable instructions are transmission media. Thus, by way of example, and not limitation, embodiments of the invention can comprise at least two distinctly different kinds of computer-readable media: physical storage media and transmission media.

[0028] Physical storage media includes RAM, ROM, EEPROM, CD-ROM or other optical disk storage, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store desired program code means in the form of computer-executable instructions or data structures and which can be accessed by a general purpose or special purpose computer.

[0029] With this description and following claims, a "physical network" is defined as one or more data links that enable the transport of electronic data between computer systems and/or modules and/or other electronic devices.

[0030] Within this description and in the following claims, an "overlay network" is defined as a computer network that is built on top of another network (e.g., a physical network or another overlay network). Nodes on an overlay network can be viewed as being connected by virtual or logical links, each of which corresponds to a path, perhaps through many physical networks and/or data links, in an underlying network. For example, many peer-to-peer networks are overlay networks because they run on top of the Internet. Overlay networks can be constructed in order to permit routing messages to destinations not specified by an IP address. For example, distributed hash tables can be used to route messages to a node having specific logical address, whose IP address is not known in advance.

[0031] When information is transferred or provided over a network or another communications connection (either hardwired, wireless, or a combination of hardwired or wireless) to a computer, the computer properly views the connection as a transmission medium. Transmissions media can include a network and/or data links which can be used to carry or desired program code means in the form of computer-executable instructions or data structures and which can be accessed by a general purpose or special purpose computer. Combinations of the above should also be included within the scope of computer-readable media.

[0032] Further, it should be understood, that upon reaching various computer system components, program code means in the form of computer-executable instructions or data structures can be transferred automatically from transmission media to physical storage media (or vice versa). For example, computer-executable instructions or data structures received over a network or data link can be buffered

in RAM within a network interface module (e.g., a "NIC"), and then eventually transferred to computer system RAM and/or to less volatile physical storage media at a computer system. Thus, it should be understood that physical storage media can be included in computer system components that also (or even primarily) utilize transmission media.

[0033] Computer-executable instructions comprise, for example, instructions and data which cause a general purpose computer, special purpose computer, or special purpose processing device to perform a certain function or group of functions. The computer executable instructions may be, for example, binaries, intermediate format instructions such as assembly language, or even source code. Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the described features or acts described above. Rather, the described features and acts are disclosed as example forms of implementing the claims.

[0034] Those skilled in the art will appreciate that the invention may be practiced in network computing environments with many types of computer system configurations, including, personal computers, desktop computers, laptop computers, message processors, hand-held devices, multi-processor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, mobile telephones, PDAs, pagers, routers, switches, and the like. The invention may also be practiced in distributed system environments where local and remote computer systems, which are linked (either by hardwired data links, wireless data links, or by a combination of hardwired and wireless data links) through a network, both perform tasks. In a distributed system environment, program modules may be located in both local and remote memory storage devices.

[0035] Figure 1A illustrates a view of network architecture 100 that facilitates fitness based routing. As depicted in Figure 1A, network architecture 100 includes ring 151. Ring 151 is a bi-directional doubly linked list of 2^9, or 512, locations that may be occupied by a node, providing a doubly linked ring topology (i.e., an

overlay network). Ring 151 includes a plurality of nodes including node 101 and other nodes listed by ID number in routing table 102. Generally, ring 151 represents logical connectively between nodes that may be physically connected over various and different underlying networks and connections. For example, each of the nodes can be physically connected to one another over a system bus and/or over (or be part of) one or more underlying networks, such as, for example, a Local Area Network ("LAN"), a Wide Area Network ("WAN"), and even the Internet.

[0036] Nodes on ring 151 can use various overlay protocols to communicate with another. Overlay protocols can build and/or be based on protocols of the underlying physical networks. Thus, each of the depicted nodes as well as any other connected nodes, can create message related data and exchange message related data (e.g., Internet Protocol ("IP") datagrams and other higher layer protocols that utilize IP datagrams, such as, Transmission Control Protocol ("TCP"), Hypertext Transfer Protocol ("HTTP"), Simple Mail Transfer Protocol ("SMTP"), etc.) over the underlying networks.

[0037] Node 101 maintains routing table 102 that stores a plurality of nodes (hereinafter referred to as "routing nodes"). Node 101 can communicate directly with each of the routing nodes in routing table 102. Node 101 can send a message to a routing node to delivery the message to a node that is closer to the routing node than node 101. For example, to delivery a message to the node having ID = 403, node 101 can send the message to the routing node having ID = 401

[0038] Node 101 can maintain a routing table 102 in accordance with routing table policy 161. Routing table policy 161 can define a maximum number of routing nodes that are to be included in routing table 102. Routing table policy 161 can also indicate how to divide nodes within routing table 102 into a number of ranges and define a maximum number of routing nodes per range. For example, routing table policy 161 can dictate that routing table 102 can be divided into ranges 191, 192, 193, 194, 195, 196, 197 and 198.

[0039] The range IDs includes in a range can be varied such that ranges closer to node 101 are more densely populated with routing nodes and ranges further from

node 101 are less densely populated with routing nodes. For example, range 192 includes four routing nodes in a range from +32 IDs to +64 IDs from node 101. Range 194 includes four routing nodes in a range from +128 to +246. Thus, both range 192 and 194 include four routing nodes. However, range 192 is more densely populated since the four routing nodes in range 192 are spread over a smaller distance on ring 151. On the other hand, range 194 is less densely populated since the four routing nodes in range 192 are spread over a greater distance on ring 151.

[0040] Routing table policy 161 can also define a range for nearby neighborhood nodes. For example, nodes within +16 to -16 node IDs from node 101 can automatically be include in routing table 102 as neighborhood nodes. Neighborhood nodes can be split between predecessor and successor neighborhoods. For example, routing table 102 includes successor neighborhood 181 and predecessor neighborhood 182 each having three nodes relatively close to (within 16 node IDs of) node 101.

[0041] Figure 1B illustrates a view of node 101 maintaining routing table 102. Referring now to Figure 1B, node 101 includes fitness calculation module 103 and routing table manager 104. Fitness calculation module 103 is configured to calculate fitness values for routing nodes. Fitness calculation module 103 can implement a fitness calculation algorithm to calculate fitness values for routing nodes. Generally, a fitness value indicates a routing node's ability to process and route messages.

[0042] A fitness value can be calculated from a variety of different types of data that indicate, at least to some extent, a node's ability to process and route messages. For example, a fitness value can be calculated from one or more of: routing table size of the routing node, latency between node 101 and the routing node, the number of messages that a routing node sends, and the number of messages that a routing node receives. Different types of data can be weighted differently when calculating a fitness value. In some embodiments, a single type of data, such as, for example, routing table size, is used to calculate a fitness value.

