US 20160352631A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2016/0352631 A1**

Medved et al. (43) **Pub. Date: Dec. 1, 2016**

(54) **DYNAMICALLY GENERATING APPLICATION-LAYER TRAFFIC OPTIMIZATION PROTOCOL MAPS**

(71) Applicant: **Juniper Networks, Inc.**, Sunnyvale, CA (US)

(72) Inventors: **Jan Medved**, Pleasanton, CA (US); **Hannes Gredler**, Pillberg (AT); **David Ward**, Los Gatos, CA (US); **Satish Raghunath**, Sunnyvale, CA (US)

(73) Assignee: **Juniper Networks, Inc.**

(21) Appl. No.: **15/231,525**

(22) Filed: **Aug. 8, 2016**

Related U.S. Application Data

(63) Continuation of application No. 14/252,526, filed on Apr. 14, 2014, now Pat. No. 9,413,847, which is a continuation of application No. 13/110,987, filed on May 19, 2011, now Pat. No. 8,700,801.

(60) Provisional application No. 61/418,793, filed on Dec. 1, 2010, provisional application No. 61/449,499, filed on Mar. 4, 2011.

Publication Classification

(51) **Int. Cl.**

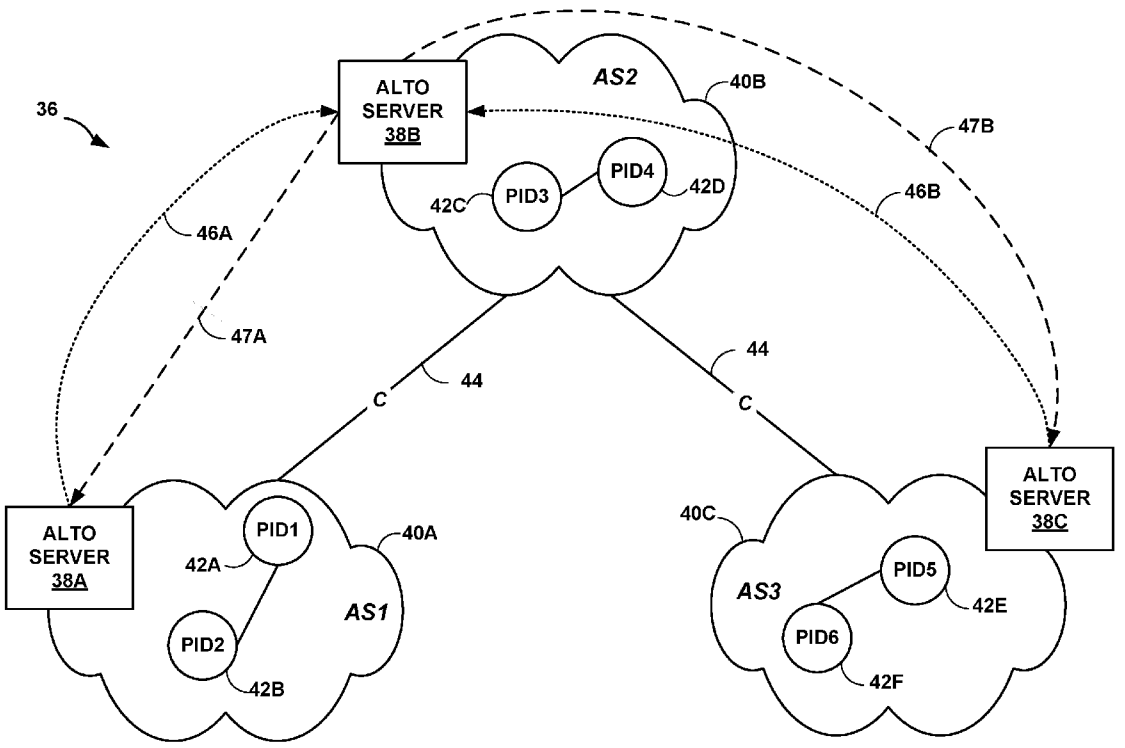
H04L 12/717	(2006.01)
H04L 12/751	(2006.01)
H04L 29/08	(2006.01)

(52) **U.S. Cl.**

CPC ***H04L 45/42*** (2013.01); ***H04L 69/329*** (2013.01); ***H04L 45/02*** (2013.01); ***H04L 69/321*** (2013.01)

(57) **ABSTRACT**

In general, techniques are described for using routing information obtained by operation of network routing protocols to dynamically generate network and cost maps for an application-layer traffic optimization (ALTO) service. For example, an ALTO server of an autonomous system (AS) receives routing information from routers of the AS by listening for routing protocol updates outputted by the routers and uses the received topology information to dynamically generate a network map of PIDs that reflects a current topology of the AS and/or of the broader network that includes the AS. Additionally, the ALTO server dynamically calculates inter-PID costs using received routing information that reflects current link metrics. The ALTO server then assembles the inter-PID costs into a cost map that the ALTO server may provide, along with the network map, to clients of the ALTO service.



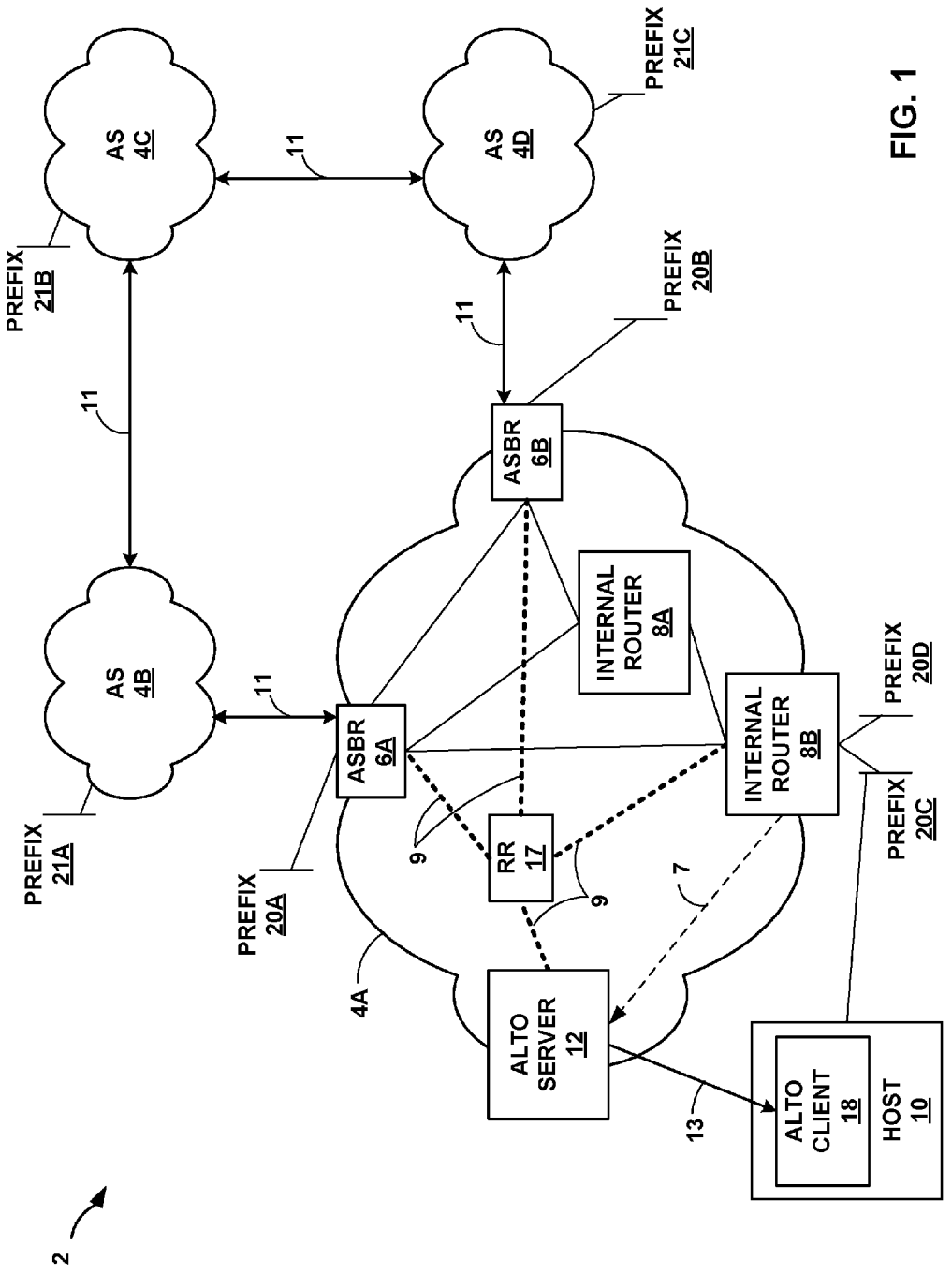

AS 4C
AS 4D
AS 4B
11
PREFIX 21B
PREFIX 21C
PREFIX 20B
ASBR 6B
INTERNAL ROUTER 8A
ASBR 6A
INTERNAL ROUTER 8B
PREFIX 20D
PREFIX 20C
PREFIX 21A
PREFIX 20A
9
RR 17
7
4A
ALTO SERVER 12
13
ALTO CLIENT 18
HOST 10
2

FIG. 1

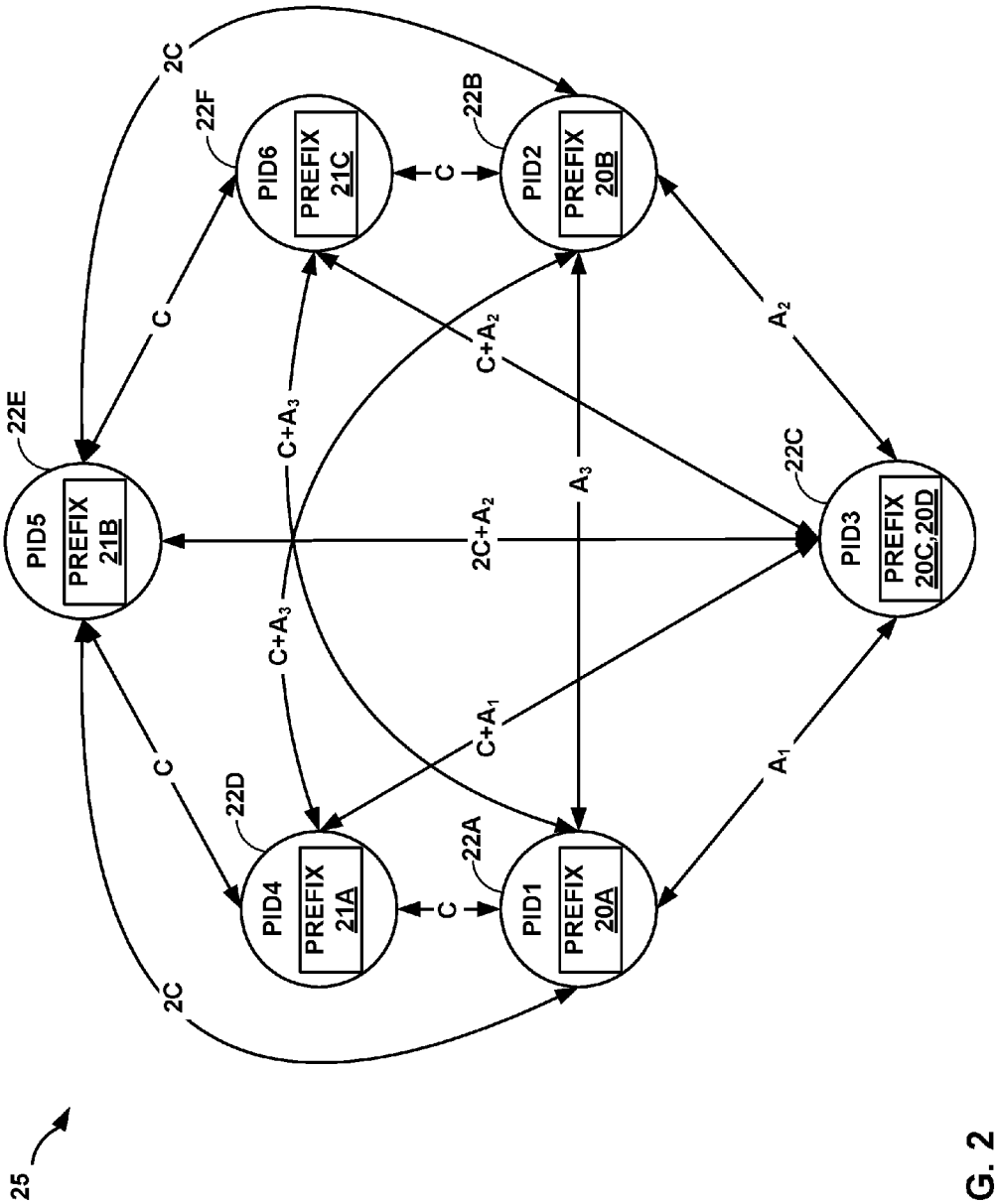

25
PID5
PREFIX
21B
22E
PID6
PREFIX
21C
22F
PID2
PREFIX
20B
22B
PID3
PREFIX
20C,20D
22C
PID4
PREFIX
21A
22D
PID1
PREFIX
20A
22A
2C
C
C+A3
C+A2
A2
A3
2C+A2
C+A3
C+A1
A1
2C
C
C

FIG. 2

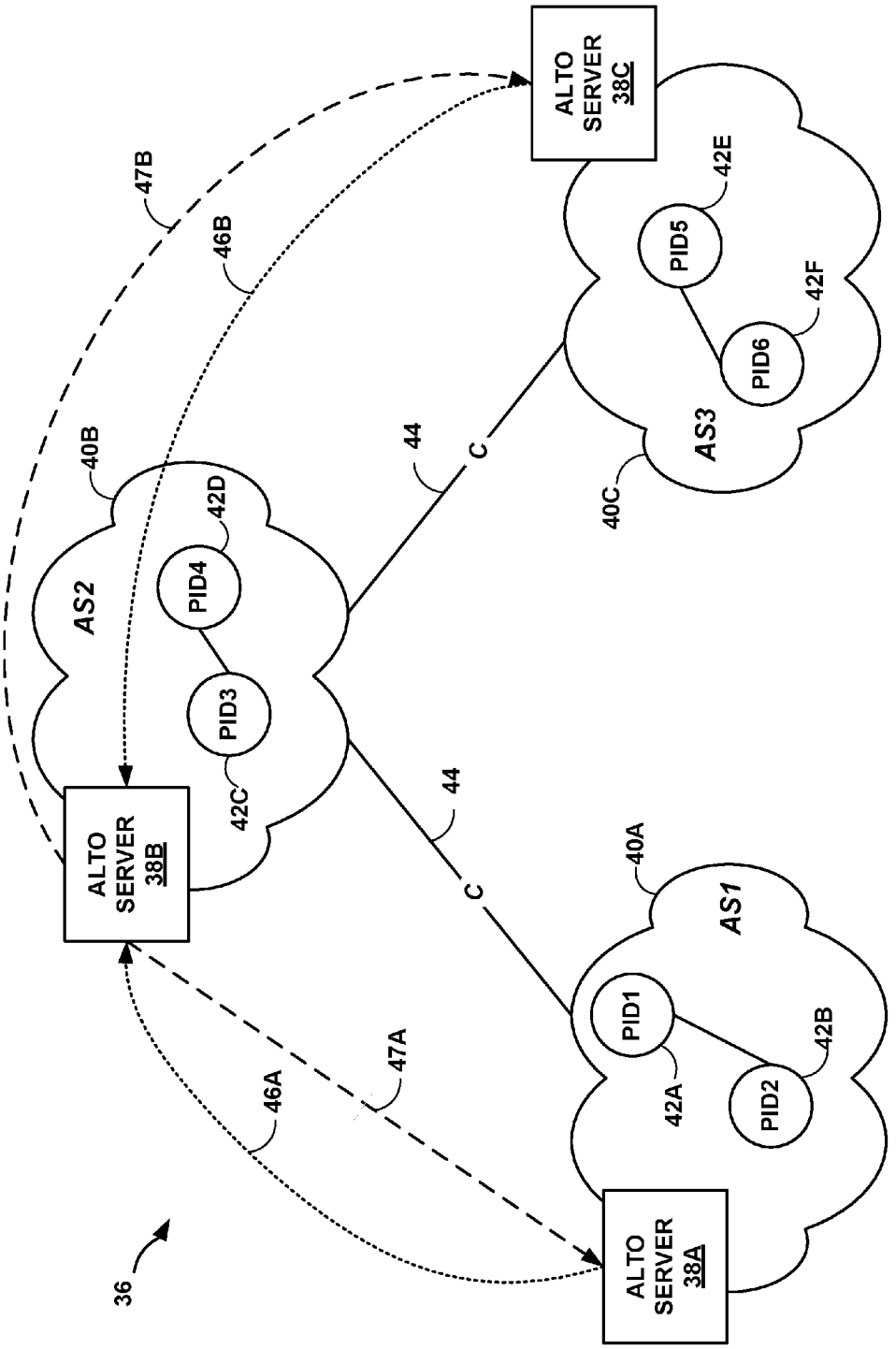

ALTO SERVER 38C
ALTO SERVER 38B
ALTO SERVER 38A
AS1
AS2
AS3
40A
40B
40C
PID1
PID2
PID3
PID4
PID5
PID6
42A
42B
42C
42D
42E
42F
44
C
44
C
47A
47B
46A
46B
36