[0043] Routing table manager 104 is configured to put nodes into and remove nodes from routing table 102 in accordance with a routing table policy 161. Routing table manager 104 can also compare fitness values between routing nodes to determine what routing nodes to keep and what routing nodes remove.

[0044] From time to time, node 101 communicates with nodes on ring 151. During this communication, the nodes can include node information about themselves as well as other nodes on ring 151. The node information can include routing table sizes of other nodes, latency between node 101 and a node, the number of messages that a node has sent, and the number of messages a node has received. Based on received node information, fitness calculation module can calculate a fitness value for one or more nodes.

[0045] Some of the nodes can be retained in routing table 102 along with their corresponding fitness values. For example, within range 193: node ID 76 has fitness = 100, node ID 90 has fitness = 1000, node ID 99 has fitness = 50, and node ID 112 has fitness = 5000. Routing table manager 104 can permit the routing table 102 and its various ranges to fill up with routing nodes to specified maximum values. After reaching specified maximum values, routing table manager 104 can interoperate with fitness calculation module 103 to maintain the size of routing table 102 and its various ranges at the specified maximum values.

[0046] Figure 2 illustrates a flow chart of method 200 for maintaining a routing table. Method 200 will be described with respect to the components and data in Figures 1A and 1B.

[0047] Method 200 includes an act of receiving node information for another node that exists at a specified location within the overlay network, the node information including fitness information for the other node (act 201). For example, node 101 can receive node information 107. Node information 107 includes characteristics about node 121. For example, node information 107 indicates that node 121 has routing table size of 45. That is, the routing table of node 121 includes 45 routing nodes.

[0048] Method 200 includes an act of accessing a routing table that includes one or more nodes, each node in the routing table being a node that the computer

system can send a message to to delivery the message to a destination node within the overlay network, each node in the routing table having a fitness metric value representing an ability of the node to transfer and process messages within the overlay network (act 202). For example, routing table manager 104 can access routing table 102. Each node in routing table 102 is a routing node that node 101 can send a message to to delivery the message to a destination node (on ring 151). Each routing node in routing table 102 also has a fitness value. For example, within range 193: node ID 76 has fitness = 100, node ID 90 has fitness = 1000, node ID 99 has fitness = 50, and node ID 112 has fitness = 5000.

[0049] Prior to accessing routing table 102, routing table manager 104 can lock routing table 102. Locking routing table 102 prevents other access to routing table 102 (e.g.., routing determinations). Thus, locking mitigates the potential for routing errors due to accessing routing table 102 when routing table 102 is being altered.

[0050] Method 200 includes an act of calculating a fitness metric value for the other node, the fitness metric value representing the other node's ability to transfer and process messages within the overlay network, the fitness metric value based at least in part on the fitness information for the other node (act 203). For example, fitness calculation module 103 can calculate fitness value 107 for node 121 based at least in part on node information 107. Fitness value 107 can be calculated based on one or more characteristics of node 121, including routing table size. Fitness metric value 107 indicates the ability of node 121 to transfer and process messages within ring 151.

[0051] Method 200 includes an act of inserting the other node into the routing table (act 204). For example, routing table manager 104 can insert node ID 97 into routing table 102. Method 200 includes an act of dividing the routing table into a plurality of ranges, each range corresponding to a portion of the overlay network (act 205). For example, routing table manager 104 can divide routing table 102 into ranges 191 – 198.

[0052] Method 200 includes an act of assigning each node in the routing table to a specified range based on the location of the node in the overlay network (act 206).

For example, routing table manager can assign each routing node of routing table 102 to a specified range, selected from among ranges 191-198, based on the location of the routing node within ring 151. Routing table 102 can be divided and routing nodes assigned to ranges as depicted in Figure 1. Additionally, node 121 can be assigned to range 193 as depicted in Figure 1B.

[0053] Method 200 includes an act of identifying the range that includes the most nodes (act 207). For example, routing table manager 104 can identify that range 193 includes the most nodes. Range 193 includes five nodes, while ranges 192, and 194-197 include four nodes (neighborhood nodes can be excluded from the identification).

[0054] Method 200 includes an act of identifying the node within the identified range that is least able to transfer and process messages within the overly network based on fitness metric values of the nodes in the identified range (act 208). For example, routing tale manager 104 can identify node ID 99 as having the lowest fitness metric value in range 193. Within computer architecture 100, a lower fitness metric value can indicate a lesser ability to transfer and process messages. Accordingly, routing table manager 104 can identify node ID 99 as the routing node least able to transfer and process messages within ring 151. (However, depending on configuration, a higher fitness metric value, or some other mechanism for determining messaging processing and transfer abilities from fitness metric values, can be used to indicate a lesser ability to transfer and process messages).

[0055] Method 200 includes an act of removing the identified node from the routing table (act 209). For example, routing table manager 104 can remove node ID 99 from routing table 102. Subsequent, to removing node ID 99, routing table manager 104 can unlock routing table 102. Unlocking routing table 102 permits other access to routing table 102. Thus, node 101 can used routing table 102 to identify routing nodes after any alteration is complete.

[0056] In some embodiments, a node receives a message that includes node information for a plurality of other nodes. Figures 1C and 1D illustrates another view of node 101 maintaining a routing table when node information for a plurality

of nodes is received. Figure 3 illustrates a flow chart of a method 300 for maintaining a routing table. Method 300 will be described with respect to the components and data in Figures 1C and 1D.

[0057] Method 300 includes an act of receiving a message from another node in the overlay network, the message including node information for a plurality of further nodes in the overlay network, the node information including fitness information for each of the plurality of further nodes (act 301). For example, node 101 can receive message 111 from another node on ring 151. Message 111 includes node information for a plurality of other nodes on ring 151, including fitness information for each of the plurality of nodes. For example, message 111 includes routing table size and other characteristics (e.g., number of sent messages, number of received messages, etc.) for node IDs 46, 144, 242, and 460. The contents of message 111 can represent the other node's routing table or a portion thereof.

[0058] Method 300 includes an act of accessing the computer system's routing table, the computer system's routing table including a plurality of nodes that the computer system can communicate with to route messages to destination nodes within the overlay network, each of the plurality of nodes in the computer system's routing table having a fitness metric value representing an ability to transfer and process messages within the overlay network (act 302). For example, routing table manger 104 can access routing table 102. Each routing node in routing table 102 has a fitness metric value represent the ability of the routing node to transfer and process messages within ring 151.

[0059] Prior to accessing routing table 102, routing table manager 104 can lock routing table 102. Locking routing table 102 prevents other access to routing table 102 (e.g., routing determinations). Thus, locking mitigates the potential for routing errors due to accessing routing table 102 when routing table 102 is being altered.

[0060] For each of the plurality of further nodes, method 300 includes an act of calculating a fitness metric value for the further node, the fitness metric value representing the further node's ability to transfer and process messages within the overlay network, the fitness metric value based at least in part on the fitness

information for the further node (act 303). For example, fitness calculation module 103 can calculate fitness metric values 40, 200, 2000, and 500 for node IDs 46, 144, 242, and 460 respectively. Fitness metric values can be calculated based on routing table sizes and/or other characteristics indicated in message 111.

[0061] For each of the plurality of further nodes, method 300 includes an act of inserting the further node into the computer system's routing table (act 304). For example, routing table manager 104 can insert node IDs 46, 144, 242, and 460 into routing table 102.

[0062] Method 300 includes an act of determining that the number of nodes in the computer system's routing table exceeds a specified number (act 305). For example, routing table manager 104 can refer to routing table policy 161 to access a maximum routing node count for routing tables 102. Routing table manger 104 can determine that the insertion of node IDs 46, 144, 242, and 460 has caused the number of nodes in routing table 102 to exceed the maximum routing node count.

[0063] Method 300 includes an act of dividing the routing table into a plurality of ranges, each range corresponding to a portion of the overlay network (act 306). For example, routing node manager 104 can divide routing table 102 into ranges 191-198. Method 300 includes an act of assigning each node in the routing table to one of the plurality of ranges based on the location of the node in the overlay network (act 307). For example, each of the nodes in routing table 102 can be assigned to one of the ranges 191-198 based on their location in ring 151. Node ID 46 can be assigned to range 193, node IDs 144 and 242 can be assigned to range 195, and node ID 460 can assigned to range 197.

[0064] Method 300 includes an act of identifying the range with the most nodes (act 308). For example, routing table manager can identify range 194 as including the most (six) routing nodes. Method 300 includes of removing the node that is least fit to transfer and process messages, among the nodes in the identified range, from the routing table based on fitness metric values (act 309). For example, routing table manager 102 can remove node ID 135 from routing table 102. Routing table manager 104 can identify node ID 135 as having the lowest fitness metric value in range 193. Accordingly, routing table manager 104 determines that

node ID 135 is the routing node least able to transfer and process messages within ring 151. (However, as previously described depending on configuration, a higher fitness metric value, or some other mechanism for determining messaging processing and transfer abilities from fitness metric values, can be used to indicate a lesser ability to transfer and process messages).

[0065] In some embodiments, acts 308 and 309 can be repeated until the number of nodes routing table 102 no longer exceeds the specified number. For example, after the removal of a node, routing table manager 104 can again check the number of nodes in routing table 102 to the maximum routing node count. If the number of routing nodes in routing table 102 still exceeds the maximum routing node count, acts 308 and 309 can be repeated.

[0066] For example, after removal of node ID 135, routing table manager 104 can determine that range 192 includes five nodes. Routing tale manager 104 can identify node ID 37 the routing node least able to transfer and process messages within ring 151 based on fitness metric values. Accordingly, routing table manager 104 removes node ID 37 from routing table 102. Subsequent similar checks can be performed resulting in the removal of node ID 144 (in range 193) and node ID 463 (in range 197). Thus, routing table 102 is eventually returned to the size before message 111 was received. Figure 1D shows the resulting contents of routing table 102 subsequent to processing message 111.

[0067] Subsequent to updating routing table 102, routing table manager 104 can unlock routing table 102. Unlocking routing table 102 permits other access to routing table 102. Thus, node 101 can used routing table 102 to identify routing nodes after any alteration is complete. As such, subsequent to processing message 111, node 101 can send a message towards a destination node in ring 151. Based on the location of the destination node, node 101 can select a routing node (from routing table 102) in closer proximity to the destination node and send the message to the proximally closer routing node.

[0068] Embodiments of the invention significantly improve the likelihood that routing nodes contained in routing table have adequate (or even relatively increased) ability to transfer and process messages in an overlay network. Thus,

when the node is to make a routing decision for a message, the node has some assurances that any selected routing node is adequate (or is at least the best currently available). Further, a sending node can take preference to routing nodes with higher fitness metric values when sending a message. Preference to higher fitness metric values further insures that messages are adequately transferred and processed. Accordingly, embodiments of the invention can be used to route messages in a manner that optimizes bandwidth and provides efficient routing capability.

[0069] The present invention may be embodied in other specific forms without departing from its spirit or essential characteristics. The described embodiments are to be considered in all respects only as illustrative and not restrictive. The scope of the invention is, therefore, indicated by the appended claims rather than by the foregoing description. All changes which come within the meaning and range of equivalency of the claims are to be embraced within their scope.

CLAIMS

What is claimed:

1. At a computer system, the computer system included as a node (101) in an overlay network (100), the overlay network (100) also including a plurality of other nodes (121), a method for maintaining a routing table at the computer system, the method comprising:

an act receiving node information (107) for another node (121) that exists at a specified location within the overlay network (100), the node information (107) including fitness information for the other node (121);

an act of accessing a routing table (102) that includes one or more nodes, each node in the routing table (102) being a node that the computer system can send a message to to delivery the message to a destination node within the overlay network (100), each node in the routing table (102) having a fitness metric value representing an ability of the node to transfer and process messages within the overlay network (100);

an act of calculating a fitness metric value (107) for the other node, the fitness metric value (107) representing the other node's ability to transfer and process messages within the overlay network (100), the fitness metric value (107) based at least in part on the fitness information (107) for the other node;

an act of inserting the other node (121) into the routing table (102);

an act of dividing the routing table into a plurality of ranges (191, 192, 193), each range corresponding to a portion of the overlay network (100);

an act of assigning each node in the routing table to a specified range based on the location of the node in the overlay network;

an act of identifying the range (194) that includes the most nodes;

an act of identifying the node within the identified range that is least able to transfer and process messages within the overly network based on fitness metric values of the nodes in the identified range; and

an act of removing the identified node from the routing table (102).

2. The method as recited in claim 1, wherein the act of receiving node information for another node comprises an act of receiving a node table size for the other node along with one or more other characteristics of the other node.

3. The method as recited in claim 1, wherein the act accessing a routing table comprises an act out accessing a routing table containing nodes that can be used to route a message in a ring topology.

4. The method as recited in claim 1, wherein the act of calculating a fitness metric value for the other node comprises an act of calculating a fitness metric value based on the routing table size of the other node.

5. The method as recited in claim 1, wherein the act of dividing the routing table into a plurality of ranges comprises an act of dividing the routing table in ranges, each range corresponding to an exponentially smaller or exponentially larger portion of the overlay network depending on the distance of the range from the computer system, ranges closer to the computer being smaller, and range further from the computer system being larger.

6. The method as recited in claim 1, wherein an act of identifying the node within the identified range that is least able to transfer and process messages within the overly network comprises an act of identifying the node with the smallest routing table.