FIG. 3

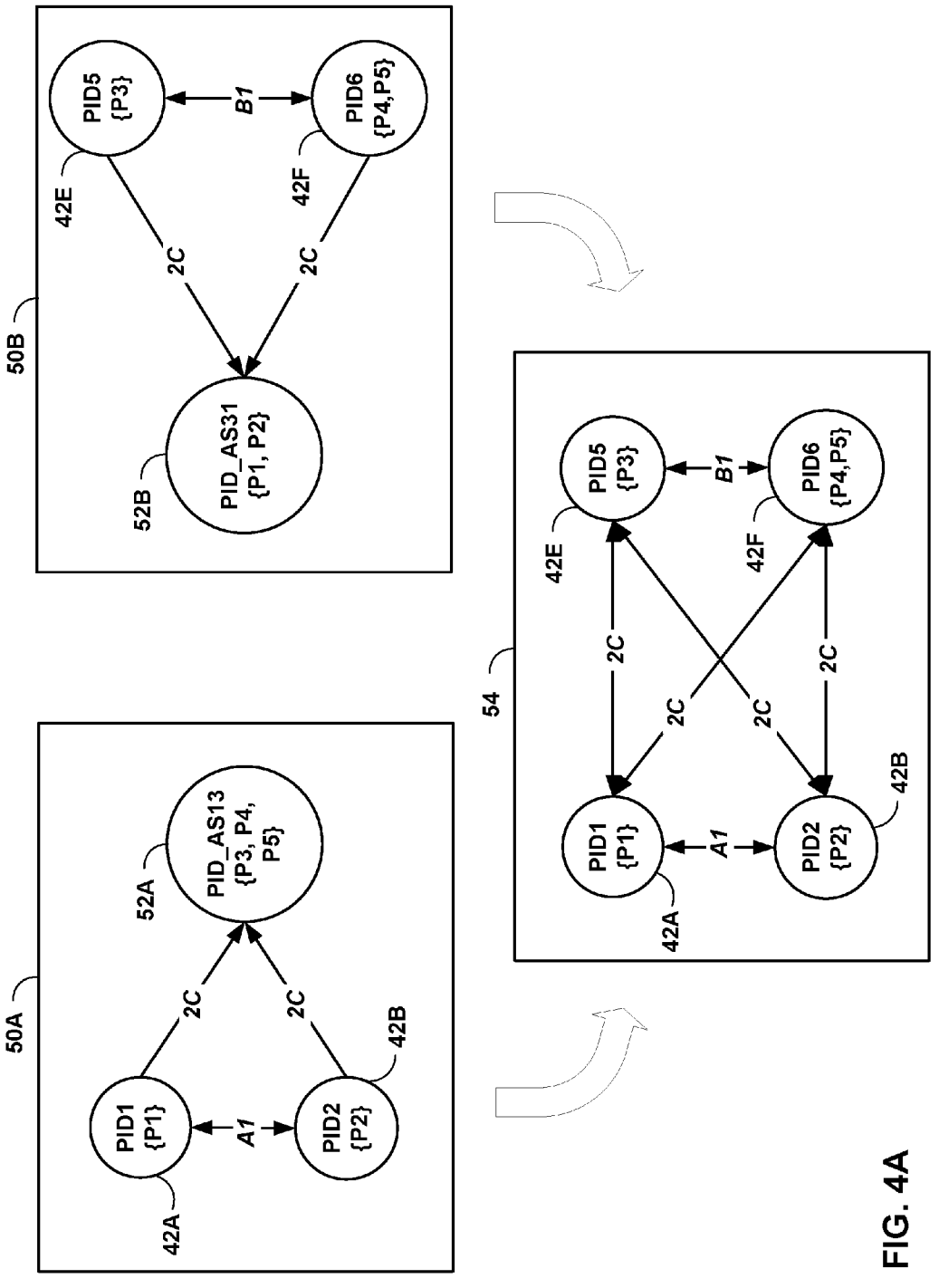

50A
PID1 {P1}
42A
A1
PID2 {P2}
42B
2C
2C
PID_AS13 {P3, P4, P5}
52A
50B
PID5 {P3}
42E
B1
PID6 {P4,P5}
42F
2C
2C
PID_AS31 {P1, P2}
52B
54
PID5 {P3}
42E
B1
PID6 {P4,P5}
42F
2C
2C
2C
2C
PID1 {P1}
42A
A1
PID2 {P2}
42B

FIG. 4A

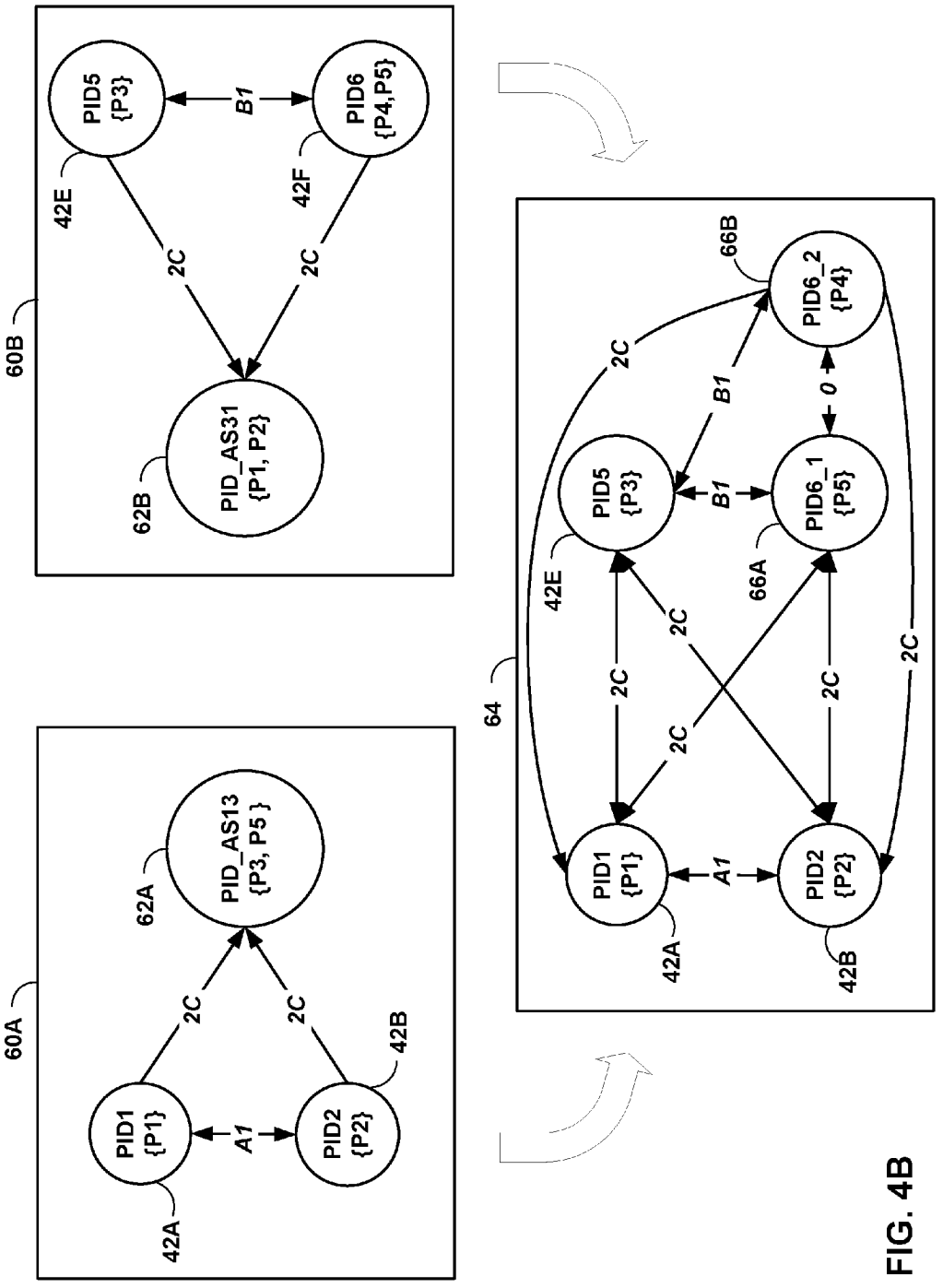

60A
PID_AS13
{P3, P5 }
62A
2C
2C
PID1
{P1}
A1
PID2
{P2}
42A
42B
60B
PID5
{P3}
B1
PID6
{P4,P5}
42E
42F
2C
2C
PID_AS31
{P1, P2}
62B
64
PID6_2
{P4}
66B
2C
B1
0
PID5
{P3}
B1
PID6_1
{P5}
42E
66A
2C
2C
2C
2C
PID1
{P1}
A1
PID2
{P2}
42A
42B

FIG. 4B

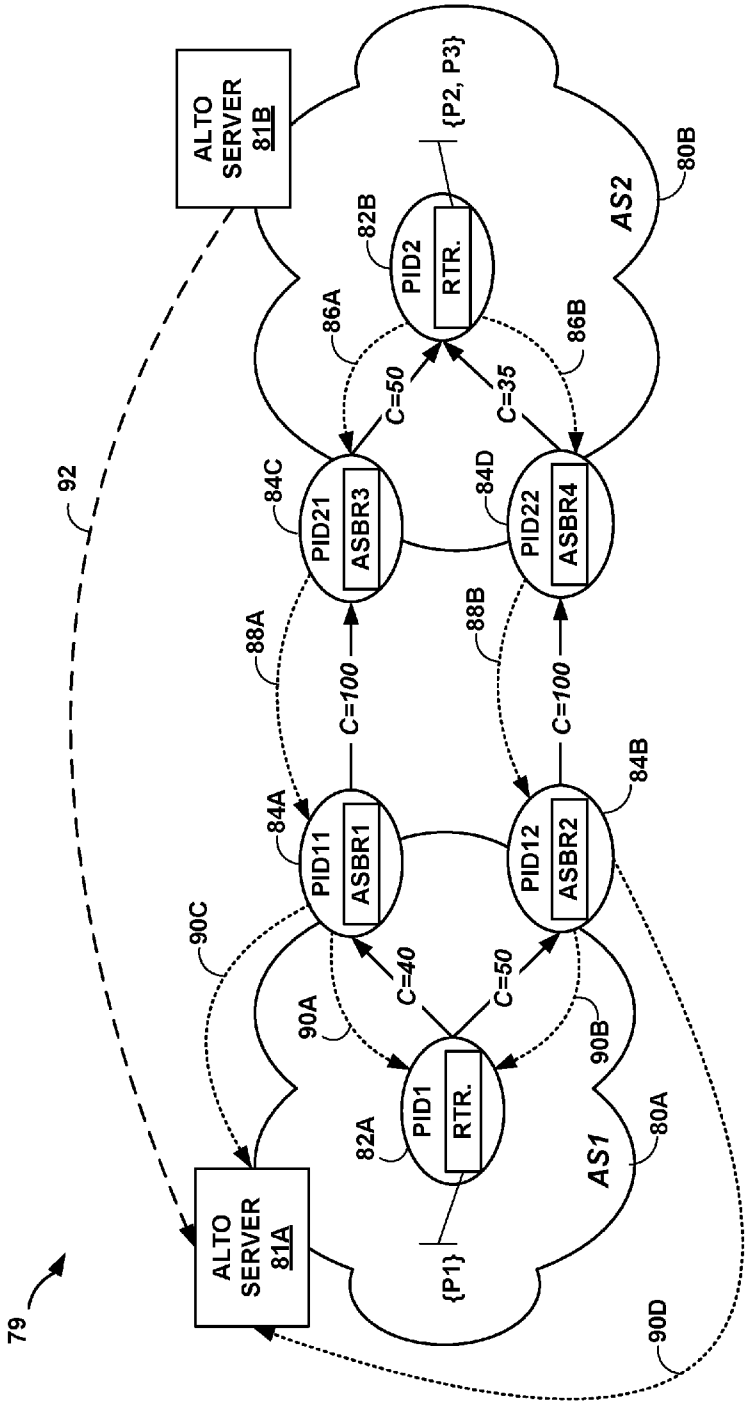

79
ALTO SERVER 81A
ALTO SERVER 81B
92
AS1
80A
AS2
80B
PID1
RTR.
82A
{P1}
PID2
RTR.
82B
{P2, P3}
PID11
ASBR1
84A
PID12
ASBR2
84B
PID21
ASBR3
84C
PID22
ASBR4
84D
C=40
C=50
C=50
C=35
C=100
C=100
86A
86B
88A
88B
90A
90B
90C
90D

FIG. 5

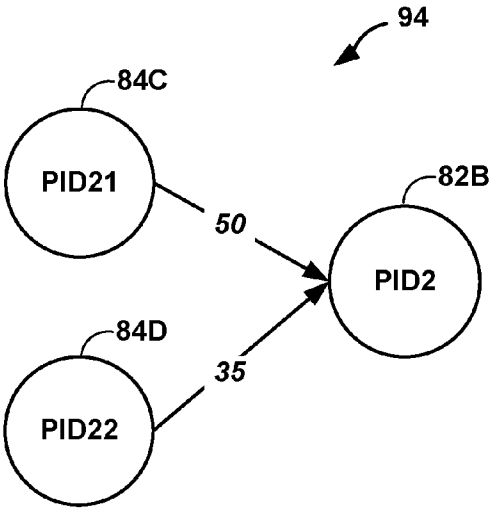

94
84C
PID21
82B
50
PID2
84D
35
PID22

FIG. 6A

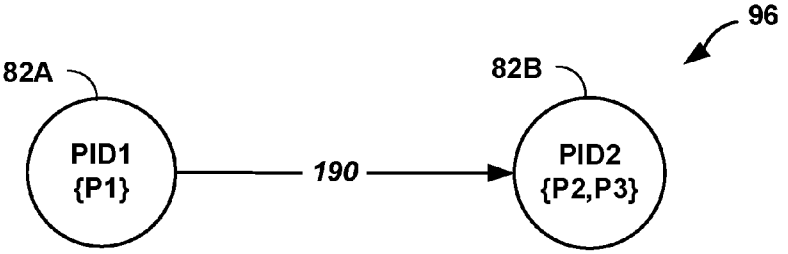

96
82A
82B
PID1
{P1}
190
PID2
{P2,P3}

FIG. 6B

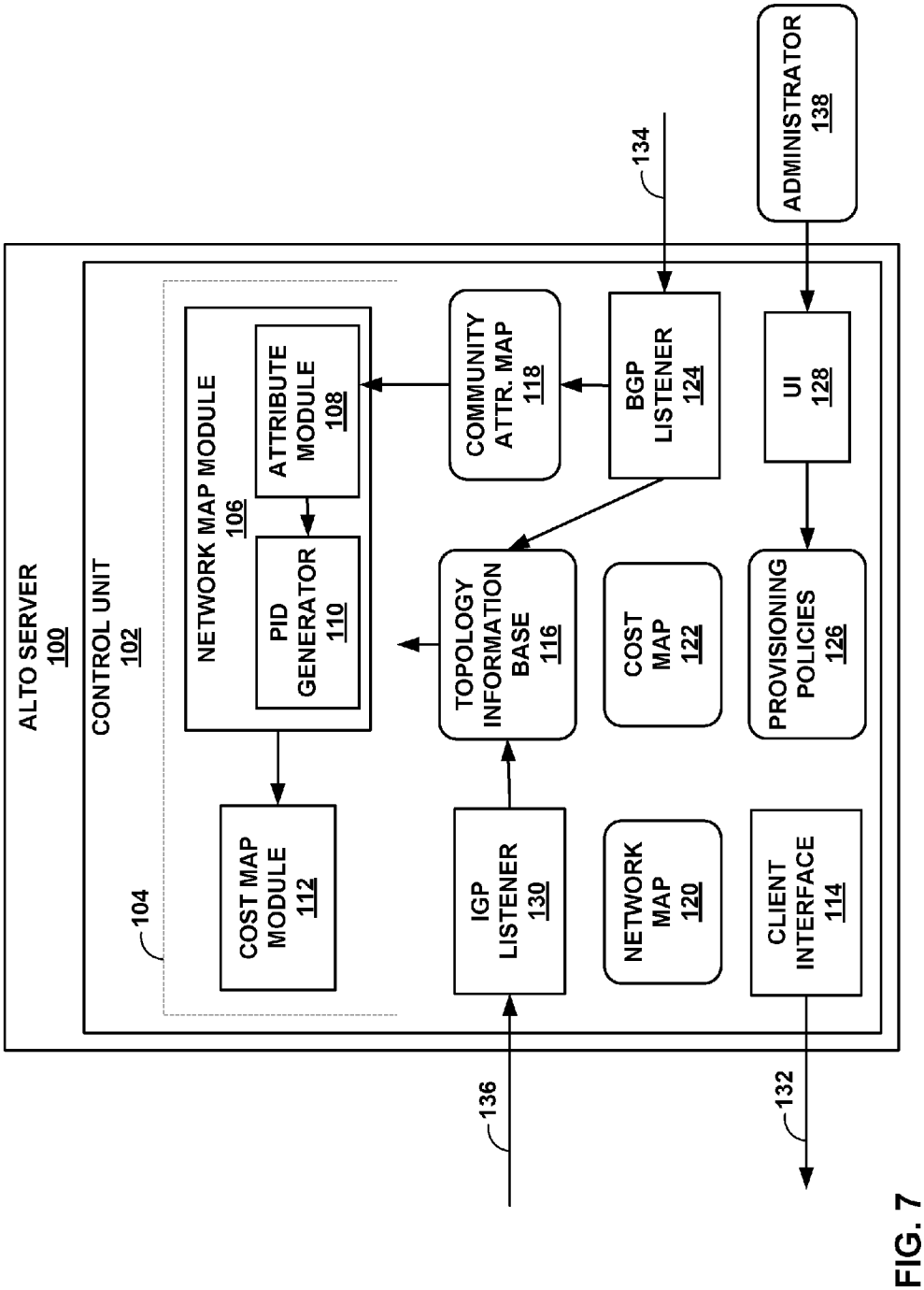

ALTO SERVER
100
CONTROL UNIT
102
104
NETWORK MAP MODULE
106
ATTRIBUTE MODULE
108
PID GENERATOR
110
COST MAP MODULE
112
COMMUNITY ATTR. MAP
118
BGP LISTENER
124
134
ADMINISTRATOR
138
UI
128
TOPOLOGY INFORMATION BASE
116
COST MAP
122
PROVISIONING POLICIES
126
IGP LISTENER
130
NETWORK MAP
120
CLIENT INTERFACE
114
136
132

FIG. 7

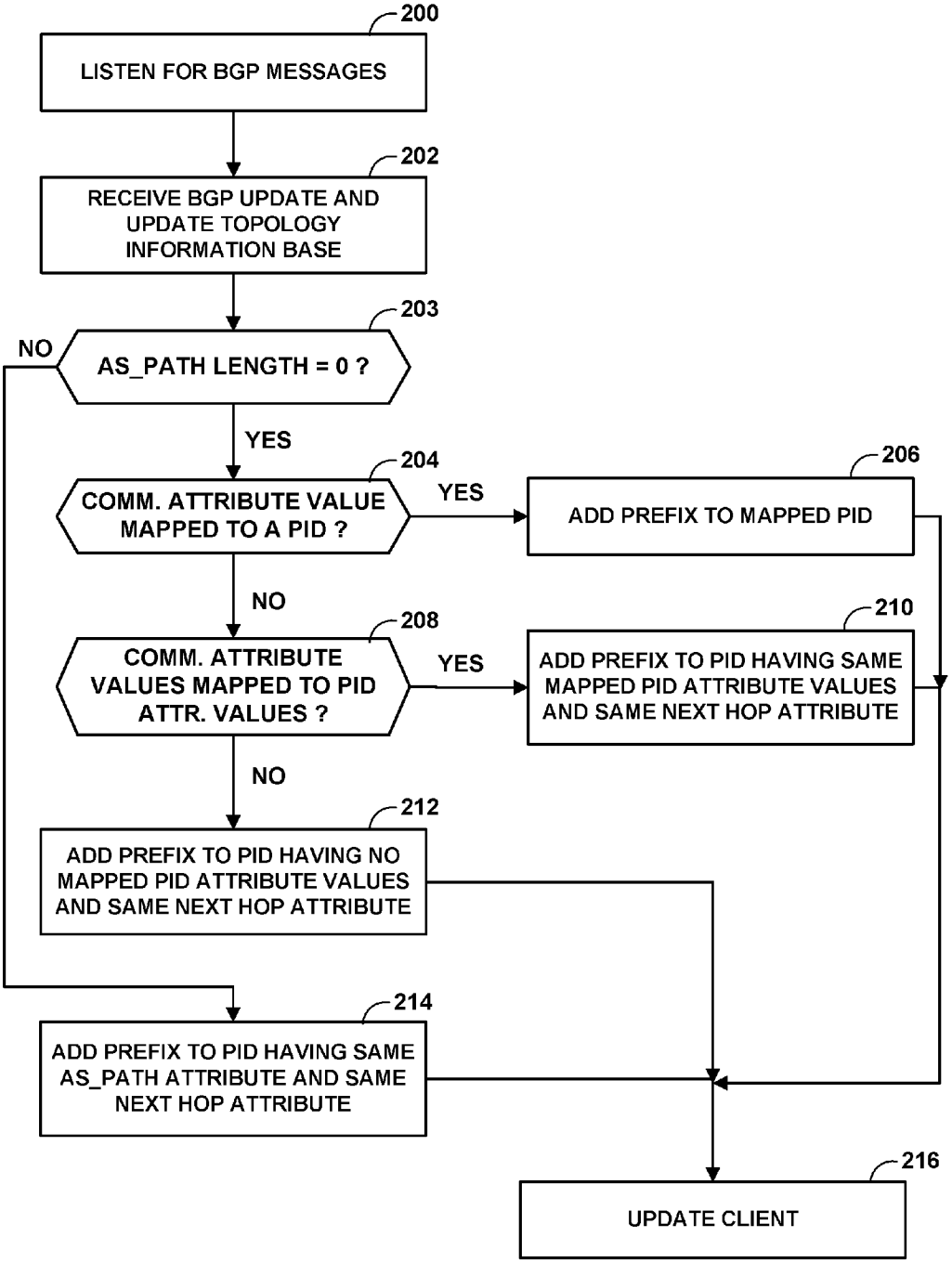

200
LISTEN FOR BGP MESSAGES
202
RECEIVE BGP UPDATE AND UPDATE TOPOLOGY INFORMATION BASE
203
AS_PATH LENGTH = 0 ?
NO
YES
204
COMM. ATTRIBUTE VALUE MAPPED TO A PID ?
YES
206
ADD PREFIX TO MAPPED PID
NO
208
COMM. ATTRIBUTE VALUES MAPPED TO PID ATTR. VALUES ?
YES
210
ADD PREFIX TO PID HAVING SAME MAPPED PID ATTRIBUTE VALUES AND SAME NEXT HOP ATTRIBUTE
NO
212
ADD PREFIX TO PID HAVING NO MAPPED PID ATTRIBUTE VALUES AND SAME NEXT HOP ATTRIBUTE
214
ADD PREFIX TO PID HAVING SAME AS_PATH ATTRIBUTE AND SAME NEXT HOP ATTRIBUTE
216
UPDATE CLIENT

FIG. 8

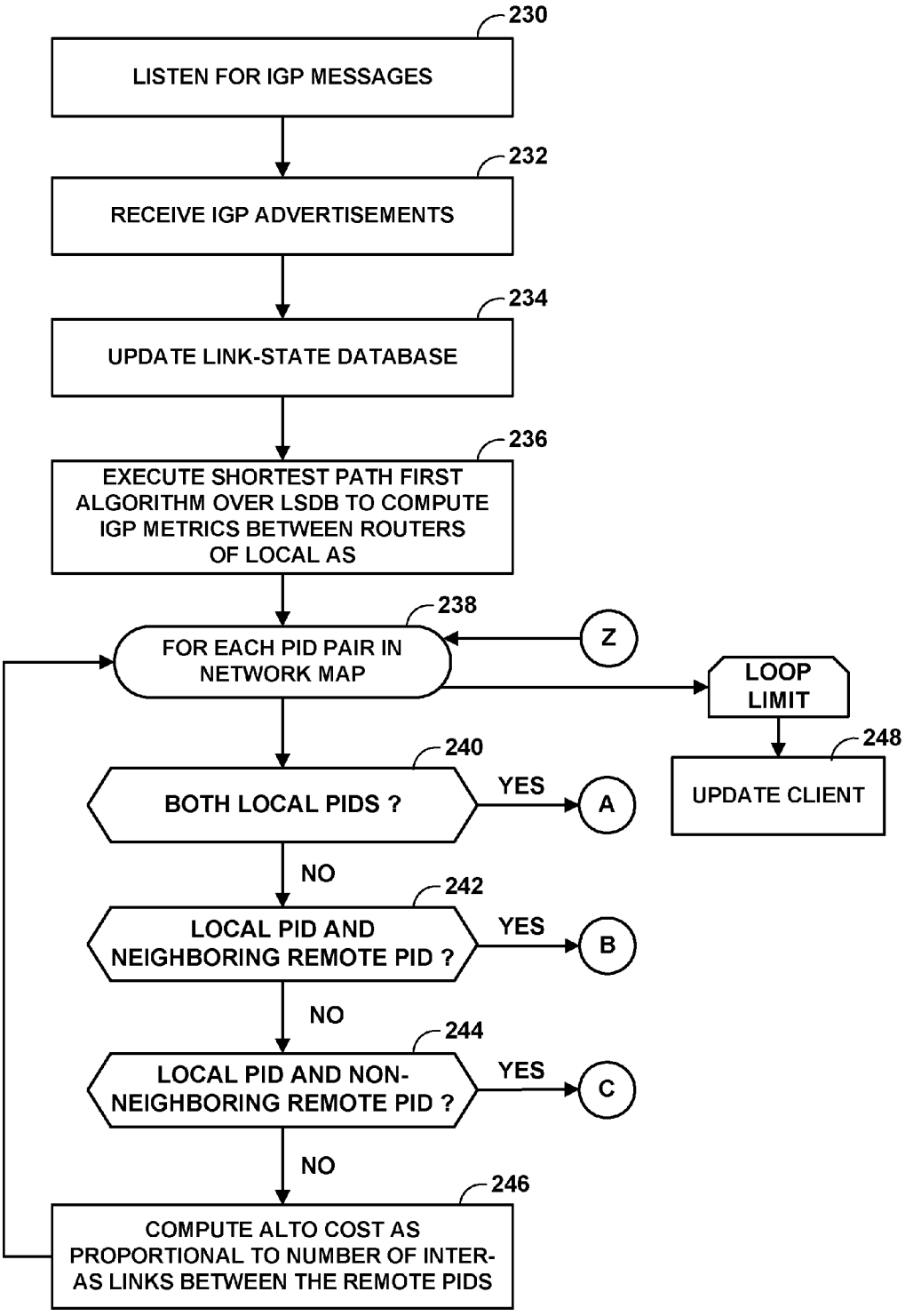

230
LISTEN FOR IGP MESSAGES
232
RECEIVE IGP ADVERTISEMENTS
234
UPDATE LINK-STATE DATABASE
236
EXECUTE SHORTEST PATH FIRST ALGORITHM OVER LSDB TO COMPUTE IGP METRICS BETWEEN ROUTERS OF LOCAL AS
238
FOR EACH PID PAIR IN NETWORK MAP
Z
LOOP LIMIT
248
UPDATE CLIENT
240
BOTH LOCAL PIDS ?
YES
A
NO
242
LOCAL PID AND NEIGHBORING REMOTE PID ?
YES
B
NO
244
LOCAL PID AND NON-NEIGHBORING REMOTE PID ?
YES
C
NO
246
COMPUTE ALTO COST AS PROPORTIONAL TO NUMBER OF INTER-AS LINKS BETWEEN THE REMOTE PIDS

FIG. 9A

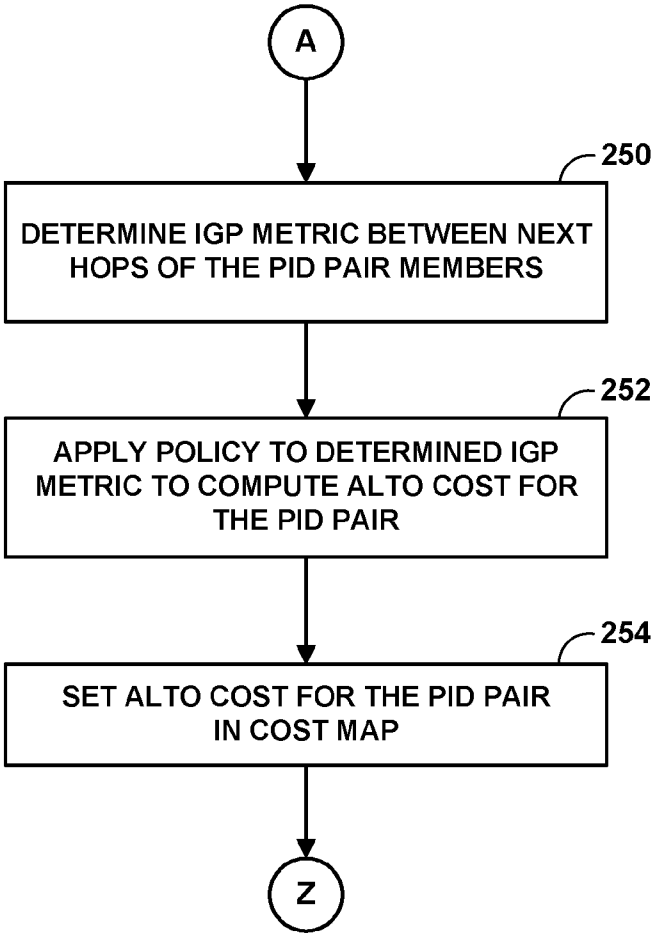

A
250
DETERMINE IGP METRIC BETWEEN NEXT HOPS OF THE PID PAIR MEMBERS
252
APPLY POLICY TO DETERMINED IGP METRIC TO COMPUTE ALTO COST FOR THE PID PAIR
254
SET ALTO COST FOR THE PID PAIR IN COST MAP
Z

FIG. 9B

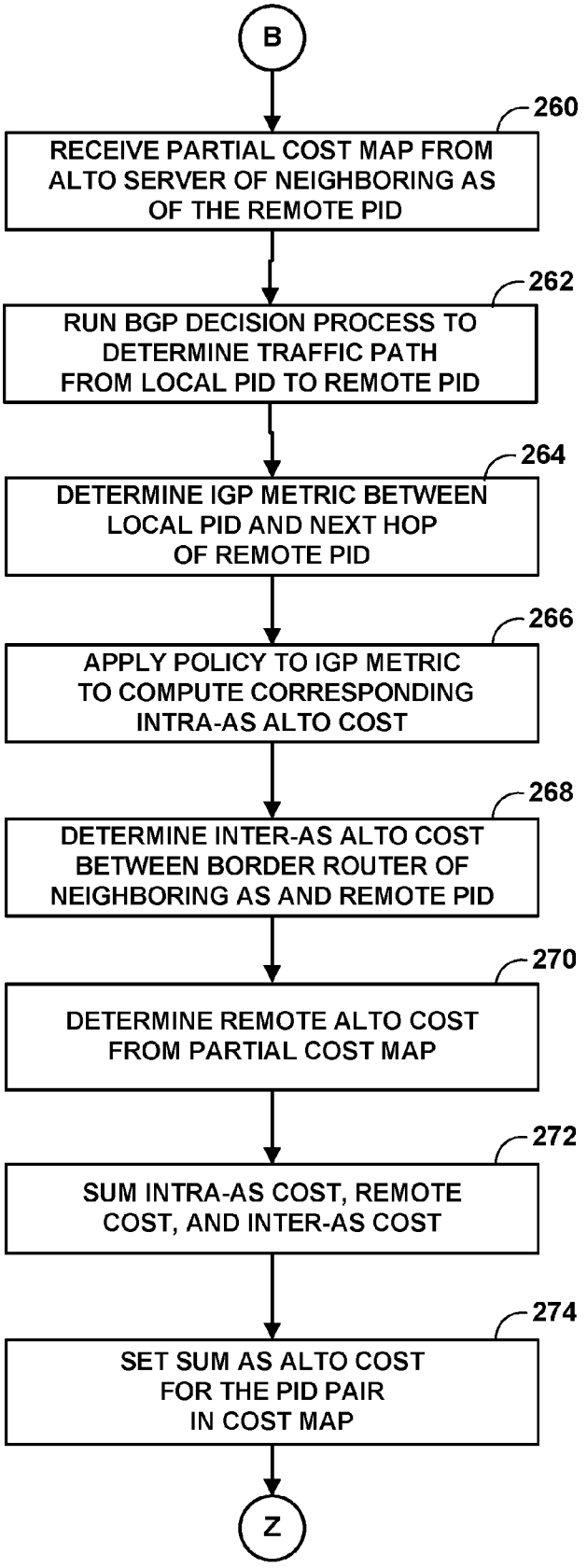

B
260
RECEIVE PARTIAL COST MAP FROM ALTO SERVER OF NEIGHBORING AS OF THE REMOTE PID
262
RUN BGP DECISION PROCESS TO DETERMINE TRAFFIC PATH FROM LOCAL PID TO REMOTE PID
264
DETERMINE IGP METRIC BETWEEN LOCAL PID AND NEXT HOP OF REMOTE PID
266
APPLY POLICY TO IGP METRIC TO COMPUTE CORRESPONDING INTRA-AS ALTO COST
268
DETERMINE INTER-AS ALTO COST BETWEEN BORDER ROUTER OF NEIGHBORING AS AND REMOTE PID
270
DETERMINE REMOTE ALTO COST FROM PARTIAL COST MAP
272
SUM INTRA-AS COST, REMOTE COST, AND INTER-AS COST
274
SET SUM AS ALTO COST FOR THE PID PAIR IN COST MAP
Z

FIG. 9C

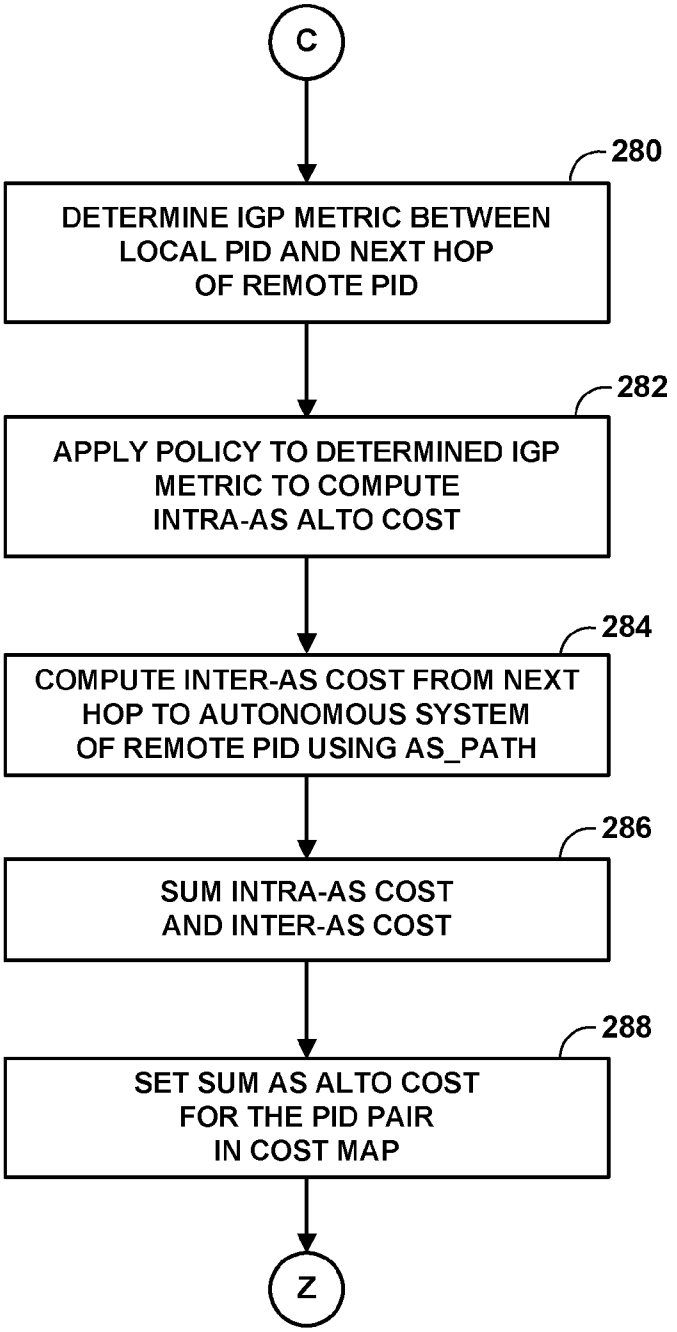

C
280
DETERMINE IGP METRIC BETWEEN LOCAL PID AND NEXT HOP OF REMOTE PID
282
APPLY POLICY TO DETERMINED IGP METRIC TO COMPUTE INTRA-AS ALTO COST
284
COMPUTE INTER-AS COST FROM NEXT HOP TO AUTONOMOUS SYSTEM OF REMOTE PID USING AS_PATH
286
SUM INTRA-AS COST AND INTER-AS COST
288
SET SUM AS ALTO COST FOR THE PID PAIR IN COST MAP
Z

FIG. 9D

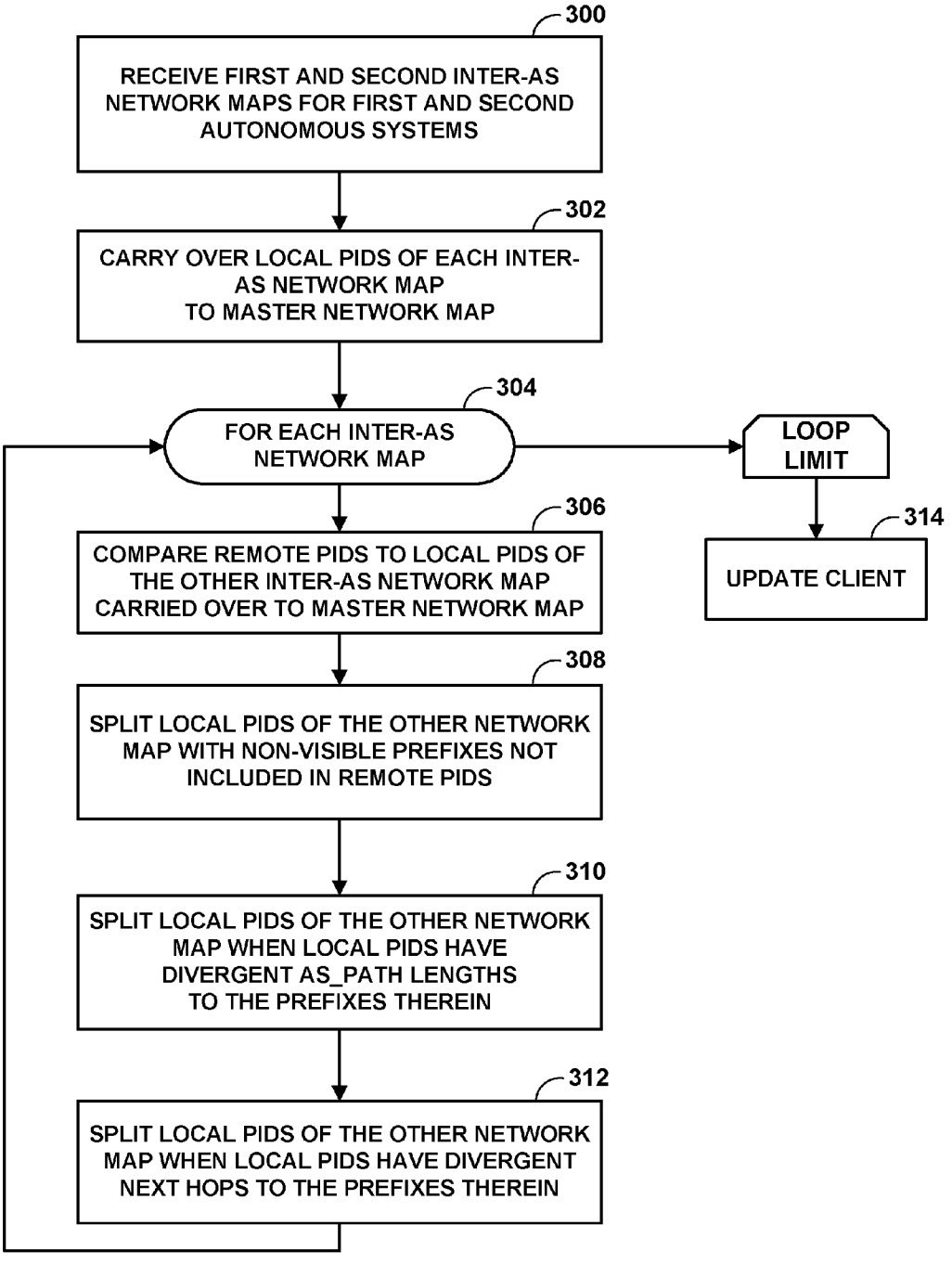

300
RECEIVE FIRST AND SECOND INTER-AS NETWORK MAPS FOR FIRST AND SECOND AUTONOMOUS SYSTEMS
302
CARRY OVER LOCAL PIDS OF EACH INTER-AS NETWORK MAP TO MASTER NETWORK MAP
304
FOR EACH INTER-AS NETWORK MAP
LOOP LIMIT
314
UPDATE CLIENT
306
COMPARE REMOTE PIDS TO LOCAL PIDS OF THE OTHER INTER-AS NETWORK MAP CARRIED OVER TO MASTER NETWORK MAP
308
SPLIT LOCAL PIDS OF THE OTHER NETWORK MAP WITH NON-VISIBLE PREFIXES NOT INCLUDED IN REMOTE PIDS
310
SPLIT LOCAL PIDS OF THE OTHER NETWORK MAP WHEN LOCAL PIDS HAVE DIVERGENT AS_PATH LENGTHS TO THE PREFIXES THEREIN
312
SPLIT LOCAL PIDS OF THE OTHER NETWORK MAP WHEN LOCAL PIDS HAVE DIVERGENT NEXT HOPS TO THE PREFIXES THEREIN