7. The method as recited in claim 1, wherein the act of removing the identified node from the routing table comprises an act of removing the node having the smallest routing table from the computer system's routing table.

8. The method as recited in claim 1, further comprising:

an act of locking the routing table prior to accessing the routing table; and

an act of unlock the routing table subsequent to removing the identified node.

9. The method as recited in claim 1, further comprising:

an act of selecting a routing node from the routing table subsequent to removing the identified node; and

an act of sending a message to the selected routing node.

10. At a computer system, the computer system included as a node (101) in an overlay network, the overlay network also including a plurality of other nodes, a method for maintaining a routing table (102) at the computer system, the method comprising:

an act of receiving a message (111) from another node in the overlay network, the message (111) including node information for a plurality of further nodes in the overlay network (100), the node information including fitness information for each of the plurality of further nodes;

an act of accessing the computer system's routing table, the computer system's routing table (102) including a plurality of nodes that the computer system can communicate with to route messages to destination nodes within the overlay network (100), each of the plurality of nodes in the computer system's routing table (102) having a fitness metric value representing an ability to transfer and process messages within the overlay network;

for each of the plurality of further nodes:

an act of calculating a fitness metric value for the further node, the fitness metric value representing the further node's ability to transfer and process messages within the overlay network (100), the fitness metric value based at least in part on the fitness information for the further node; and

an act of inserting the further node into the computer system's routing table (102);

an act of determining that the number of nodes in the computer system's routing table exceeds a specified number (161);

an act of dividing the routing table into a plurality of ranges (191, 192, 193), each range corresponding to a portion of the overlay network (100);

an act of assigning each node in the routing table to one of the plurality of ranges based on the location of the node in the overlay network (100);

an act of identifying the range that includes the most nodes (194); and

an act of removing the node that is least fit to transfer and process messages, among the nodes in the identified range, from the routing table (102) based on fitness metric values.

11. The method as recited in claim 10, wherein the an act of receiving a message from another node in the overlay network comprises an act of receiving a message that includes routing table size for each of the plurality of further nodes along with one or more other characteristics for each of the plurality of further nodes.

12. The method as recited in claim 10, wherein the act of calculating a fitness metric value for the further node comprises an act of calculating fitness metric based on routing table size of the further node.

13. The method as recited in claim 10, wherein the act of calculating a fitness metric value for the further node comprises an act of calculating fitness metric based solely on routing table size of the further node.

14. The method as recited in claim 10, wherein the act of determining that the number of nodes in the computer system's routing table exceeds a specified number comprises an act of referring to a routing table policy.

15. The method as recited in claim 10, further comprising subsequent to removing the node that is least fit to transfer and process messages from among the nodes in the identified range:

repeating until the number of nodes in the computer system's routing table no longer exceeds the specified number:

an act of identifying a next range that includes the most nodes; and

an act of removing the node that is least fit to transfer and process messages, among the nodes in the next identified range, from the routing table, based on fitness metric values.

16. The method as recited in claim 10, further comprising:

an act of locking the routing table prior to accessing the routing table; and

an act of unlock the routing table subsequent to removing the identified node.

17. The method as recited in claim 10, further comprising:

an act of selecting a routing node from the routing subsequent to removing the identified node; and

an act of sending a message to the selected routing node.

18. The method as recited in claim 10, wherein the overlay network is a ring topology.

19. The method as recited in claim 10, wherein the overlay network is peer-to-peer network overlaid on the Internet.

20. At a computer system, the computer system included as a node in an overlay network having a doubly linked ring topology, the overlay network also including a plurality of other nodes, a method for maintaining a routing table at the computer system, the method comprising:

an act of receiving a message from another node in the doubly linked ring topology, the message including node information for a plurality of further nodes in the other node's routing table, the node information indicating routing table size for each further node in the other node's routing table;

an act of locking the computer system's routing table to prevent any routing determinations subsequent to receiving the message, each node in the computer system's routing table being a node that the computer system can send messages to to delivery the messages to a destination node within the doubly linked ring topology, each of the plurality of nodes in the computer system's routing table stored along with a an indication of a corresponding routing table size of the node;

an act of integrating the plurality of further nodes into the computer system's routing table;

an act of determining that the number of nodes in the computer system's routing table exceeds a specified number;

an act of dividing the routing table into a plurality of ranges, each range corresponding to a portion of doubly linked ring topology;

an act of assigning each node in the routing table to one of the plurality of ranges based on the location of the node in the doubly linked ring topology;

repeating until the number of nodes in the computer system's routing table no longer exceeds the specified number:

an act of identifying the range that includes the most nodes;

an act of identifying the node within the identified range that has the smallest indicated routing table size; and

an act of removing the identified node from the computer system's routing table; and

an act of unlocking the computer system's routing table so that the computer system's routing table can again be used in routing determinations subsequent to the number of nodes in the computer system's routing table no longer exceeding the specified number.