FIG. 10

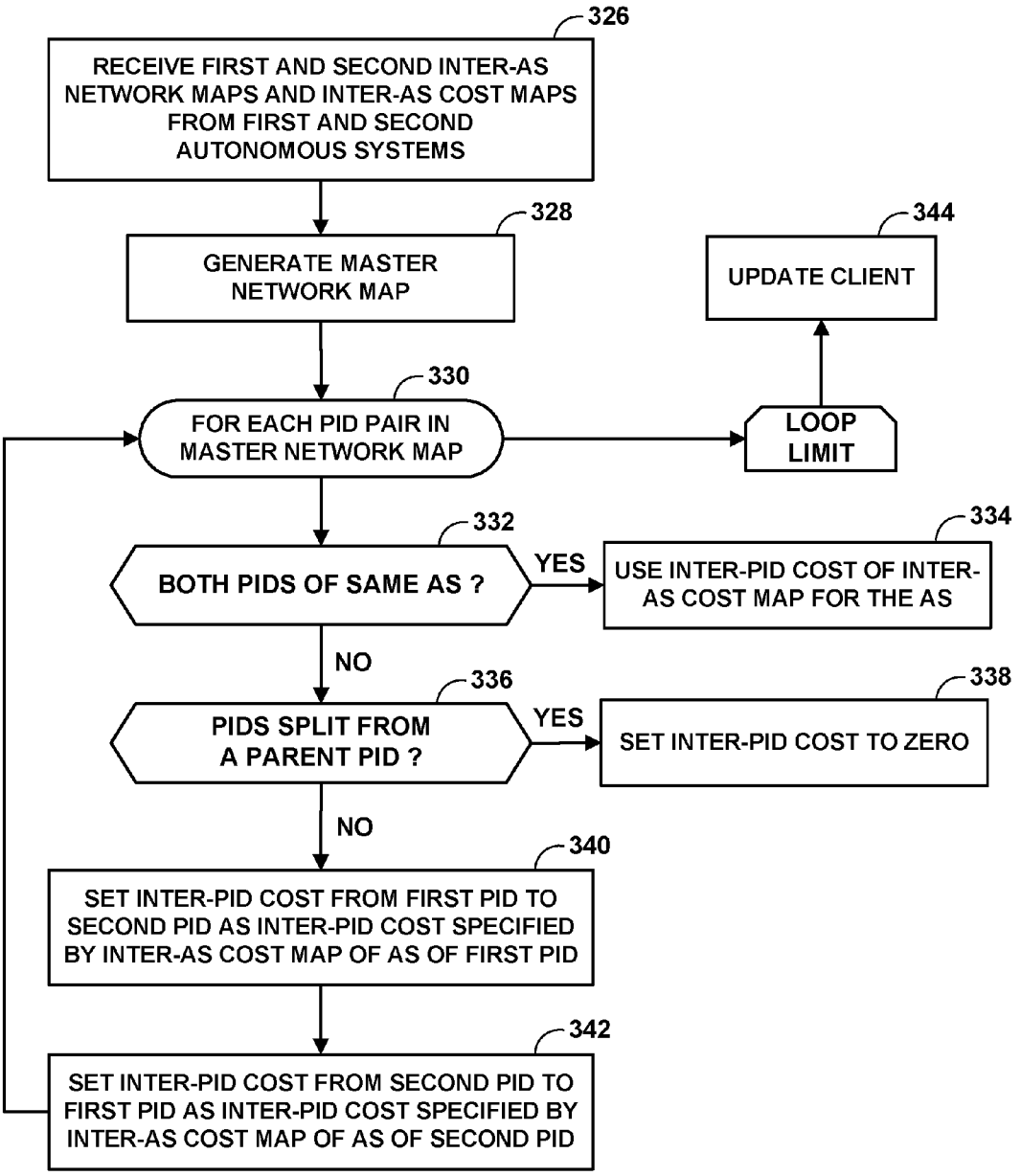

326
RECEIVE FIRST AND SECOND INTER-AS NETWORK MAPS AND INTER-AS COST MAPS FROM FIRST AND SECOND AUTONOMOUS SYSTEMS
328
GENERATE MASTER NETWORK MAP
330
FOR EACH PID PAIR IN MASTER NETWORK MAP
LOOP LIMIT
344
UPDATE CLIENT
332
BOTH PIDS OF SAME AS ?
YES
334
USE INTER-PID COST OF INTER-AS COST MAP FOR THE AS
NO
336
PIDS SPLIT FROM A PARENT PID ?
YES
338
SET INTER-PID COST TO ZERO
NO
340
SET INTER-PID COST FROM FIRST PID TO SECOND PID AS INTER-PID COST SPECIFIED BY INTER-AS COST MAP OF AS OF FIRST PID
342
SET INTER-PID COST FROM SECOND PID TO FIRST PID AS INTER-PID COST SPECIFIED BY INTER-AS COST MAP OF AS OF SECOND PID

FIG. 11

DYNAMICALLY GENERATING APPLICATION-LAYER TRAFFIC OPTIMIZATION PROTOCOL MAPS

[0001] This application is a continuation of U.S. patent application Ser. No. 14/252,526, filed Apr. 14, 2014, which is a continuation of U.S. patent application Ser. No. 13/110,987, filed May 19, 2011, now U.S. Pat. No. 8,700,801, issued Apr. 15, 2014, which claims the benefit of U.S. Provisional Application No. 61/418,793, filed Dec. 1, 2010 and of U.S. Provisional Application No. 61/449,499, filed Mar. 4, 2011, the entire content of each of which is incorporated by reference herein.

TECHNICAL FIELD

[0002] The invention relates to computer networks and, more specifically, to enhancing content delivery.

BACKGROUND

[0003] Peer-to-peer (P2P) and content delivery networks (CDNs) applications serve large amounts of data and generate significant amounts of network traffic. These applications leverage multiple copies of data content populated to multiple different network nodes to allow a requesting agent to obtain portions of the data content from one or more of many possible data sources. Distributing data content to multiple nodes for delivery on behalf of applications such as file sharing, real-time communication, and on-demand media streaming improves application performance and scalability.

[0004] P2P and CDN application clients often select resources naively, that is, without incorporating network topology information or related details. Rather, clients rely on heuristics to approximate such information. As a result, network data traffic exchanged using these applications may congest network links, cross service provider network boundaries multiple times, and generally transit the communication network in a manner that is suboptimal from a user-standpoint and undesirable from the point of view of the service provider. For instance, while two peers may be members of the same service provider network, an overlay link connecting the peers may nevertheless traverse multiple network boundaries, which unnecessarily increases the inter-peer transit costs to the service provider. Furthermore, although distributed applications capitalize on excess bandwidth at the data sources to improve throughput and reduce latencies for end-users while also reducing the burden of content providers to provision application servers, the ability to cheaply distribute data content comes at the expense of service providers, which bear the cost of inefficiently transporting network data.

[0005] A service provider, content provider, or a third party may provide an Application-Layer Traffic Optimization (ALTO) protocol service to provide guidance to application clients and content request routers regarding selection of a particular resource from which to obtain data content. The ALTO service provides information that incorporates provider preferences with regard to network resources to influence network resource consumption patterns while maintaining or improving application performance. In one example, a service provider provisions an ALTO server for a service provider network with network topology and topology link cost information. Application clients and content request routers send ALTO requests to the ALTO server to obtain a network map and a corresponding cost map. The network map specifies a set of topological groupings, or "PIDs," defined by the ALTO server for the network. A particular PID within a network map may represent a single device or device component, a collection of devices such as a network subnet identified by a network address prefix, a service provider network, or some other grouping. A cost map for a corresponding network map defines provider preferences respecting inter-PID routing costs for connections among the various PIDs of the network map. Using the network map and cost map provided by the ALTO server, application clients and content request routers select serving resources to minimize costs, as specified by the ALTO maps, between content requesting clients and available resources.

[0006] As a result, service providers provisioning the ALTO server may direct application clients and request routers to select resources according to service provider preferences, which may include optimizing throughput and/or user experience, for instance, reducing costs to the service provider, or promoting other provider objectives. The ALTO service and ALTO protocol is described in further detail in J. Seedorf et al., RFC 5693, "Application-Layer Traffic Optimization (ALTO) Problem Statement," Network Working Group, the Internet Engineering Task Force draft, October 2009; and R. Alimi et al., "ALTO Protocol: draft-ietf-alto-protocol-06.txt," ALTO Working Group, the Internet Engineering Task Force draft, October 2010, each of which is incorporated herein by reference in its entirety.

SUMMARY

[0007] In general, techniques are described for using routing information obtained by operation of network routing protocols to dynamically generate network and cost maps for an application-layer traffic optimization service, such as that provided by the Application-Layer Traffic Optimization (ALTO) protocol. In one example, an ALTO server of an autonomous system (AS) receives routing information, such as topology, link state, and link metrics information, from routers of the AS by listening for routing protocol updates output by the routers. In other words, the ALTO server may execute layer three (L3) routing protocols so as to snoop or otherwise receive routing protocol update messages exchanged between L3 routing devices within the network. Based on the routing protocol update messages, the ALTO server assembles topology information representative of the network. Topology information snooped by the ALTO server may include information that describes the intra-AS topology of the AS that includes the receiving ALTO server, as well as information that describes intra-AS topologies of neighboring autonomous systems and of an inter-AS topology of multiple, interconnected autonomous systems. The ALTO server uses the assembled topology information to dynamically generate an ALTO network map of PIDs that reflects a current topology of the AS that includes the ALTO server and/or of the broader network that includes additional ASes. In some cases, the ALTO server functionality may be incorporated within an L3 routing device of the network that operates as a peer to other routers within the network.

[0008] Link metrics (e.g., distance or throughput) received by the ALTO server in routing protocol updates, autonomous system path lengths, and administratively configured data determine current inter-PID costs among the PIDs of the dynamically generated network map. The ALTO server

dynamically calculates inter-PID costs using received routing information that reflects current link metrics. The ALTO server then assembles the inter-PID costs into an ALTO cost map that the ALTO server may provide, along with the ALTO network map, to ALTO clients.

[0009] Additionally, a master ALTO server may interact with ALTO servers of other federated autonomous systems to receive network and cost maps generated in the AS-internal ALTO Servers according to their respective network topology perspectives. Using the detailed topology and cost information included within such maps for remote areas of the broader network, the master ALTO server generates master network ALTO and cost maps that detail a more complete perspective of available PIDs and inter-PID costs of the broader multi-AS network. The ALTO server may then pass the master ALTO network and cost maps to ALTO clients and/or to federated ALTO servers to improve inter-AS node selection by applications.

[0010] The described techniques may present one or more advantages. For example, dynamically generating and updating ALTO network and costs maps with an ALTO server using routing information received from network elements synchronizes the ALTO maps to an ever-changing network environment. As a result, the ALTO network and cost maps provided by the ALTO server to ALTO clients may reflect recent updates to the network topology and/or utilization and may thus improve node selection and increase application performance. Moreover, automatically creating network and cost maps may reduce an ALTO server configuration burden on an operator, making an ALTO server implementation viable in a large-scale service provider environment. Additionally, the techniques may use both intra-AS (intra-domain) and inter-AS (inter-domain) routing information received from network elements, such as Border Gateway Protocol and Interior Gateway Protocol speakers. An ALTO server that implements the techniques may therefore receive and incorporate in ALTO network and costs maps routing information from remote ASes that may not be administratively configurable within the ALTO server.

[0011] In one embodiment, the invention is directed to a method comprising executing a routing protocol on an application-layer traffic optimization (ALTO) server to receive layer three (L3) network topology information defining routes to endpoints of a network, and aggregating, with the ALTO server, the endpoints into a set of one or more subsets of topological groupings (PIDs), wherein each PID is associated with a different subset of the endpoints. The method further comprises receiving, with the routing protocol, a topology information advertisement that specifies one or more routes and includes network address information identifying one or more of the endpoints. The method further comprises aggregating, with the ALTO server, the identified endpoints into a first one of the set of PIDs. The method also comprises generating, with the ALTO server, an ALTO network map that includes a PID entry to describe each of the PIDs, and sending the ALTO network map from the ALTO server to an ALTO client.

[0012] In another embodiment, the invention is directed to a method comprising executing an interior gateway protocol on an application-layer traffic optimization (ALTO) server, and receiving, with the interior gateway protocol, routing information for an autonomous system that includes the ALTO server. The method also comprises computing, with the ALTO server, an ALTO cost for a pair of a set of one or more subsets of topological groupings (PIDs) of an ALTO network map based at least on the routing information, wherein each one of the set of PIDs is associated with a different subset of endpoints of a network that comprises the autonomous system. The method further comprises generating an ALTO cost map that includes an entry that specifies the ALTO cost between the first member and second member of the pair of PIDs, and sending the ALTO cost map from the ALTO server to an ALTO client.

[0013] In another embodiment, the invention is directed to a method comprising receiving a first inter-AS network map and a first inter-AS cost map for a first autonomous system with a master application-layer traffic optimization (ALTO) server, wherein the first inter-AS network map comprises a first set of one or more local and remote subsets of topological groupings (PIDs), wherein each local and remote PID of the first inter-AS network map is associated with a different subset of endpoints of a network, wherein the local PIDs of the first inter-AS network map specify network address prefixes of the first autonomous system and remote PIDs of the first inter-AS network map specify network address prefixes of a second autonomous system, wherein the first inter-AS cost map specifies ALTO costs for pairs of PIDs of the first inter-AS network map. The method also comprises receiving a second inter-AS network map for the second autonomous system with the master ALTO server, wherein the second inter-AS network map comprises a second set of one or more local and remote PIDs, wherein each local and remote PID of the second inter-AS network map is associated with a different subset of the endpoints of the network, wherein the local PIDs of the second inter-AS network map specify network address prefixes of the second autonomous system and remote PIDs of the second inter-AS network map specify network address prefixes of the first autonomous system, wherein the second inter-AS cost map specifies ALTO costs for pairs of PIDs of the second inter-AS network map. The method further comprises generating, with the master ALTO server, a master ALTO network map for the network based at least on the first inter-AS network map and the second inter-AS network map, and outputting the master ALTO network map from the master ALTO server.

[0014] In another embodiment, the invention is directed to an application-layer traffic optimization (ALTO) server comprising a control unit having one or more processors and a topology information base. A Border Gateway Protocol (BGP) listener of the control unit that executes a routing protocol to receive layer three (L3) network topology information defining routes to endpoints of a network that includes an autonomous system that includes the ALTO server. A PID generator of the control unit that aggregates the endpoints into a set of one or more subsets of topological groupings (PIDs), wherein each PID is associated with a different subset of the endpoints, wherein the BGP listener receives a topology information advertisement that specifies one or more routes and includes network address information identifying one or more of the endpoints, wherein the BGP listener stores the one or more routes to the topology information base, and wherein the PID generator aggregates the identified endpoints into a first one of the set of PIDs. The ALTO server also includes a network map module of the control unit that generates an ALTO network map that

includes a PID entry to describe each of the PIDs, and a client interface that sends the ALTO network map to an ALTO client.

[0015] In another embodiment, the invention is directed to an application-layer traffic optimization (ALTO) server comprising a control unit having one or more processors. An interface of the control unit that receives first inter-AS network map and a first inter-AS cost map for a first autonomous system, wherein the first inter-AS network map comprises a first set of one or more local and remote subsets of topological groupings (PIDs), wherein each local and remote PID of the first inter-AS network map is associated with a different subset of endpoints of a network, wherein the local PIDs of the first inter-AS network map specify network address prefixes of the first autonomous system and remote PIDs of the first inter-AS network map specify network address prefixes of a second autonomous system, wherein the first inter-AS cost map specifies ALTO costs for pairs of PIDs of the first inter-AS network map, wherein the interface receives a second inter-AS network map for the second autonomous system, wherein the second inter-AS network map comprises a second set of one or more local and remote PIDs, wherein each local and remote PID of the second inter-AS network map is associated with a different subset of the endpoints of the network, wherein the local PIDs of the second inter-AS network map specify network address prefixes of the second autonomous system and remote PIDs of the second inter-AS network map specify network address prefixes of the first autonomous system, wherein the second inter-AS cost map specifies ALTO costs for pairs of PIDs of the second inter-AS network map. The ALTO server also comprises a network map module of the control unit that generates a master ALTO network map for the network based at least on the first inter-AS network map and the second inter-AS network map, and a client interface of the control unit that sends the master ALTO network map to an ALTO client.

[0016] The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

[0017] FIG. **1** is a block diagram illustrating an example network that dynamically generates and updates, using advertised routing information, network and cost maps in the manner described herein for use in an Application-Layer Traffic Optimization (ALTO) service.

[0018] FIG. **2** is a block diagram illustrating an example graph that represents a combined ALTO inter-AS network and cost map that are dynamically generated by an ALTO server for a network according to the described techniques.

[0019] FIG. **3** is a block diagram illustrating a network comprising ALTO servers that perform federated ALTO techniques described herein.

[0020] FIGS. **4**A-**4**B are block diagrams illustrating an example graphs that represents a partial combined master network and cost map for a network that is generated by a master ALTO server, in accordance with the described techniques, from multiple partial combined inter-AS network and cost maps for the network as generated by and from the perspective of respective ALTO servers.

[0021] FIG. **5** is a block diagram illustrating an example network comprising ALTO servers that perform federated ALTO techniques for neighboring autonomous systems in the manner described herein.

[0022] FIG. **6**A is a block diagram illustrating an example graph that represents a partial combined intra-AS network and cost map for that is generated by an ALTO server for an autonomous system according to the described techniques.

[0023] FIG. **6**B is a block diagram illustrating an example graph that represents a partial combined master network and cost map generated for a network in accordance with the techniques described herein.

[0024] FIG. **7** is a block diagram illustrating, in detail, an example ALTO server that receives routing information and dynamically generates network and cost maps in accordance with the techniques described herein.

[0025] FIG. **8** is a flowchart illustrating an example mode of operation of an ALTO server for dynamically generating or modifying an inter-AS network map, according to the described techniques, for an autonomous system served by the ALTO server.

[0026] FIGS. **9**A-**9**D present a flowchart that illustrates an example operation of an ALTO server to dynamically generate or modify an inter-AS cost map for an autonomous system served by the ALTO server according to the techniques described herein.

[0027] FIG. **10** is a flowchart illustrating an example operation of an ALTO server to generate a master ALTO network map according to the techniques set forth herein.

[0028] FIG. **11** is a flowchart illustrating an example operation of an ALTO server to generate master ALTO network and cost maps according to the techniques set forth herein.

DETAILED DESCRIPTION

[0029] FIG. **1** is a block diagram illustrating an example network **2** that dynamically generates and updates, using advertised routing information, network and cost maps for use in an Application-Layer Traffic Optimization (ALTO) service. Network **2** comprises autonomous systems **4**A-**4**D (illustrated as "ASes **4**A-**4**D" and collectively referred to herein as "autonomous systems **4**") interconnected by external communication links. The term "communication link," as used herein, comprises any form of transport medium, wired or wireless, and can include intermediate nodes such as network devices. Network **2** may in some embodiments represent the Internet or any other publicly accessible computer network, a private network, or a virtual private network (VPN), that transports content for delivery to requesting devices. While described with respect to Internet Protocol networks, the techniques are also applicable to other types of delivery networks, such as Asynchronous Transfer Mode (ATM) networks.

[0030] Each of autonomous systems **4** run one or more interior gateway protocols (IGPs), such as Open Shortest Path First (OSPF), Routing Information Protocol (RIP), Intermediate System-to-Intermediate System (IS-IS), Interior Gateway Routing Protocol (IGRP), Enhanced IGRP (EIGRP), and Interior Border Gateway Protocol (iBGP), and each of autonomous systems **4** includes a set of one or more routers operating within a single administrative domain according to a routing policy. Autonomous systems **4** each have an identification number provided by an Internet registry or by an Internet service provider (ISP) that uniquely

identifies the autonomous system to other autonomous systems. In some instances, the identification number may be drawn from a private identifier number space and therefore unique only within a private network comprising ASes **4**. In various embodiments, each of autonomous systems **4** may represent a service provider network, an enterprise or campus network, a content access network (CAN), or a content delivery network (CDN), for example. In addition, one or more service providers, content provider, or enterprise/campus network administrators may administer any one or more of autonomous systems **4**.

[0031] Routers of autonomous systems **4** execute an Internet Protocol (e.g., IPv4 or IPv6) to route packets from a source network addresses to destination network addresses, and each of autonomous system **4** offers network packet delivery to a network (or subnet) of one or more endpoints identified by a network address prefix that encompasses the network address range defined by the network addresses of endpoints. For example, AS **4**B-**4**D offer packet delivery to/from respective remote network address prefixes **21**A-**21**C ("remote prefixes **21**"), while AS **4**A offers packet delivery to/from local network address prefixes **20**A-**20**D ("local prefixes **20**"). The terms "remote" and "local" with respect to the prefixes refer to the network perspective of AS **4**A, illustrated in greater detail that ASes **4**B-**4**D. The various routers illustrated in autonomous system **4**A are interconnected by internal communication links.

[0032] An AS that offers packet delivery to a prefix is the "originating domain" for the prefix, also known as the origin AS, and the BGP routing protocol running in all autonomous systems **4** propagates the identification number of the origin AS for the prefix as reachability information for the prefix. As a result, autonomous systems **4** route packets destined for an endpoint having a network address that is within a particular prefix to the origin AS for the prefix. For instance, autonomous systems **4**A-**4**C route packets destined for an endpoint of prefix **21**C to AS **4**D.

[0033] Host **10** accesses AS **4**A to issue content requests to, e.g., other hosts or CDN nodes located in any of autonomous systems **4**, and to receive application-related content for hosted applications. Host **10** may represent, for example, a workstation, desktop computer, laptop computer, cellular or other mobile device, Personal Digital Assistant (PDA), gaming console, television set-top box, or any other device capable of accessing a computer network via a wireless and/or wired connection. Host **10** has a network address within local prefix **20**C and is thus directly served by internal router **8**B of AS **4**A. In some embodiments, host **10** is a member of a customer network served by AS **4**A.

[0034] The other devices (not shown in FIG. **1**) to which host **10** issues content requests may be served by any of ASs **4**A-**4**D. For example, host **10** may issue a content request to a CDN node having a network address within remote prefix **21**B served by AS **4**C. As another example, host **10** may issue a content request to an additional host having a network address within local prefix **20**A and executing a peer-to-peer (P2P) application for exchanging content among the various hosts, including host **10**, which also execute the P2P application.

[0035] In the illustrated embodiment, AS **4**A includes autonomous system boundary routers **6**A-**6**B ("ASBRs **6**") that connect AS **4**A to ASes **4**B, **4**D, respectively, over one of communication links (ASBRs of ASes **4**B-**4**D not shown for ease of illustration). ASBRs **6** execute an exterior gateway protocol (EGP) to exchange, in one of external peering sessions **11** with ASBRs of ASes **4**A and **4**D, routing information for respective autonomous systems **4**. For example, ASBRs **6** provides routing information that describes an internal topology of autonomous system **4**A and/or reachability of local prefixes **20**. Additionally, ASBRs **6** receive routing information that describes reachability of remote prefixes **21** via ASes **4**B-**4**D. In some instances, ASBRs **6** may receive routing information originated by one of ASes **4**B-**4**D that describes an internal topology of the originating autonomous system.

[0036] Routing information for AS **4**A outputted by one of ASBRs **6** in a peering session typically includes topology information received from one or more interior routing protocol speakers of autonomous system **4**A executing an IGP, such as Internal BGP (iBGP), or received in one of external peering sessions **11** from another one of ASes **4**. Topology information may also include administratively configured routes or other information on ASBRs **6**. The EGP used for external peering sessions **11** may comprise, for instance, Exterior Border Gateway Protocol (BGP). Each external peering session **11** may comprise a Transmission Control Protocol (TCP) session.

[0037] Autonomous system **4**A includes reachability protocol speakers that peer with one another to internally advertise topology information, e.g., routes, for local prefixes **20** and remote prefixes **21** to other routers within AS **4**A. Reachability protocol speakers of AS **4**A include ASBRs **6**, route reflector **17**, and internal router **8**B. ASBRs **6** and internal router **8**B each establish a peering session **9** with which to exchange routes with route reflector **17**. Route reflector **17** is a reachability protocol speaker that re-advertises (or "reflects") routes received from other gateway protocol speakers to enable the reachability protocol speakers to avoid forming a full mesh of peering sessions **9**. However, in embodiments of network **2** that do not include route reflector **17**, the reachability protocol speakers may form a full mesh of peering sessions. The reachability protocol with which reachability protocol speakers advertise routes may comprise, for example, the Interior Border Gateway Protocol (IBGP). Topology information advertisements may comprise, for example, route advertisements such as IBGP UPDATE messages. In general, a topology information advertisement associates a prefix with a NEXT_HOP for the prefix and a list of autonomous systems that must be traversed to reach the prefix ("AS_PATH" in IBGP UPDATE messages).

[0038] The service provider or other administrator for AS **4**A deploys Application-Layer Traffic Optimization (ALTO) server **12** to AS **4**A to provide an application-layer traffic optimization service over autonomous systems **4**. The application-layer traffic optimization may in some instances conform to the ALTO protocol. In general, the ALTO service enables service and/or content providers to influence the node selection process by applications to further service/content provider objectives, which may include improving a user experience by selecting the most geographically proximate serving node to requesting host **10**, reducing transmission costs to the provider, load balancing, service-level discrimination, accounting for bandwidth constraints, decreasing round-trip delay between host **10** and the serving node, and other objectives. Further details regarding the use of an ALTO service in CDNs may be found in R. Penno et al., "ALTO and Content Delivery Networks: draft-penno-

alto-cdn-03," Network Working Group, the Internet Engineering Task Force draft, March 2011, which is incorporated herein by reference in its entirety. Furthermore, while generally described with respect to the ALTO service and ALTO servers as described in Seedorf et al., incorporated above, the techniques are applicable to any form of application-layer traffic optimization.

[0039] ALTO server **12** generates a network map and cost map for network **2** from the perspective of the ALTO server and provides these maps to ALTO clients in ALTO maps update message **13**, such as ALTO client **18** of host **10**. A network map contains network location identifiers, or PIDs, that each represents one or more network devices in a network. In general, a PID may represent a single device or device component, a collection of devices such as a network subnet, one of ASes **4**, or some other grouping. A cost map contains cost entries for pairs of PIDs represented in the network map and an associated value that represents a cost to traverse a network path between the members of the PID pair. The value can be ordinal (i.e., ranked) or numerical (e.g., actual). ALTO client **18** of host **10** uses the network map and cost map to determine an optimal endpoint for use by an application executing on host **10**. For example, host **10** may execute a P2P application that requires particular content from a peer of the P2P network. ALTO client **18** of host **10** determines, from the network map and cost map, to determine an optimal peer from which the P2P application may request and download the content. In some embodiments, host **10** represents a request router that receives content requests from client applications executing on nodes of network **2**, determines an optimal server for the requesting clients using the network map and cost map provided by ALTO server **12**, and returns a network address or other identifier for the optimal servers to the requesting nodes. In still further embodiments, host **10** executes a client of a client-server application that performs one of the techniques described above. Further details regarding generating network and cost maps for a multi-domain network are found in Penno et al., U.S. patent application Ser. No. 12/861,645, entitled "APPLICATION-LAYER TRAFFIC OPTIMIZATION SERVICE SPANNING MULTIPLE NETWORKS," filed Aug. 23, 2010, the entire contents of which are incorporated herein by reference. Additional details regarding ALTO map updates are found in Raghunath et al., U.S. patent application Ser. No. 12/861,681, entitled "APPLICATION-LAYER TRAFFIC OPTIMIZATION SERVICE MAP UPDATES," filed Aug. 23, 2010, the entire contents of which are incorporated herein by reference.

[0040] In some embodiments, ALTO server **12** provides an endpoint cost service. In such embodiments, ALTO client **18** of host **10** (operating as a peer for a P2P application) provides a list of one or more endpoints and an identifier for hosts **10** to ALTO server **12**. In response, ALTO server **12** uses network and cost maps generated in accordance with the techniques described herein to determine an optimal endpoint in the list of endpoints received for the receiving host. ALTO server **12** returns the optimal endpoint to ALTO client **18** for use by an application executing on host **10**. In some embodiments of ALTO server **12** that implement an endpoint cost service, ALTO server **12** returns a list of the endpoints in a rank ordering according to cost or returns a list of the endpoints with associated costs for each endpoint.

[0041] ALTO server **12** may comprise, for example, a high-end server or other service device, a content indexer for a P2P application, or a service card or programmable interface card (PIC) insertable into a network device, such as a router or switch. ALTO server **12** may operate as an element of a service plane of a router to provide ALTO services in accordance with the techniques of this disclosure. Additional details regarding providing ALTO services as an element of a service plane of a router are found in Raghunath et al., incorporated above.

[0042] In accordance with the techniques described herein, ALTO server **12** establishes peering session **9**, which may comprise an IBGP session, with route reflector **17** of AS **4**A. In this way ALTO Server **12** receives, in peering session **9**, routing information for AS **4**A originated or forwarded by the reachability protocol speakers of AS **4**A. In this instance, ALTO server **12** is a passive reachability protocol listener that receives routing information reflected by route reflector **17**. That is, in this example, because ALTO server **12** does not (in its capacity as an ALTO server) originate or forward routes, route packets, or perform other such routing functions, ALTO server **12** merely receives routing information. Peering session **9** may comprise a Transmission Control Protocol (TCP) session between route reflector **17** and ALTO server **12**. In instances that do not include a route reflector, ALTO server **12** may establish peering session **9** to form a full protocol mesh with each of the reachability protocol speakers, e.g., ASBRs **6** and internal router **8**B.

[0043] In accordance with the techniques described herein, ALTO server **12** generates an intra-AS ALTO network map for AS **4**A using routing information received in peering session **9**. An intra-AS network map includes prefixes that constitute or are served by AS **4**A (i.e., originated by AS **4**A) aggregated by ALTO server **12** into one or PIDs. The intra-AS network map does not include any prefixes originated external to AS **4**A.

[0044] In some instances, an administrator or other entity may configure reachability protocol speakers, including ASBRs **6** and internal router **8**B, to incorporate a prefix attribute into routing information outputted in peering sessions **9**. In instances where the reachability protocol is IBGP, the prefix attribute may comprise a BGP Community Attribute or BGP Extended Community Attribute. An AS **4**A administrator, or another entity, assigns within the corresponding one of ASBRs **6** and/or internal router **8**B, a prefix attribute to one or more prefixes **20** for which the reachability protocol speaker originates or forwards routing information. For each prefix **20** typed in this manner, the reachability protocol speaker adds the prefix attribute to a topology information advertisement that includes topology information that traverses or originates with the respective reachability protocol speaker and that identifies the prefix.

[0045] For example, an administrator may configure internal router **8**B to incorporate a particular prefix attribute (e.g., "PREFIX_1") into a first topology information advertisement that advertises prefix **20**C and into a second topology information advertisement that advertises prefix **20**D. As another example, an administrator may configure internal router **8**B to incorporate a first particular prefix attribute (e.g., "PREFIX_1") into a first topology information advertisement that advertises prefix **20**C and into a second topology information advertisement that advertises prefix **20**D. The administrator may additionally configure internal router **8**B to incorporate a second particular prefix attribute (e.g., "CDN NODE") into the second topology information advertisement to identify the range of network addresses defined

by prefix **20**D as including endpoints that are CDN nodes of a CDN. Internal router **8**B then generates/modifies and outputs topology information advertisements in accordance with the specified configuration to route reflector **17** in a peering session **9**. ALTO server **12** therefore receives topology information advertisements that may include one or more prefix attributes for advertised prefixes **20**.

[0046] A prefix attribute in a topology information advertisement may comprise a string, bitstring, or integer, for instance. Further details regarding using prefix attributes of topology information advertisements to specify endpoint types in an ALTO context may be found in Medved et al., U.S. patent application Ser. No. 12/982,153, entitled "DYNAMICALLY GENERATING APPLICATION-LAYER TRAFFIC OPTIMIZATION PROTOCOL ENDPOINT ATTRIBUTES," filed Dec. 30, 2010, the entire contents of which are incorporated by reference herein.

[0047] ALTO server **12** includes one or more network map policies that specify PID aggregation or PID attributes for topology information received in topology information advertisements in a peering session **9**. For example, the network map policies may define mappings of prefix attributes incorporated within topology information advertisements to PID attributes. A PID attribute value may be different than the prefix attribute value to which it is mapped. As another example, the network map policies may define mappings of prefix attributes incorporated within topology information advertisements to specified PID identifiers to direct ALTO server **12** to aggregate one or more prefixes tagged with a particular prefix attribute into the specified PID. In some instances, network map policies that specify PID aggregation may include additional information that defines a cost between PIDs. The specified cost may comprise a constant value or a formula for computing a cost based on one or more parameters.

[0048] When ALTO server **12** receives topology information advertisements, the ALTO server dynamically generates or modifies an intra-AS network map in accordance with the network map policies described above. In one example implementation, ALTO server **12** generates and/or updates PIDs of an intra-AS network map according to the following rules that reference the network map policies. First, if the advertised prefix originates from outside of AS **4**A, ALTO server **12** ignores the advertised prefix.

[0049] For an advertised prefix that originates with AS **4**A, if the topology information advertisement that carries the prefix includes a prefix attribute that maps to a specified PID within an ALTO map policy, then ALTO server **12** adds the prefix to the specified PID in the network map.

[0050] If a prefix attribute is not specified in any of the network map policies of ALTO server **12**, the ALTO server creates/modifies a PID having various prefixes associated with the same NEXT_HOP attribute in topology information advertisements for the prefixes. In some instances, ALTO server **12** may receive from multiple reachability protocol speakers different topology information advertisements for a particular prefix that each specifies a different NEXT_HOP attribute for the prefix. In such instances, ALTO server **12** first selects one of the advertisements (e.g., one of the routes) according to a decision process, such as the BGP decision process described in Rekhter et al., "A Border Gateway Protocol 4 (BGP-4)," Request for Comments 4271, January, 2006, section 9.1 of which is incorporated by reference herein. In some instances, route reflector **17** performs the decision process prior to providing topology information advertisements in peering session **9** to ALTO server **12**. The ALTO server **12** then uses the NEXT_HOP attribute for the selected topology information advertisement to aggregate the prefix therein with other prefixes also associated with the same NEXT_HOP attribute in respective topology information advertisements (provided the advertisements do not include a prefix attribute specified in a network map policy). In this way, ALTO server **12** groups the prefixes according to the next hop router for the prefixes when network map policies of the ALTO server do not associate the prefixes to a PID or PID attribute.

[0051] For received topology information advertisements that do not fit any of the above categories yet originate with AS **4**A and include one or more prefix attributes that map to one or more PID attributes in a network map policy, ALTO server **12** aggregates into PIDs advertised prefixes associated in the advertisements with the same set of prefix attributes and also with the same NEXT_HOP attribute. According to this rule, ALTO server **12** may group into separate PIDs nodes that are both topologically proximate and that also exhibit similar functionality (e.g., CDN nodes attached to AS **4**A at the same NEXT_HOP). In some instances, ALTO server **12** may perform the decision process described above to select one of multiple topology information advertisements for use in network map generation/modification. In some instances, however, route reflector **17** performs the decision process prior to providing topology information advertisements in peering session **9** to ALTO server **12**. In some instances, ALTO server **12** may receive overlapping routes, as described in Rekhter et al., incorporated above. In such instances, ALTO server **12** may append the advertised prefix to multiple PIDs, which is permitted according to Alimi et al., incorporated above.

[0052] In addition to the dynamic PID aggregation process, ALTO server **12** dynamically assigns PID attributes to PIDs of the intra-AS network map according to network map policies that define mappings of prefix attributes incorporated within topology information advertisements to PID attributes. ALTO server **12** also specifies for each PID in the intra-AS network map, as a PID attribute, the NEXT_HOP for the PID prefixes.

[0053] Application of the above rules by ALTO server **12** to topology information advertisements produces an intra-AS network map that reflects a current topology of AS **4**A. That is, the intra-AS network generated/updated by the ALTO server **12** includes PIDs dynamically aggregated and PID attributes dynamically assigned in accordance with the techniques described above. ALTO server **12** provides the intra-AS network map to ALTO client **18** in ALTO maps update message **13** for use by host **10** in node selection.