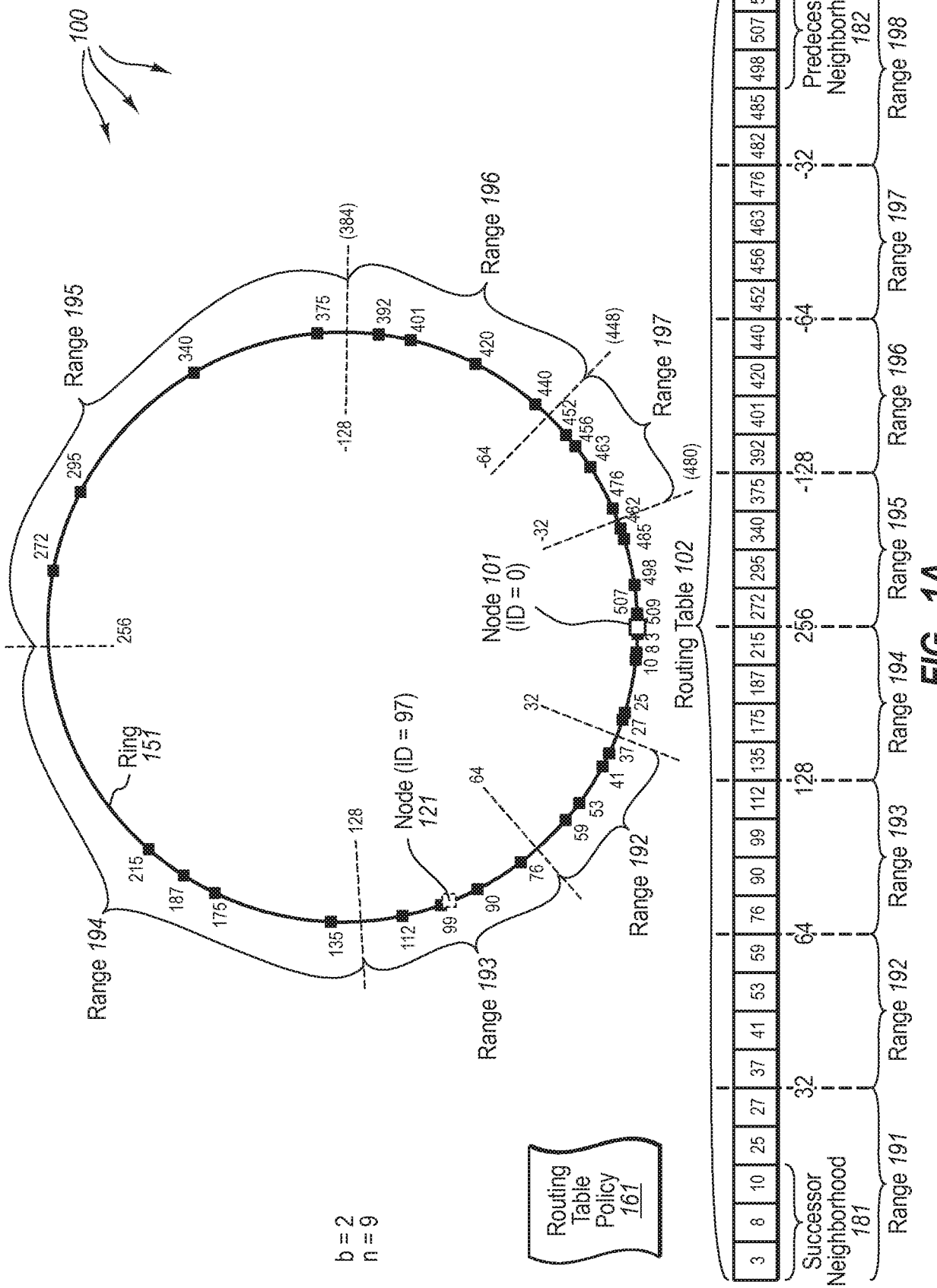

100
Range 195
Range 196
Range 197
Range 198
Range 194
Range 193
Range 192
Range 191
Ring 151
Node (ID = 97) 121
Node 101 (ID = 0)
Routing Table 102
Routing Table Policy 161
b = 2
n = 9
Successor Neighborhood 181
Predecessor Neighborhood 182
3 8 10 25 27 37 41 53 59 76 90 99 112 135 175 187 215 272 295 340 375 392 401 420 440 452 456 463 476 482 485 498 507 509
32 64 128 256 -128 -64 -32

FIG. 1A

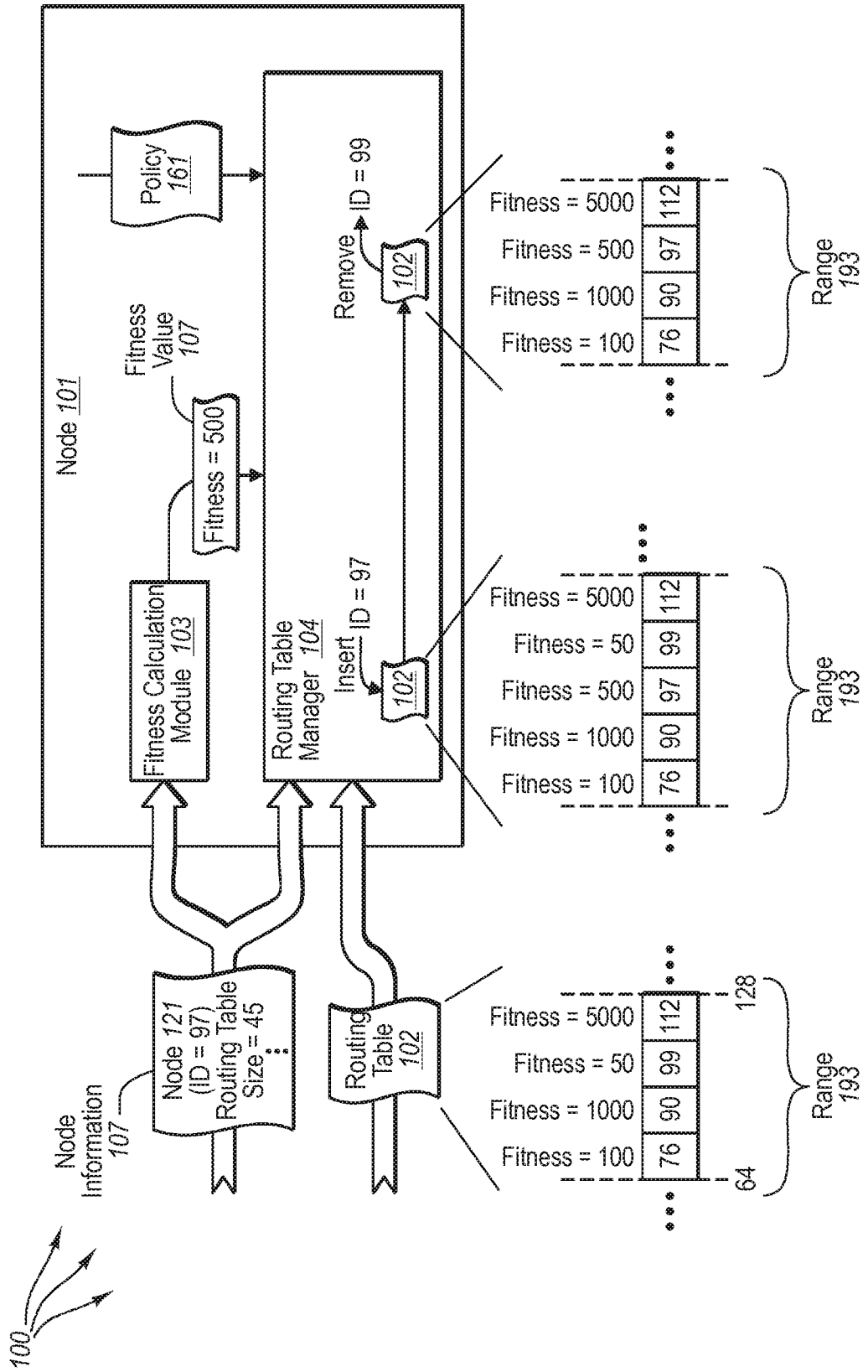
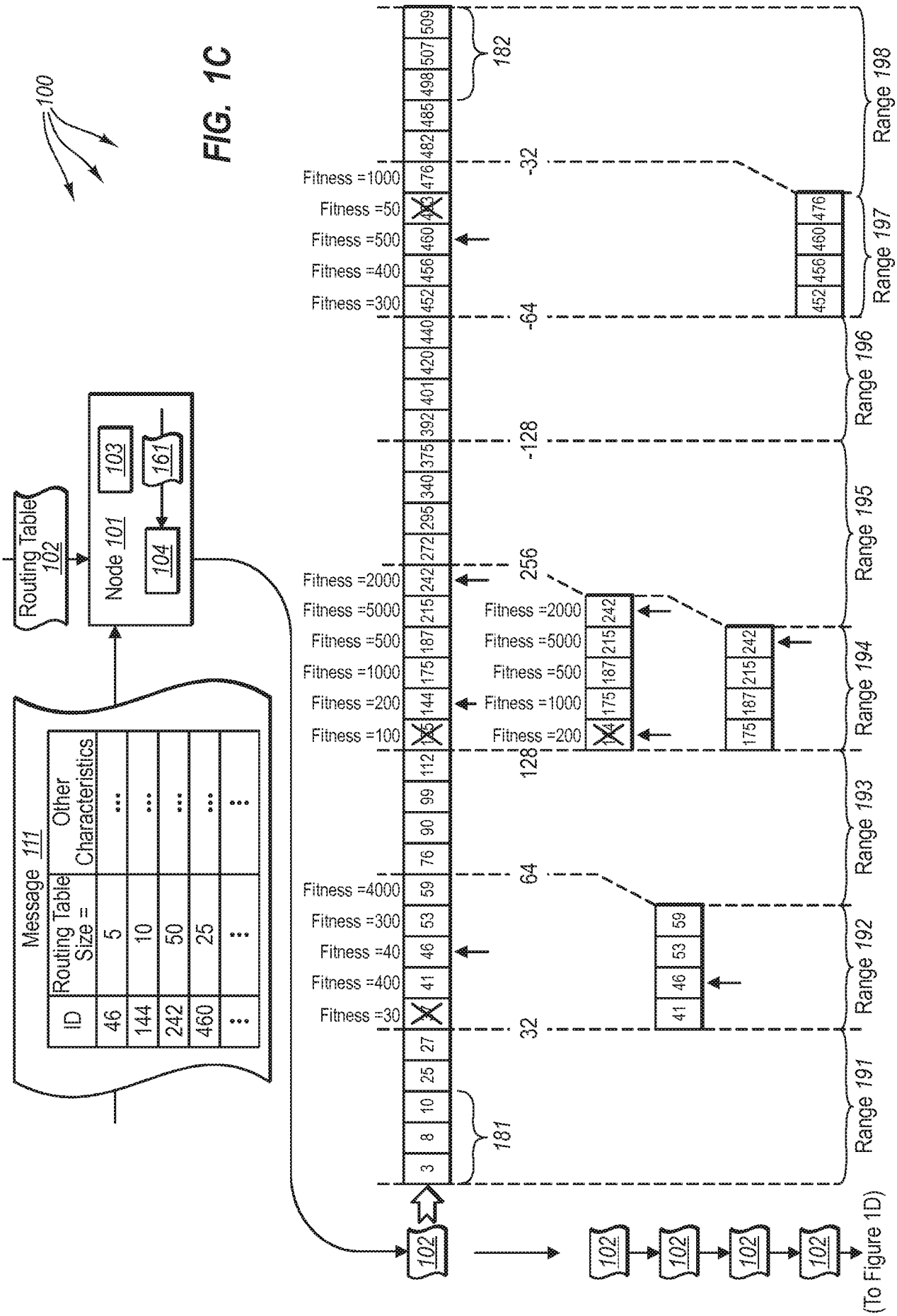

100
Node Information 107
Node 121 (ID = 97) Routing Table Size = 45
Routing Table 102
Node 101
Fitness Value 107
Fitness = 500
Fitness Calculation Module 103
Routing Table Manager 104
Insert ID = 97
102
Remove ID = 99
102
Policy 161
Fitness = 5000
Fitness = 50
Fitness = 500
Fitness = 1000
Fitness = 100
76
90
97
99
112
128
64
Range 193

FIG. 1B

FIG. 1C


100
Routing Table 102
Node 101
103
161
104
Message 111
ID
Routing Table Size =
Other Characteristics
46 5 ...
144 10 ...
242 50 ...
460 25 ...
... ...
Fitness =1000
Fitness =50
Fitness =500
Fitness =400
Fitness =300
Fitness =2000
Fitness =5000
Fitness =500
Fitness =1000
Fitness =200
Fitness =100
Fitness =4000
Fitness =300
Fitness =40
Fitness =400
Fitness =30
3 8 10 25 27 41 46 53 59 76 90 99 112 144 175 187 215 242 272 295 340 375 392 401 420 440 452 456 460 476 482 485 498 507 509
181
182
32
64
128
256
-32
-64
-128
Range 191
Range 192
Range 193
Range 194
Range 195
Range 196
Range 197
Range 198
102
(To Figure 1D)