[0054] Routers of AS **4**A, including ASBRs **6** and internal routers **8**A-**8**B, execute an interior gateway protocol (IGP) to disseminate routing information that allows the routers to select routes between any two nodes on a computer network. One type of routing protocol, referred to as a link state protocol, allows routers to exchange and accumulate link state information, i.e., information describing the various communication links that interconnect routers within the AS **4**A. With a typical the link state routing protocol, the routers exchange information related to available interfaces, metrics and other variables associated with network links. This allows a router to construct its own topology or map of the network. Metrics may include, for example, latency, link

throughput, link availability and reliability, path length, load, and communication cost (i.e., price). These metrics are typically expressed as simple integers.

[0055] Link state protocols include OSPF and IS-IS. Through application of the link state protocol, the routers exchange link information with other adjacent routers via link state advertisements (LSAs). A router generating an LSA typically floods the LSA throughout the network such that every other router receives the LSA. In this way, the receiving routers may construct and maintain their own network topologies in a routing table (e.g., a link-state database (LSDB)) using the link information exchanged via the LSAs.

[0056] Another type of routing protocol, referred to as a distance vector routing protocol, allows routers to exchange "vectors" of routing information that include, in each vector, a distance and direction. The distance refers to a metric, while the direction specifies a next hop router for the route advertised by the vector. In general, each router learns routes from neighboring routers and advertises routes from its own perspective. Distance vector protocols include, for example, Routing Information Protocol (RIP), Interior Gateway Routing Protocol (IGRP), and Enhanced IGRP (EIGRP).

[0057] ALTO server **12** operates as a passive IGP listener by peering with internal router **8**B in IGP peering session **7**. That is, ALTO server **12** receives routing information from internal router **8**B in IGP peering session **7** but does not originate or forward routing information, because ALTO server **12** does not route packets (in its capacity as an ALTO server). In some instances, internal router **8**B may set an overload (OL) bit in link-state advertisements (LSAs) to ALTO server **12** to prevent the ALTO server from returning routing information to internal router **8**B.

[0058] IGP peering session **7** may represent, for example, an OSPF neighbor relationship (or "adjacency") or may simply represent movement of current routing information from internal router **8**B to ALTO server **12**. In various configurations, ALTO server **12** may peer with other routers of AS **4**A in addition, or alternatively, to internal router **8**B.

[0059] If routers of AS **4**A execute a single-area OSPF (that is, if AS **4**A is a single-area OSPF network), ALTO server **12** may peer with any router of AS **4**A that executes OSPF to obtain the required routing information. Similarly, if routes of AS **4**A execute Level 2 IS-IS, ALTO server **12** may peer with any router of AS **4**A that executes IS-IS to obtain the required routing information.

[0060] In some instances, AS **4**A may comprise a multi-area OSPF network. In such instances, ALTO server **12** establishes IGP peering session **7** with at least one backbone (i.e., area **0**) router to receive high-level routing information that describes links between the backbone and backbone routers having at least one interface connected to the backbone. ALTO server **12** may use the high-level routing information alone to estimate IGP metrics between next hops of PID pairs, where a next hop of a PID is a router that is a next hop for prefixes of the PID. PID next hops may be specified in ALTO network maps as next hop attributes of PID entries. ALTO server **12** may establish additional IGP peering sessions with other IGP routers in one or more non-backbone areas to receive lower-level routing information for links encompassed by the respective non-backbone area. The ALTO server **12** may then use a combination of high-level and lower-level routing information to compute IGP metrics between next hops of PID pairs. Each IGP peering session between ALTO server **12** and an IGP router in a non-backbone area may comprise a virtual link such as, for example, a Generic Routing Encapsulation (GRE) tunnel.

[0061] In some instances, AS **4**A may comprise a Level 1/Level 2 IS-IS network. In these instances, ALTO server **12** establishes IGP peering session **7** with at least one Level 2 router to receive high-level routing information that describes links between the backbone and backbone routers having at least one interface connected to the backbone. ALTO server **12** may use the high-level routing information alone to estimate IGP metrics between next hops of PID pairs. ALTO server **12** may establish additional IGP peering sessions with other IGP routers in one or more Level 1 areas to receive lower-level routing information for links encompassed by the respective Level 1 area. The ALTO server **12** may then use a combination of high-level and lower-level routing information to compute IGP metrics between next hops of PID pairs. Each IGP peering session between ALTO server **12** and an IGP router in a non-backbone area may comprise a virtual link such as, for example, a Generic Routing Encapsulation (GRE) tunnel. In this instance, the remote Level 1 or Level 2 router supports configuration of an IS-IS adjacency over a GRE tunnel interface.

[0062] In some instances, ALTO server **12** receives, in peering session **9** with route reflector **17**, traffic engineering data distributed by BGP speakers of AS **4**A and/or ASes **4**B-**4**D. The traffic engineering data may include link attributes such as local/remote IP addresses, local/remote interface indices, metrics, link bandwidth, reservable bandwidth, per CoS class reservation state, preemption and Shared Risk Link Groups (SRLG). The traffic engineering data may be encoded within BGP advertisements from the BGP speakers. Further details regarding exchanging traffic engineering data using BGP are described in U.S. Provisional Patent Appl. No. 61/449,499, incorporated above. In these instances, ALTO server **12** receives topology and routing information, including reachability and link information for links connecting autonomous systems **4** router pairs, from BGP speakers in other IGP areas and may therefore avoid IGP peering with internal router **8**B to receive IGP information for AS **4**A. In this way, ALTO server **12** may have a unified interface to AS **4**A and the broader network encompassing ASes **4**.

[0063] Upon receiving routing information, ALTO server **12** uses the information to computes routes and calculates costs between different next hop pairs of AS **4**A, then assigns these costs as ALTO costs to PID pairs of the intra-AS network map that have respective next hop attributes that correspond to the next hop pairs. In other words, for each PID pair of the intra-AS network map, comprised of a first PID having a first next hop and a second PID having a second next hop, ALTO server **12** uses received routing information to compute a route between the routers referred to by the first and second next hop values (e.g., IP addresses), computes a path cost for the route using link metrics (or distances), and assigns a function of the path cost as an ALTO cost for the PID pair. The ALTO cost for a PID pair represents a cost to traverse AS **4**A from the next hop of the first PID of the PID pair to the next hop of the second PID of the PID pair. ALTO server **12** may calculate separate costs to traverse AS **4**A from the first PID to the second PID (i.e., "forward metric") and from the second PID to the first PID (i.e., "backward metric").

[0064] For example, ALTO server **12** may receive a current LSDB from internal router **8**B executing a link state routing protocol, apply a shortest path first (SPF) algorithm to the LSDB to compute a shortest path tree for a PID of the intra-AS network map, and use path costs associated with the branches of the shortest path trees to calculate ALTO costs between the PID and other PIDs of the intra-AS network map. As another example, ALTO server **12** may build a routing table using routing information obtained from internal router **8**B by operating as a passive listener of a distance vector protocol. The ALTO server **12** uses the routes and distances specified in the routing table to calculate ALTO costs between the PID and other PIDs of the intra-AS network map.

[0065] ALTO server **12** may include one or more ALTO cost policies that define formulas for calculating ALTO cost values for a particular metric. Such policies may incorporate other parameters, in addition to the IGP metric, into the formulas. For example, ALTO server **12** may receive traffic engineering parameters as configuration data or within extended LSAs of OSPF, for instance. For example, ALTO server **12** may receive path costs that represent traffic engineering parameters encoded as Network Layer Reachability Information (NLRI) attributes of a BGP UPDATE message. In general, routers use traffic engineering parameters to create the traffic engineering database, which is used by Constrained Shortest Path First (CSPF) to compute Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs). ALTO cost policies may specify incorporating information within the traffic engineering database regarding various links and paths interconnecting routers of AS **4**A when computing ALTO costs for PID pairs of the intra-AS network map. Other parameters for AS **4**A incorporated by ALTO server **12** may include link delays or load on router interfaces for routers of AS **4**A. Some parameters may be received from sources external to AS **4**A, such as application or other content servers attached to AS **4**A in one of prefixes **20**. These parameters may include load and response times for the servers, for instance.

[0066] ALTO server **12** assembles the calculated ALTO cost values into an intra-AS cost map for AS **4**A. The intra-AS cost map defines inter-PID routing costs for connections among the various PIDs of the intra-AS network map. ALTO server **12** then provides the intra-AS cost map to ALTO client **18** in ALTO maps update message **13**. Using the intra-AS network map and cost map provided by the ALTO server, application clients and content request routers (e.g., executing on host **10**) select serving resources to minimize costs, as specified by the ALTO maps, between content requesting clients and available resources of AS **4**A.

[0067] In some instances, ALTO server **12** may generate different costs for multiple cost maps that reflect a particular service provider objective. For example, ALTO server **12** may generate a first intra-AS cost map that specifies ALTO costs calculated by the ALTO server to minimize delay. Such a cost map may improve the performance of voice applications, for example, when provided to ALTO client **18** of host **10**. ALTO server **12** may generate second intra-AS cost map that specifies ALTO costs calculated by the ALTO server to maximize bandwidth. This cost map may improve performance of video or other high-bandwidth applications.

[0068] In some embodiments, ALTO server **12** additionally generates ALTO network and cost maps that incorporate additional elements of network **2** beyond AS **4**A. ALTO server **12** generates these "inter-AS" network and cost maps from the perspective of AS **4**A, for that autonomous system encompasses ALTO server **12**, and routing information known to ASBRs **6** of AS **4**A may become known to ALTO server **12**. As stated previously, network **2** may represent the Internet. In contrast to intra-AS network and cost maps, which may include a detailed rendering of application-relevant topology and costs within AS **4**A, inter-AS network maps may not include all Internet prefixes, and the division of such prefixes that are included into PIDs may be coarse-grained. Further, inter-AS cost maps may be sparse due to limited external routing information available to AS **4**A. As with intra-AS cost maps, ALTO server **12** may generate different costs for multiple inter-AS cost maps that reflect a particular service provider objective.

[0069] ASBRs **6** execute an exterior gateway protocol (e.g., BGP) to exchange, in one of external peering sessions **11** with ASBRs of ASes **4**A and **4**D, routing information for respective autonomous systems **4**. The exterior gateway protocol is hereinafter described with respect to exterior BGP (BGP). ASBRs **6** thus receive topology information advertisements in the form of BGP UPDATE messages from ASBRs of ASes **4**A and **4**D that include routes to prefixes **21**. A BGP UPDATE message associates one of prefixes **21** with a NEXT_HOP for the prefix and a list of autonomous systems that must be traversed to reach the prefix (the AS_PATH). The AS_PATH and identifiers for prefixes **21** may be expressed within BGP UPDATE messages within Network Layer Reachability Information (NLRI). ASBRs **6** redistribute routing information received in BGP UPDATE messages to an IGP via peering sessions **9**. As stated above, the IGP may comprise IBGP and in such instances peering sessions **9** comprise IBGP peering sessions. In the illustrated embodiment, route reflector **17** re-advertises the routing information to ALTO server **12**. As a result of internal distribution of externally originated routing information by IGP routers of AS **4**A, ALTO server **12** receives the routing information that includes routes to prefixes **21** of ASes **4**B-**4**D. ALTO server **12** may also be administratively configured with routing information, such as routes to prefixes **21**, as well as metrics for inter-AS links (e.g., a cost for a communication link that links AS **4**B to AS **4**C, or a default cost for all inter-AS links). ALTO server **12** may comprise a routing table with which to store routing information.

[0070] In some instances, ASBRs **6** receive traffic engineering data advertised by BGP speakers of ASes **4**B-**4**D and encoded with BGP messages. ASBRs **6** provide these messages to route reflector **17**, which re-advertises the traffic engineering data to ALTO server **12**. As a result, ALTO server **12** may receive respective intra-AS and/or inter-AS routing and topology information for ASes **4**B-**4**D.

[0071] ALTO server **12** uses current intra-AS and inter-AS routing information received in a peering session **9** (or administratively configured) to generate/update an inter-AS network map for network **2**. In some instances, ALTO server **12** generates/updates the inter-AS network map by first determining an optimal route for prefixes **20**, **21** for which the ALTO server **12** is aware. For local prefixes **20**, ALTO server **12** performs the techniques described above for generating inter-AS network and cost maps. As a result, an inter-AS network map generated by ALTO server **12** may comprise the substance of an intra-AS network map for AS **4**A generated as in the manner described above.

[0072] To aggregate prefixes **21** originated in remote ASes **4**B-**4**D into PIDs, ALTO server **12** aggregates into a separate PID all prefixes **21** that have, according to the routing information, the same AS_PATH attribute and NEXT_HOP attribute. In other words, ALTO server **12** aggregates into a separate PID any prefixes **21** that are reachable by the same route (AS_PATH) and are also advertised within AS **4**A by the same one of ASBRs **6** (NEXT_HOP). In the example embodiment, ALTO server **12** assigns each of prefixes **21** to a separate PID, because each of the prefixes have a different AS_PATH. In various embodiments, however, multiple different prefixes advertised in separate topology advertisements may be assigned to the same PID according to the above aggregation rules. ALTO server **12** assembles the PIDs into an inter-AS network map for network **2**.

[0073] Dynamically generating an inter-AS cost map for network **2** from the perspective of AS **4**A (and ALTO server **12**) may require inter-AS costs, i.e., a cost between PIDs with prefixes located in different ones of ASes **4**. ALTO server **12** may be configured with a policy or other configuration data with inter-AS costs. In some embodiments, the inter-AS cost applied by ALTO server **12** is identical for all pairs of ASes **4**. In other words, the inter-AS cost is a default inter-AS cost. In some embodiments, however, ALTO server **12** may include specific inter-AS costs for one or more pairs of ASes **4**. For example, ALTO server **12** may be configured with a value for an inter-AS cost between ASes **4**B and **4**C.

[0074] A cost between two remote PIDs (e.g., between a PID in AS **4**B and a PID in AS **4**C) is deemed proportional to the number of inter-AS links between the two remote PIDs when the inter-AS cost is a default value that is larger than the largest intra-AS cost. To determine the number of inter-AS links between the two remote PIDs, ALTO server **12** reads the AS_PATH attribute of the two remote PIDs. The final autonomous system identification number in the AS_PATH attribute value for a PID is the autonomous system that serves the PID (i.e., the autonomous system that originates advertisements for prefixes encompassed by the PID). For example, from the perspective of AS **4**A, an AS_PATH for a PID including prefix **21**B comprises [(Identification number for AS **4**B), (Identification number for AS **4**C)].

[0075] If the AS_PATH attribute for either of the two remote PIDs includes the identification number for the autonomous system that serves the other remote PID, ALTO server **12** computes the inter-AS cost as the product of the default inter-AS cost (C) and the number of inter-AS links, according to the AS_PATH attribute, that must be traversed to reach the PID from the other remote PID. In the above example, according to the AS_PATH attribute for the PID comprising prefix **21**B of AS **4**C, a PID comprising prefix **21**A of AS **4**B traverses one inter-AS link to reach the PID comprising prefix **21**B. The inter-AS cost between the two PIDs is therefore equal to 1×C.

[0076] ALTO server **12** computes costs in this manner between each of the remote PIDs represented in the inter-AS network map and stores the costs into a corresponding inter-AS cost map. In some instances, ALTO server **12** includes a policy or other configured data that specifies actual, rather than default, inter-AS costs between pairs of ASes **4**. In such instances, ALTO server **12** may sum the inter-AS costs between pairs of ASes **4** represented in an AS_PATH attribute of a remote PID to compute a total cost between two remote PIDs.

[0077] To determine costs between two PIDs each encompassing one or more of local prefixes **20** (i.e., "local" PIDs from the perspective of AS **4**A), ALTO server **12** performs the techniques described above for generating intra-AS cost maps and incorporated computed costs for local PID pairs into the inter-AS cost map. As a result, an inter-AS cost map generated by ALTO server **12** may comprise the substance of an intra-AS cost map for AS **4**A generated as in the manner described above.

[0078] For a mixed local-remote PID pair combination, ALTO server **12** supplements the inter-AS cost from AS **4**A to the remote PID with the intra-AS cost from local PID to the NEXT_HOP of the remote PID. In the illustrated embodiment, the NEXT_HOP of the remote PID is one of ASBRs **6**. In other words, ALTO server **12** combines the inter-AS cost with the intra-AS (IGP) cost to determine a total cost to traverse a route from the local PID of the PID pair to the remote PID. The inter-AS cost from AS **4**A, where ALTO server **12** applies the default inter-AS cost, C, to inter-AS links, is the product of the length of the AS_PATH attribute value (i.e., the number of ASes in the AS_PATH) for the remote PID and C. For example, the inter-AS cost between a local PID and a remote PID comprising prefix **21**B is equal to 2×C, for the AS_PATH includes AS **4**B, **4**C and thus has length two.

[0079] ALTO server **12** sums the intra-AS and inter-AS cost to determine a total cost, then add the total cost for the local-remote PID pair to the inter-AS cost map for network **2**. In some instances, ALTO server **12** may include ALTO cost policies that define formulas for calculating total cost values based on inter-AS and intra-AS costs. Such policies may incorporate other parameters, in addition to these costs, as described above with respect to using ALTO cost policies in the exclusive intra-AS context.

[0080] Dynamically generating and updating ALTO network and costs maps with an ALTO server using routing information received from network elements synchronizes the maps to an ever-changing network environment. As a result, the network and cost maps provided by ALTO server **12** to ALTO client **18** may reflect recent updates to the network topology and/or utilization and may thus improve node selection and increase performance by one or more applications of host **10**. Moreover, automatically creating network and cost maps may reduce an ALTO server **12** configuration burden on the administrator of AS **4**A, making an ALTO server implementation viable in a large-scale service provider environment. The configuration burden may be, rather, distributed to respective administrators of the various ASes **4**. In this way, the administrator "close to" a particular one of subnets **20**, **21** can be responsible for the treatment of that subnet by ALTO server **12** and ALTO client **18**. Because the respective administrators of remote ASes **4**B-**4**D typically have a fuller knowledge of the configuration of the remote ASes than does the administrator of AS **4**A, the techniques may thus improve subsidiarity among network entities and operators that cooperate to facilitate content distribution, thereby relieving configuration pressures on the administrator of ALTO server **12**.

[0081] FIG. **2** is a block diagram illustrating an example graph **25** that represents a combined ALTO inter-AS network and cost map for network **2** of FIG. **1** that are dynamically generated by ALTO server **12** according to the described techniques. PIDs **22**A-**22**F ("PIDs **22**") each encompass one or more prefixes and thus represent the PIDs

of an inter-AS network map for network **2**. Edges interconnecting PIDs and each annotated with an inter-PID cost represent an inter-AS cost map for network **2**.

[0082] Performing the techniques described above with respect to FIG. **1**, ALTO server **12** aggregates prefix **20**A into PID **22**A because the next hop for prefix **20**A is ASBR **6**A. Similarly, ALTO server **12** aggregates both prefixes **20**C, **20**D into PID **22**C because the shared next hop for these prefixes is internal router **8**B. Topology advertisements originated by respective ones of ASes **4**B-**4**D include prefixes that are placed by ALTO server **12** into a separate PID. In other words, in this instance, ALTO server **12** aggregates all prefixes served by the same one of ASes **4**B-**4**D into the same PID. For example, PID **22**E includes prefix **21**B advertised by AS **4**C and, in other configurations, would further include any other prefixes advertised by AS **4**C.