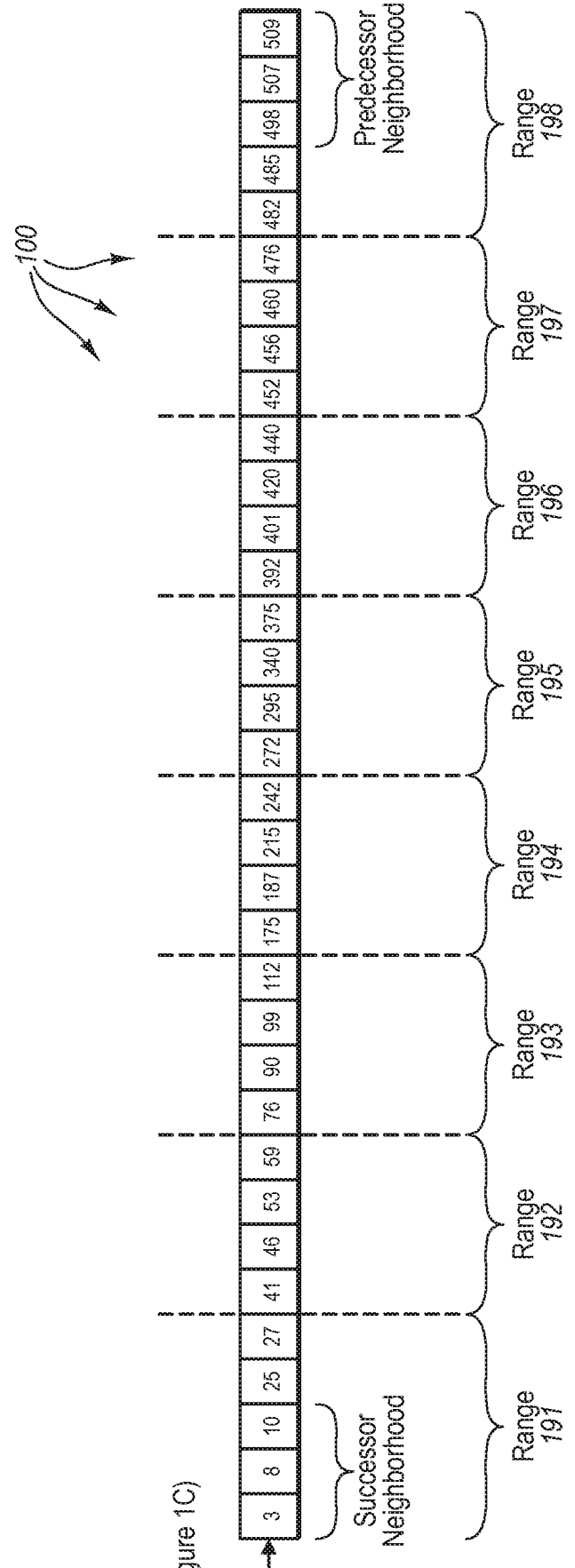

100
(From Figure 1C)
102
3
8
10
25
27
41
46
53
59
76
90
99
112
175
187
215
242
272
295
340
375
392
401
420
440
452
456
460
476
482
485
498
507
509
Successor Neighborhood
Predecessor Neighborhood
Range 191
Range 192
Range 193
Range 194
Range 195
Range 196
Range 197
Range 198

FIG. 1D

ms

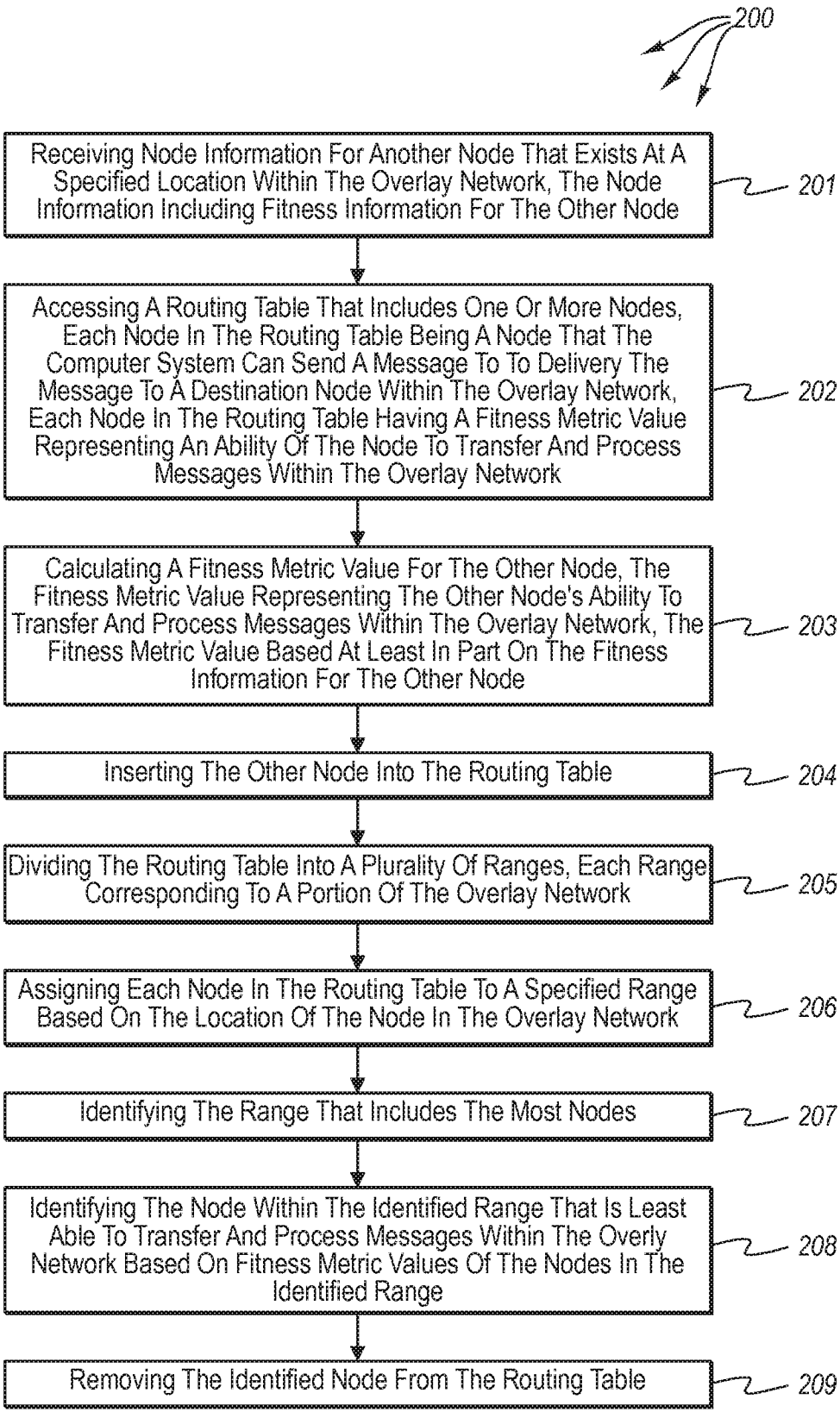

200
Receiving Node Information For Another Node That Exists At A Specified Location Within The Overlay Network, The Node Information Including Fitness Information For The Other Node
201
Accessing A Routing Table That Includes One Or More Nodes, Each Node In The Routing Table Being A Node That The Computer System Can Send A Message To To Delivery The Message To A Destination Node Within The Overlay Network, Each Node In The Routing Table Having A Fitness Metric Value Representing An Ability Of The Node To Transfer And Process Messages Within The Overlay Network
202
Calculating A Fitness Metric Value For The Other Node, The Fitness Metric Value Representing The Other Node's Ability To Transfer And Process Messages Within The Overlay Network, The Fitness Metric Value Based At Least In Part On The Fitness Information For The Other Node
203
Inserting The Other Node Into The Routing Table
204
Dividing The Routing Table Into A Plurality Of Ranges, Each Range Corresponding To A Portion Of The Overlay Network
205
Assigning Each Node In The Routing Table To A Specified Range Based On The Location Of The Node In The Overlay Network
206
Identifying The Range That Includes The Most Nodes
207
Identifying The Node Within The Identified Range That Is Least Able To Transfer And Process Messages Within The Overly Network Based On Fitness Metric Values Of The Nodes In The Identified Range
208
Removing The Identified Node From The Routing Table
209

FIG. 2

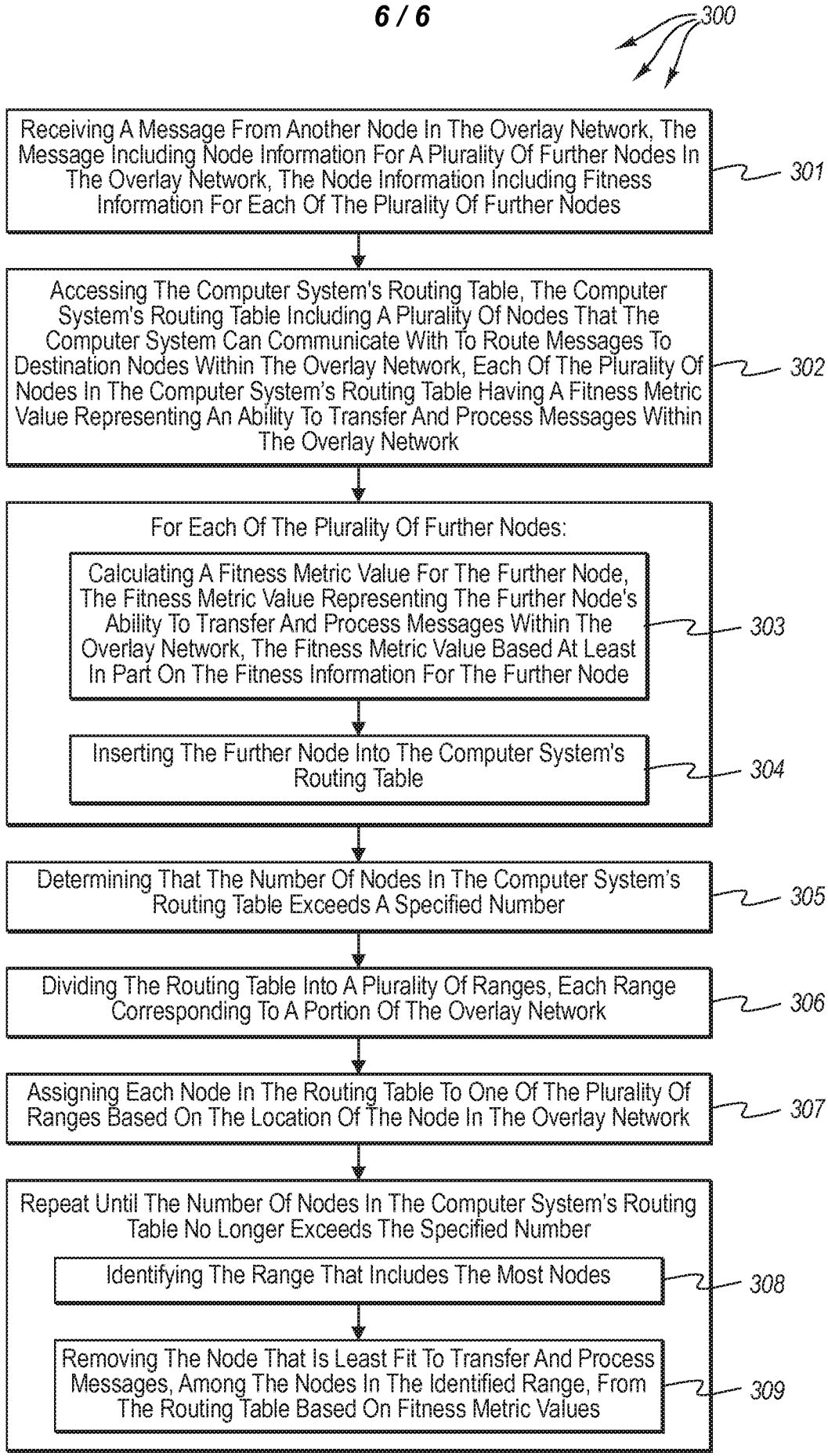

300
Receiving A Message From Another Node In The Overlay Network, The Message Including Node Information For A Plurality Of Further Nodes In The Overlay Network, The Node Information Including Fitness Information For Each Of The Plurality Of Further Nodes
301
Accessing The Computer System's Routing Table, The Computer System's Routing Table Including A Plurality Of Nodes That The Computer System Can Communicate With To Route Messages To Destination Nodes Within The Overlay Network, Each Of The Plurality Of Nodes In The Computer System's Routing Table Having A Fitness Metric Value Representing An Ability To Transfer And Process Messages Within The Overlay Network
302
For Each Of The Plurality Of Further Nodes:
Calculating A Fitness Metric Value For The Further Node, The Fitness Metric Value Representing The Further Node's Ability To Transfer And Process Messages Within The Overlay Network, The Fitness Metric Value Based At Least In Part On The Fitness Information For The Further Node
303
Inserting The Further Node Into The Computer System's Routing Table
304
Determining That The Number Of Nodes In The Computer System's Routing Table Exceeds A Specified Number
305
Dividing The Routing Table Into A Plurality Of Ranges, Each Range Corresponding To A Portion Of The Overlay Network
306
Assigning Each Node In The Routing Table To One Of The Plurality Of Ranges Based On The Location Of The Node In The Overlay Network
307
Repeat Until The Number Of Nodes In The Computer System's Routing Table No Longer Exceeds The Specified Number
Identifying The Range That Includes The Most Nodes
308
Removing The Node That Is Least Fit To Transfer And Process Messages, Among The Nodes In The Identified Range, From The Routing Table Based On Fitness Metric Values
309

FIG. 3

ms

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2008/072250

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/28(2006.01)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)

IPC 8: H04L, H04B, H04Q

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 since 1975.
Japanese utility models and applications for utility models since 1975.

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

eKIPASS(KIPO internal), IEEEXplore: "overlay network", "routing", "routing table", "heterogeneous", "link capacity", "doubly link", "ring", "remov* node", "remov* entry"

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2006-246205 A (NIPPON TELEGR. & TELEPH. CORP.) 14 September 2006 See abstract; figure 1; page 4, paragraph 14 - page 6, paragraph 26	1-20
A	Zhi Li and Prasant Mohapatra, "QRON: QoS-Aware Routing in Overlay Networks," IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 22, NO. 1, pp. 29-39, January 2004	1-20
A	EP 1164753 A1 (TELEFONAKTIEBOLAGET L M ERICSSON) 19 December 2001 See abstract; figure 1; column 5, paragraph 19 - column 7, paragraph 39	1-20
A	EP 1164754 A1 (TELEFONAKTIEBOLAGET L M ERICSSON) 15 June 2000 See abstract; figure 1, column 5, paragraph 20 - column 6, paragraph 38	1-20
A	US 2007-0153782 A1 (FLETCHER, G. et al.) 5 June 2007 See abstract; figure 4; claim 1	1-20

☐ 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 citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 DECEMBER 2008 (18.12.2008)	**18 DECEMBER 2008 (18.12.2008)**

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Lee Hyoung Il

Telephone No. 82-42-481-8199

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2008/072250

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
JP 2006-246205 A	14.09.2006	None	
EP 1164753 A1	19.12.2001	None	
EP 1164754 A1	15.06.2000	None	
US 2007-0153782 A1	05.06.2007	None	