[0083] ALTO server **12** computes intra-AS costs for local PIDs, i.e., PIDs that encompass prefixes **20** served by AS **4**A, using routing information received in an IGP session. Graph **25** annotates intra-AS costs on edges using an instance of A_i. For example, the intra-AS cost between local PID **22**A and local PID **22**B is A_3. Additionally, ALTO server **12** in this instance uses a default inter-AS cost, C, as the transit routing cost between each pair of neighboring ASes **4**. For example, a cost between PID **22**D (here aggregating AS **4**B) and PID **22**E (here aggregating AS **4**C) is C because the represented ASes **4**B, **4**C are neighbors. As another example, a cost between PID **22**E and PID **22**A (the next hop attribute of which is the next hop to AS **4**B, i.e., ASBR **6**A) is 2×C.

[0084] For a mixed local-remote PID **22** pair combination, ALTO server **12** supplements the inter-AS cost from the next hop of AS **4**A to the remote PID with the intra-AS cost from local PID to the NEXT_HOP of the remote PID. In this instance, ALTO server **12** sums the inter-AS cost and intra-AS costs values to compute a total inter-PID cost. For example, a cost, $C+A_1$, between PID **22**D and PID **22**C is a sum of the cost, C, between PID **22**D and PID **22**A and the cost, A_1, between PID **22**A and PID **22**C.

[0085] Graph **25** does not include an edge to connect PID **22**D and PID **22**F because ALTO server **12** does not receive routes that include an AS_PATH comprising both AS **4**B and AS **4**D. ALTO server **12** is therefore unable to determine the costs between remote AS **4**B and remote AS **4**D.

[0086] FIG. **3** is a block diagram illustrating network **36** comprising ALTO servers **38**A-**38**C ("ALTO servers **38**") that perform federated ALTO techniques described herein. Network **36** includes autonomous systems **40**A-**40**C ("ASes **40**") that each includes a respective one of ALTO servers **38**. Autonomous systems **40** are interconnected via external communication links **44**. Each of ASes **40** may represent an example embodiment of one of autonomous systems **4** of FIG. **1**. Each of ALTO servers **38** may represent an example embodiment of ALTO server **12** of FIG. **1**. Various combinations of ASes **40** may be under administrative control of one or more administrators or service providers.

[0087] Each of ALTO servers **38** performs dynamic network map generation and modification techniques described in this disclosure to aggregate prefixes (not shown in FIG. **3**) served by associated ASes **40** into one or more of PIDs **42** and assemble the PIDs into an inter-AS network map and, in some instances, into an intra-AS network map. As illustrated, for example, ALTO server **38**A aggregates prefixes served by AS **40**A into PIDs **42**A, **42**B, prefixes of which are connected by an intra-AS communication link. That is, prefixes of PID **42**A are connected by an inter-AS communication link to prefixes of PID **42**B. Each of ALTO servers **38** additionally performs dynamic cost map generation and modification techniques described herein to produce an inter-AS cost map that includes costs for various combinations of local and remote PIDs associated with the network maps produced by the ALTO server. As a result, each of ALTO servers **38** produces local network and local cost maps that represent a topology of network **36** from the perspective of the associated one of autonomous systems **40** for the ALTO server.

[0088] ALTO servers **38** federate with one another in an ALTO federation to share information to foster fine PID prefix granularity throughout network **36** and thereby improve node selection. ALTO server **38**B is a master ALTO server that, in addition to its local functions (i.e., generating network and cost maps from the perspective of AS **40**B), generates master network and cost maps for network **36** using the various local intra-AS and/or inter-AS network and cost maps generated by each of ALTO servers **38**. ALTO server **38**B (hereinafter, "master ALTO server **38**B") may be administratively configured as a master ALTO server **38**B or may be elected during a negotiation between eligible ALTO servers **38**. In some configurations, multiple ALTO servers **38** may operate as a master ALTO server.

[0089] To enable master ALTO server **38**B to determine, within the many network maps master ALTO server **38**B generates and receives, the originating one of autonomous systems **40** for each PID in the network maps, ALTO servers **38** add an autonomous system identifier ("AS ID") attribute to PIDs of their respective network maps. The autonomous system identifier may comprise the registered identification number for the originating one of ASes **40** for each of the PIDs (including PIDs **42**). In other words, each of PID that encompasses a particular one or more prefixes is "tagged" with an AS ID for the one of ASes **40** that originated the prefixes. Each of ALTO servers **38** may therefore tag PIDs in its network maps with different AS IDs. For example, ALTO server **38**A tags PIDs **42**A, **42**B in the inter-AS network map generated by the ALTO server with an AS ID for AS **40**A. In some instances, ALTO server **38**A may additionally tag any remote PIDs of the inter-AS network map with AS IDs for the autonomous systems that originated prefixes of the remote PIDs.

[0090] Each one of local PIDs **42** in an inter-AS network map thus carries an identity of the one of ASes **40** that includes the one of ALTO servers **38** that generated the PID. Because the local perspective of each prefix included in one of PIDs **42** represents the most precise topological grouping for the prefix, PIDs from the originating one of ASes **40** should be migrated to the master network map. Put another way, by identifying the originating one of ASes **40** for each of PIDs **42**, PIDs at the finest level of prefix granularity may be identified and selected by master ALTO server **38**B for inclusion in a master network map. ALTO servers **38**A, **38**C send their respective, inter-AS network and cost maps to ALTO server **38**B in respective upload messages **46**A, **46**B.

[0091] In some embodiments, rather than adding an AS ID attribute to each PID in inter-AS network maps, each of ALTO servers **38** generate both intra-AS and inter-AS network maps. ALTO servers **38**A, **38**C then send their respective, network and cost maps to ALTO server **38**B in respective upload messages **46**A, **46**B. In these embodi-

ments, upload messages **46**A, **46**B additionally include an AS ID for the one of ASes **40** that includes the sending one of ALTO servers **38**A, **38**C. For example, ALTO server **38**A sends inter-AS and intra-AS network maps, an inter-AS cost map, and an AS ID for AS **40**A in upload message **46**A to master ALTO server **38**B. When master ALTO server **38**B receives respective upload message **46**A, **46**B from one of ALTO servers **38**A, **38**C in these embodiments, the master ALTO server **38**B identifies local PIDs within the inter-AS network map by correlating the prefixes of PIDs within the inter-AS network map to prefixes of PIDs within the intra-AS network map. When a prefix is included within a PID of the intra-AS network map, any PID of the inter-AS network map that also includes the prefix is a local PID. Master ALTO server **38**B then tags identified local PIDs with the AS ID received along with the local network maps. This technique may reduce a size of upload messages **46**.

[0092] ALTO servers **38**A, **38**C send their respective, local network and cost maps to ALTO server **38**B in respective upload messages **46**A, **46**B. Master ALTO server **38**B generates a local network and cost map for AS **40**B. Responsive to receiving local network and cost maps, master ALTO server **38**B generates the master network and cost map for the ALTO federation of network **36**. Master ALTO server **38**B may correlate perspectives from each of the local network and cost maps received or generated into a single, consolidated master network and cost map. Master ALTO server **38**B then sends the master network and cost maps to ALTO server **38**A, **38**C in respective download messages **47**A, **47**B.

[0093] FIG. **4**A is a block diagram illustrating an example graph **54** that represents a partial combined master network and cost map for network **36** of FIG. **3** that is generated by master ALTO server **38**B, according to the described techniques, from graphs **50**A, **50**B that represent partial combined inter-AS network and cost maps for network **36** generated by and from the perspective of respective ALTO servers **38**A, **38**C. Graphs **54** and **50**A-**50**B are "partial" in that they do not include elements of AS **40**B (for ease of illustration). In some instances, graphs **50**A-**50**B include additional elements to incorporate the elements of AS **40**B of FIG. **3**. In some instances, graph **54** includes additional elements to incorporate a perspective of ALTO server **38**B and/or the elements of AS **40**B included in various instances of graphs **50**A-**50**B.

[0094] ALTO server **38**A generates an inter-AS network map from a perspective of AS **40**A that include the elements illustrated in graph **58**. Specifically, the inter-AS network map includes local PIDs **42**A, **42**B. PID **42**A has PID identifier "PID**1**" and encompasses prefix P**1**. PID **42**B has PID identifier "PID**2**" and encompasses prefix P**2**. The inter-AS network map additionally includes remote PID **52**A that, in this instance, represents an aggregation of prefixes of AS **40**C that are advertised to AS **40**A in, for example, BGP UPDATE messages. These prefixes include P**3**, P**4**, and P**5**. ALTO server **38**A generates an inter-AS cost map that includes cost map entries for the inter-AS network map. Graph **50**A illustrates computed inter-PID costs as arrows. For example, a cost from PID **42**A to PID **52**A is 2C, the cost computed by ALTO server **38**A according to the techniques described herein for traversing two inter-AS (or "transit") communication links **44** between AS **40**A and AS **40**C. ALTO server **38**A additionally computes an inter-PID cost, A**1**, for the PID pair <PID **42**A, **42**B> using an IGP metric using the techniques of this disclosure.

[0095] ALTO server **38**C generates an inter-AS network map from a perspective of AS **40**C in a similar manner. Although not shown in FIG. **4**A, each of PIDs **42**A-**42**B, **42**E-**42**F, and **52**A, **52**B may include an AS ID for the originating one of ASes **40** for respective, encompassed prefixes. For example, local PID **42**E may be tagged with an AS ID for AS **40**C in the network map represented by graph **50**B. As another example, remote PID **52**A may be tagged with an AS ID for AS **40**C in the network map represented by graph **50**A.

[0096] In this example, all prefixes of AS **40**C (that is, prefixes P**3**, P**4**, and P**5**) are advertised to AS **40**A and aggregated into PID **52**A by ALTO server **38**A. Similarly, all prefixes of AS **40**A (this is, prefixes P**1** and P**2**) are advertised to AS **40**C and aggregated in PID **52**B by ALTO server **38**C.

[0097] Master ALTO server **38**B receives the inter-AS network and costs maps represented by graphs **50**A, **50**B from respective ALTO servers **38**A, **38**C and generates the master network and cost maps represented by graph **54**. To generate a master network map from multiple inter-AS network maps, master ALTO server **38**B copies all local PIDs of the inter-AS network maps into the master network map. Each of ALTO servers **38** receives the finest granular routing information for its respective one of ASes **40** and therefore has the fullest information base with which to generate local PIDs and compute local inter-PID costs.

[0098] Additionally, master ALTO server **38**B reconciles each remote PID in each one of the inter-AS network maps with local PIDs in the inter-AS network map for the originating one of ASes **40** that advertises the prefixes of the remote PID. In other words, master ALTO server **38**B intersects the prefixes of remote PIDs with prefixes of corresponding local PIDs for the originating (i.e., remote) one of ASes **40** to facilitate the finest available level of PID granularity for the PIDs of the master network map. Considering both a local PID and a remote PID as respective mathematical sets of prefixes, the resulting PIDs for a master network map consist of (1) a PID that includes the relative complement of the remote PID in the local PID (i.e., the set of prefixes in the local PID but not in the remote PID) and (2) the intersection of the remote PID and the local PID (i.e., the set of prefixes in both the remote PID and the local PID). However, if any of these resulting sets of prefixes is null, master ALTO server **38**B does not generate a corresponding PID for the null set. Moreover, if any of these resulting sets is already encompassed by a PID of the master network map, master ALTO server **38**B may not create a second PID for the same prefix set.

[0099] Typically, because ALTO servers **38** generate few remote PIDs for a remote AS, master ALTO server **38**B simply copies to the master network map the local PIDs in the inter-AS network map for the originating one of ASes **40** that advertises the prefixes of the remote. In some instances, however, master ALTO server **38**B must reallocate prefixes of a local PID of an inter-AS network map into two or more PIDs for inclusion in master network map. Examples of these instances are described in detail below with respect to FIG. **4**B.

[0100] In the illustrated example, master ALTO server **38**B reconciles remote PID **52**A of graph **50**A with local PIDs **42**E-**42**F of graph **50**B. Because PID **52**A includes

every prefix represented in PIDs **42**E-**42**F, PIDs **42**E-**42**F represent the finest level of granularity for the prefixes, are reachable from each of PIDs **42**A-**42**B with identical costs, and are thus copied by master ALTO server **38**B directly to the master network map (represented by PIDs of graph **54**).

[0101] As an example in the other network traffic direction, master ALTO server **38**B reconciles remote PID **52**B of graph **50**B with local PIDs **42**A-**42**B of graph **50**A. PID **52**B includes every prefix represented in PIDs **42**A-**42**B and is reachable from each of PIDs **42**E-**42**F at identical cost. Master ALTO server **38**B therefore directly copies PIDs **42**A-**42**B to the master network map (represented by PIDs of graph **54**).

[0102] As seen in graph **64**, master ALTO server **38**B does not include remote PIDs **52**A, **52**B in the master network map. Master ALTO server **38**B only carries over PIDs for prefixes originated in one of ASes **40** (i.e., local PIDs) into the master network map. In some instances, master ALTO server **38**B may modify PID identifiers when copying a local PID into the master network map. For example, PID **42**A in both graph **60**A and graph **64** has a PID identifier "PID**1**." In some instances, master ALTO server **38**B may provide a network PID identifier for PID **42**A in graph **64**.

[0103] To generate a master cost map from the multiple inter-AS network and cost maps provided by ALTO servers **38**, master ALTO server **38**B uses inter-PID costs computed from a perspective of an AS from which network traffic will originate. For PID pairs in which both members of the pair are local to a particular one of the inter-AS network maps, master ALTO server **38**B uses the inter-PID costs specified in the corresponding inter-AS cost map. In the illustrated example, for instance, master ALTO server **38**B sets A**1** as a cost between PIDs **42**A, **42**B of the master network map upon determining A**1** as the cost between local PIDs **42**A, **42**B of the inter-AS network map represented by graph **50**A. As in the inter-AS cost map, any PID pair of the master cost map may be associated with two unidirectional costs representing the costs to traverse a network in both directions between the member PIDs of the PID pair.

[0104] For PID pairs in which different ASes **40** originate the member PIDs of the PID pair, for each network traffic direction, master ALTO server **38**B uses the cost as specified by the one of ALTO servers **38** for the respective one of ASes **40** that originates the traffic. For example, to identify the ALTO cost to traverse the network from PID **42**A (including prefix P**1** of AS **40**A) to PID **42**E (including prefix P**3** of AS **40**C), master ALTO server **38**B queries the inter-AS cost map provided by AS **40**A to determine the inter-PID ALTO cost between local PID **42**A and remote PID **52**A (including prefix P**3** of AS **40**C). This inter-AS cost map is represented in graph **50**A, which indicates this inter-PID ALTO cost is 2C. Master ALTO server **38**B sets the cost from PID **42**A to **42**E within the master cost map to 2C.

[0105] In the other traffic direction, that is, from PID **42**E to PID **42**A, master ALTO server **38**B queries the inter-AS cost map provided by AS **40**C to determine the inter-PID ALTO cost between local PID **42**E and remote PID **52**B (including prefix P**1** of AS **40**A). Graph **50**B that represents a partial inter-AS cost map provided by AS **40**C indicates the cost is 2C. Master ALTO server **38**B therefore sets the cost from PID **42**E to **42**A within the master cost map to 2C.

[0106] Graph **54** represents all inter-PID costs as bi-directional arrows for ease of illustration because, in the illustrated embodiment, inter-PID costs in both directions are equal. In some instances, a cost to traverse a network between any two PIDs in one direction differs from the cost to traverse a network between the PIDs in the reverse direction. As a result, the master cost map may include a respective cost entry for each direction for the two PIDs.

[0107] FIG. **4**B is a block diagram illustrating variant partial combined network and cost maps of FIG. **4**A for an embodiment of network **36** of FIG. **3** in which AS **40**C does not advertise prefix P**4** to AS **40**A. Graph **64** of FIG. **4**B represents a partial combined master network and cost map for network **36** that is generated by master ALTO server **38**B, according to the described techniques, from graphs **60**A, **60**B that represent partial combined inter-AS network and cost maps for network **36** generated by and from the perspective of respective ALTO servers **38**A, **38**C.

[0108] In this illustrated embodiment, unlike the embodiment illustrated in FIG. **4**A, remote PID **62**A of graph **60**A does not include prefix P**4** because AS **40**C does not advertise prefix P**4** to AS **40**A. Autonomous system **40**A comprises ALTO server **38**A that generates the partial inter-AS network and cost maps represented by graph **60**A.

[0109] Because prefix P**4** of AS **40**C is not advertised and therefore not reachable from AS **40**A, master ALTO server **38**B allocates prefixes P**4**, P**5** of PID **42**F to separate PIDs **66**A, **66**B for the master network map represented by graph **64**. This operation comprises both the intersection and relative complement operations described above with respect to FIG. **4**A. That is, master ALTO server **38**B creates for the master network map (1) PID **66**A encompassing the set of prefixes that is an intersection of the set of prefixes of remote PID **62**A of graph **60**A and the set of prefixes of local PID **42**F of graph **60**B (i.e., prefix P**5**), and (2) PID **66**B encompassing the set of prefixes that is a relative complement of the set of prefixes of remote PID **62**A of graph **60**A in the set of prefixes of local PID **42**F of graph **60**B (i.e., prefix P**4**).

[0110] Master ALTO server **38**B sets costs in the master cost map that accord with the costs of the inter-AS cost maps. Because ALTO server **38**C aggregates prefixes P**4**, P**5** into a single PID **42**F, the ALTO cost between these prefixes is zero. Master ALTO server **38**B therefore sets the ALTO cost between PIDs **66**A, **66**B that each encompass one of prefixes P**4**, P**5** to zero in the network cost map represented by graph **64**. Similarly, the ALTO cost between PIDs **42**E, **42**F is set to B**1** according to graph **60**B. Master ALTO server **38**B therefore sets the ALTO cost between PIDs **42**E, **66**A and between PIDs **42**E, **66**B to B**1** in the network cost map represented by graph **64**, for PIDs **66**A and **66**B include reallocated prefixes of PID **44**F.

[0111] Because there is no available path from prefixes of AS **40**A to prefix P**4** of AS **40**C, the ALTO costs from PIDs **42**A, **42**B to PID **66**B is set to infinity in the master cost map. This is illustrated in graph **64** by the absence of a directional arrow from either of PIDs **42**A, **42**B to PID **66**B. However, because AS **40**A prefixes P**1**, P**2** to AS **40**C, a directional arrow illustrating cost 2C connects PID **66**B to each of PIDs **42**A, **42**B to represent these costs in the master cost map generated by master ALTO server **38**B. Some instances of master ALTO server **38**B may set the ALTO cost from PID **66**B to each of PIDs **42**A, **42**B to infinity because most applications require bi-directional communication and that an ALTO client should not therefore select either of PIDs **42**A, **42**B to serve a client in PID **66**B.

[0112] Various other embodiments of network **36** of FIG. **3** may require PID prefix reallocation from inter-AS network maps by master ALTO server **38**B when generating the master network map for network **36**. For example, in some embodiments, two or more remote PIDs of a first inter-AS network map may together encompass two or more prefixes that are aggregated at the originating one of ASes **40** for these prefixes into a single local PID of a second inter-AS network map for the originating AS. This may occur where, for example, traffic engineering parameters specify a different autonomous system path (AS_PATH) for reaching each of the remote PIDs from local PIDs of the first inter-AS network map. In other words, the prefixes in the remote PIDs are associated with different AS_PATH lengths from the autonomous system that includes the one of ALTO servers **38** that generated the first inter-AS network map.

[0113] In this instance, master ALTO server **38**B reallocates the prefixes of the single local PID of the second inter-AS network map into two or more master network map PIDs that each encompasses a different set of the prefixes, grouped according to similar AS_PATH lengths as specified by the two or more remote PIDs of the first inter-AS network map. In this way, master ALTO server **38**B aggregates all prefixes from remote PIDs of the first inter-AS network map that have the same AS_PATH length into the same PID for the master network map. Because the various prefixes are aggregated into a single PID in the second inter-AS network map, the ALTO cost between these prefixes is zero. Master ALTO server **38**B therefore sets the ALTO cost between the various pairs of the two or more master network map PIDs generated in this manner to zero. Master ALTO server **38**B sets the ALTO cost to the two or more "split" master network map PIDs from the PIDs of the AS of the first inter-AS network map according to the cost specified in the first inter-AS network map to remote PIDs of the first inter-AS network map that include prefixes of the "split" master network map PIDs from the PIDs of the AS.

[0114] As another example, two or more remote PIDs of a first inter-AS network map may together encompass two or more prefixes that are aggregated at a neighboring, originating one of ASes **40** into a single local PID of a second inter-AS network map for the neighboring, originating AS. As described in further detail below with respect to FIG. **5**, this may occur where, for example, the ALTO server **38** that generates the first inter-AS network map receives routing information that specifies different paths from local PIDs of the first inter-AS network map to the prefixes of the neighboring AS due to multiple ASBRs connecting the AS of the first inter-AS network map to the neighboring AS.

[0115] In this instance, master ALTO server **38**B reallocates the prefixes of the single local PID of the second inter-AS network map into two or more master network map PIDs that each encompasses a different set of the prefixes, grouped according to reachability from the AS of the first inter-AS network map by respective ones of the different paths. In this way, master ALTO server **38**B aggregates all prefixes from remote PIDs of the first inter-AS network map that have a same cost from the AS of the first inter-AS network map into the same PID for the master network map. Because the various prefixes are aggregated into a single PID in the second inter-AS network map, the ALTO cost between these prefixes is zero. Master ALTO server **38**B therefore sets the ALTO cost between the various pairs of the two or more master network map PIDs generated in this manner to zero. Master ALTO server **38**B sets the ALTO cost to the two or more "split" master network map PIDs from the PIDs of the AS of the first inter-AS network map according to the cost specified in the first inter-AS network map to remote PIDs of the first inter-AS network map that include prefixes of the "split" master network map PIDs from the PIDs of the AS.

[0116] As another example, a remote PID of a first inter-AS network map may encompass a prefix that encompasses sub-prefixes (e.g., subnets) that are aggregated at a neighboring, originating one of ASes **40** into multiple local PIDs of a second inter-AS network map for the neighboring, originating AS. This circumstance may occur due to prefix aggregation, for instance. In this example, the multiple local PIDs of the second inter-AS network map control PID allocation within the master network map.

[0117] FIG. **5** is a block diagram illustrating an example network **79** comprising ALTO servers **81**A-**81**B ("ALTO servers **81**") that perform federated ALTO techniques described herein. Network **79** includes autonomous systems **80**A-**80**B ("ASes **80**") that each includes a respective one of ALTO servers **81**. Autonomous systems **80** are interconnected via respective external communication links (not shown) that connect pairs of autonomous system boundary routers of ASes **80**. For example, a first external communication link communicatively couples "ASBR**2**" (encompasses by PID **84**B) of AS **80**A and "ASBR**4**" (encompassed by PID **84**D) of AS **80**B. Each of ASes **80** may represent an example embodiment of one of autonomous systems **4** of FIG. **1**. Each of ALTO servers **81** may represent an example embodiment of ALTO server **12** of FIG. **1**. Each of ASes **80** may be under administrative control of one or more administrators or service providers. While illustrated as and described as ASBRs, any of the ASBRs may comprise area boundary routers or any external BGP speaker.

[0118] Each of ALTO servers **81** performs dynamic network map generation and modification techniques described in this disclosure to aggregate prefixes advertised by routers of associated ASes **80** into one or more of PIDs **82** or PIDs **84** and assemble the PIDs into a respective inter-AS network map for the AS that includes the ALTO server. As illustrated, for example, ALTO server **81**A aggregates prefix P**1** advertised by an internal router of AS **80**A into PID **82**A. The internal router of AS **80**A may be set as a next hop attribute of PID **82**A. ALTO server **81**A additionally creates respective PIDs **84**A, **84**B for prefixes (in this case, network addresses) of each of the ASBRs of AS **80**A. ALTO server **81**B aggregates prefixes P**2**, P**3** advertised by an internal router of AS **80**B into PID **82**B. The internal router of AS **80**B may be set as a next hop attribute of PID **82**B. ALTO server **81**B additionally creates respective PIDs **84**C, **84**D for prefixes (in this case, network addresses) of each of the ASBRs of AS **80**B. V

[0119] The various PIDs **82**, **84** of each respective one of ASes **80** are connected by intra-AS communication links. Each of ALTO servers **81** additionally performs dynamic cost map generation and modification techniques described herein to produce an inter-AS cost map that includes costs for various combinations of local and remote PIDs associated with the network maps produced by the ALTO server. As a result, each of ALTO servers **81** produces local network and local cost maps that represent a topology of network **79** from the perspective of the associated one of autonomous systems **80** for the ALTO server.

[0120] ALTO servers **81** federate with one another in an ALTO federation to share information to foster fine PID prefix granularity throughout network **79** and thereby improve node selection. ALTO server **81**A is a master ALTO server that, in addition to its local functions (i.e., generating network and cost maps from the perspective of AS **80**A), generates master network and cost maps for network **79** using the various local intra-AS and/or inter-AS network and cost maps generated by each of ALTO servers **81**.

[0121] ALTO server **81**A receives routing information that advertises prefixes P**2**, P**3** of AS **80**B via a chain of topology information advertisements by routers of ASes **80**. In one example, the internal router of AS **80**B sends an IBGP UPDATE messages **86**A, **86**B that each includes prefixes P**2**, P**3** to the ASBRs of the AS **80**B. The ASBRs of AS **80**B reissues the messages as respective BGP UPDATE message **88**A, **88**B to the ASBRs of AS **80**A. The ASBRs, in turn, reissue the BGP UPDATES message **80**A, **88**B as IBGP UPDATE messages **90**A-**90**D. ALTO server **81**A receives IBGP UPDATE message **90**C from the ASBR encompassed by PID **84**A, which is a next hop for a first route for prefixes P**2**, P**3**. ALTO server **81**A receives IBGP UPDATE message **90**D from the ASBR encompassed by PID **84**B, which is a next hop for a second route for prefixes P**2**, P**3**. ALTO server **81**A also receives routing information for prefix P**1** of AS **80**A.

[0122] ALTO server **81**A also receives routing information that specifies link metrics for internal links of AS **80**A and uses these metrics to compute ALTO costs from PID **82**A to PID **84**A and from PID **82**A to PID **84**B. In addition, ALTO server **81**A is configured with inter-AS costs for each connected pair of ASBRs of the ASes **80**, which the ALTO server uses to compute ALTO costs from PID **84**A to PID **84**C and from PID **84**B to PID **84**D. These ALTO costs are illustrated in FIG. **5** as unidirectional arrows with annotated cost information.

[0123] ALTO server **81**B receives routing information for prefixes P**2**, P**3** as well as routing information that specifies link metrics for internal links of AS **80**A. ALTO server **81**B uses the metrics to compute ALTO costs from PID **84**C to PID **82**B and from PID **84**D to PID **82**B. In accordance with the techniques of this disclosure, ALTO server **81**B then uses the received routing information to generate inter-AS network and cost maps for AS **80**B. ALTO server **81**B sends the generated inter-AS network and cost maps for AS **80**B to ALTO server **81**A in upload message **92**.

[0124] Because multiple network paths from PID **82**A to PID **82**B exist, ALTO server **81**A determines the network path along which the routers of AS **80**A will forward traffic to determine an accurate ALTO cost between the PIDs. Using the routing information received from the routers of AS **80**A, ALTO server **81**A selects one of the routes to prefixes P**2**, P**3** according to a decision process, such as the BGP decision process described in Rekhter et al., referenced above. So long as the policy configuration of ALTO server **81**A and the internal router of AS **80**A are similar regarding the decision process, the decision process accurately determines the path for traffic from PID **82**A toward PID **82**B. In some embodiments, rather than perform a decision process, ALTO server **81**A accesses the Local Routing Information Base (Loc-RIB) of the internal router of AS **80**A that is the next hop for PID **82**A. The Loc-RIB specifies the route selected by the internal router. In the illustrated embodiment, the path for traffic runs to PID **84**A because of the lower internal routing cost from the internal router of PID **82**A to the ASBR of PID **84**A.

[0125] Having determined the path for traffic from PID **82**A to PID **82**B, ALTO server **81**A sums the various ALTO costs for the path links, which include the link from PID **82**A to PID **84**A, from PID **84**A to PID **84**C, and from PID **84**C to PID **82**B. The ALTO cost for the link from PID **82**A to PID **84**A is specified by the inter-AS network map generated by ALTO server **81**A for AS **80**A, as is the ALTO cost for the link from PID **84**A to PID **84**C. The ALTO cost for the link from PID **84**C to PID **82**B is specified by the inter-AS network map generated by ALTO server **81**B and sent to ALTO server **81**A in upload message **92**. In this instance, the sum of the various ALTO costs equals 190.

[0126] FIG. **6**A is a block diagram illustrating an example graph **94** that represents a partial combined intra-AS network and cost map for AS **80**B of FIG. **5** that is generated by ALTO server **81**B, according to the described techniques. Graph **94** is partial in that it does not include ALTO costs in both directions. In some embodiments, graph **94** is a combined inter-AS network and cost map that incorporates elements of AS **80**A. ALTO server **81**B sends the network and cost maps represented by graph **94** to ALTO server **81**A.

[0127] FIG. **6**B is a block diagram illustrating an example graph **96** that represents a partial combined master network and cost map for network **79** of FIG. **5**. Graph **96** is partial in that it does not include ALTO costs in both directions. ALTO server **81**A generates the master network and costs maps for network **79** using received routing information, configured inter-AS costs, and a network and cost map received from ALTO server **81**B for AS **80**B.

[0128] FIG. **7** is a block diagram illustrating, in detail, an example ALTO server **100** that receives routing information and dynamically generates network and cost maps in accordance with the techniques described herein. ALTO server **100** may represent an example embodiment of any of ALTO servers **12** of FIG. **1**, ALTO servers **38** of FIG. **3**, or ALTO servers **81** of FIG. **5**. ALTO server **100** may be a server, network controller, network switch, or other computing device or appliance that includes one or more microprocessors that provide an operating environment for one or more software modules for dynamically generating and outputting ALTO network and cost maps in accordance with the described techniques. For purpose of clarity, components, such as a microprocessor, memory, keyboard, display, an operating system, network drivers and other components commonly found in such a computing device or appliance are not shown in FIG. **7**. In some embodiments, ALTO server **100** comprises a router that includes one or more services units to apply services such as ALTO services as described herein. The services units may be distributed over one or more service cards or blades (not shown) that are inserted into rack slots of a router. The services units may communicate with the routing plane across a backplane or across a network link that enables the ALTO services to passively peer with routing daemons executing routing protocols within the routing plane of the router comprised by ALTO server **100**.

[0129] Control unit **102** of ALTO server **100** provides an operating environment for executing network map module **106**, cost map module **112**, client interface **114**, IGP listener **130**, BGP listener **124**, and user interface **128**. Control unit **102** may comprise one or more processors (not shown),

including one or more microprocessors, digital signal processors (DSPs), application specific integrated circuits (ASICs), field programmable gate arrays (FPGAs), or any other equivalent integrated or discrete logic circuitry, as well as any combinations of such components, to execute modules that implement the functionality described herein.

[0130] Map modules **104** of ALTO server **100** dynamically generate network map **120** and cost map **122** using routing information received with IGP listener **130** and reachability information, e.g., in the form of topology information advertisements, received with BGP listener **124**. In some instances, map modules **104** dynamically generate network map **120** and cost map **122** using routing information and reachability information received with BGP listener **124** in BGP advertisements that include traffic engineering data. In addition, in the illustrated embodiment, ALTO server **100** comprises topology information base **116**, a data structure that includes information regarding network topology, transmission costs for various network links, and other information that may be used by map modules **104** to generate ALTO-based maps (i.e., network map **120** and cost map **122**). Topology information base **116** may comprise one or more routing information bases (RIBs). Topology information base **116** may further comprise a traffic engineering database. Network map module **106** of map modules **104** uses topology information base **116** and community attribute map **118** (illustrated as "community attr. map **118**") to generate network map **120** according to the techniques described herein. Cost map module **112** of map modules **104** may use dynamically generated network map **120** and topology information base **116** to generate cost map **122** according to the techniques described herein. In some instances, network map module **106** additionally generates a master network map for a network using two or more inter-AS network maps, and cost map module **112** generates a master cost map for a network using two or more inter-AS cost maps.

[0131] Network map **120** may comprise an intra-AS and/or inter-AS network map dynamically generated in accordance with the described techniques. Network map **120** includes one or more PID entries that comprise a data structure to store a corresponding one or more PIDs of the network map. Each PID entry may include one or more aggregated prefixes, and PID identifier or name, and any attributes for the PID. Cost map **122** may comprise an intra-AS and/or inter-AS cost map dynamically generated according to the described techniques. Cost map **122** may comprise, for example, one or more cost map entries that each specifies two PIDs of network map **120** and an ALTO cost for the two PIDs.

[0132] User interface **128** of ALTO server **100** may comprise a command-line interface (CLI), a graphical user interface (GUI), a remote procedure call (RPC), or some other method to enable administrator **138** to configure topology information base **116** and provisioning policies **126** of ALTO server **100**. Administrator **138** may be a network operator of, e.g., a service provider network, or a software agent executing, e.g., on a network management device. Administrator **138** provisions ALTO server **100** with provisioning policies **126**, a set of policies that cause network map module **106** to generate network map **120** and cost map module **112** to generate cost map **122** in accordance with administrator preferences relating to transmission costs, load balancing, service-discrimination, PID grouping, community attribute to PID identifier mappings, or other preference areas. For example, a policy in provisioning policies **126** may direct cost map module **112** to compute ALTO costs from IGP metrics, stored in topology information base **116**, according to a particular formula. As another example, a policy in provisioning policies **126** may direct network map module **106** to aggregate prefixes tagged with a particular community attribute into a particular PID.

[0133] BGP listener **124** is a reachability protocol listener that executes BGP to peer with BGP speakers to receive BGP UPDATE messages **134** that include NLRI for installation to topology information base **116**. BGP listener **124** may store a Loc-RIB (Local Routing Information Base) and/or an Adj-RIB-In (Adjacent Routing Information Base, Incoming) for each BGP peer of ALTO server **100** (not shown). In addition, administrator **138** may add, modify, or remove routing information from topology information base **116** using UI **128**. In some instances, BGP listener **124** receives traffic engineering data encoded within BGP advertisements received from the BGP speakers. BGP listener **124** stores the traffic engineering data to a traffic engineering database (TED) of topology information base **116**, which map modules **104** use to dynamically generate network map **120** and cost map **122**.

[0134] IGP listener **130** is a routing protocol listener that executes an interior gateway protocol to receive routing information **136** from network routers for installation to topology information base **116**. In some embodiments, IGP listener **130** executes OSPF to peer with neighboring routers to receive the LSDB and link-state updates for the autonomous system that includes ALTO server **100**. In such embodiments, topology information base **116** comprises the LSDB.

[0135] Community attribute map **118** (illustrated as "community attr. map **118**") is an associative data structure, such as a table, list, or map, that maps community attribute values to corresponding PID attribute values. Community attributes and PID attributes may be derived from different namespaces. For example, a community attribute is typically a four byte integer, while a PID attribute may be a string. Community attribute map **118** operates as a lookup data structure that is keyed to community attribute values. That is, for a community attribute value (e.g., a value that specifies endpoints of type "CDN node"), community attribute map **118** provides a corresponding PID attribute value (e.g., a value that identifies PIDs associated with endpoints of type "CDN node").

[0136] PID generator **110** aggregates endpoints described in topology information base **116** into PIDs and performs other PID generation, modification, and "splitting" techniques described in this disclosure. PID generator **110** dynamically aggregates destination prefix received in BGP UPDATE messages **134** into a new PID of network map **120** or modifies an existing PID of network map **120** to incorporate the destination prefix. When individual ones of BGP UPDATE messages **134** include community attributes, attribute module **108** assigns the PID attribute value, mapped by BGP listener **124** using the community attribute value received in BGP UPDATE message **134**, to the new or modified PID that includes the destination prefix.

[0137] Network map module **106** assembles the aggregated PIDs generated by PID generator **110** into network map **120** according to the techniques described herein. Cost map module **112** applies provisioning policies **126** to net-

work map **120** and, in some instances, topology information base **116** to generate a corresponding cost map **122**. PID attributes for PIDs of network map **120** dynamically determined from BGP UPDATE messages **134** may affect determination by cost map module **122** of inter-PID costs. For example, provisioning policies **126** may include a policy requiring cost map module **122** to set to infinity inter-PID costs for PID pairs having PIDs that both include a PID attribute value of "host." In this example, an application using the ALTO service provided by ALTO server **100** that receives a content request from a host endpoint will therefore not select another host endpoint to serve the requested content. The application, in the form of a request router for example, may instead select a CDN node.

[0138] Client interface **114** exposes an ALTO server interface to enable ALTO clients to request and receive network and cost maps for an application for the network. Client interface **114** sends a copy of network map **120** and cost map **122** to a requesting client in maps upload message **132**. Any update message **132** may comprise an incremental or a complete update, as described in Raghunath et al. incorporated above.

[0139] Client interface **114** may execute one or more protocols to obtain network addresses of ALTO clients in the network, and the client interface maintains these network addresses so as to push incremental updates of the maps to the clients. Example interfaces for client interface **114** may include Simple Object Access Protocol (SOAP) or other eXtensible Markup Language (XML) interfaces, a CLI or GUI, Simple Network Management Protocol (SNMP), Remote Procedure Calls (RPCs), and a Common Object Request Broker Architecture (CORBA) interface, for example.

[0140] In some embodiments of ALTO server **100**, client interface **114** implements an endpoint cost service. When client interface **114** receives, from a client, a list of endpoints represented in network map **120**, client interface **114** returns an ordinally ranked list of the endpoints or the costs, specified by cost map **122**, between the endpoints and the client or between the endpoints and another specified source node. Alternatively, client interface **114** returns costs between each of the endpoints and the client or between each of the endpoint and another specified source node.

[0141] In some instances, ALTO server **100** may operate as a master ALTO server and generate master network and cost maps for a network. In these instances, client interface **114** operates to communicate between ALTO server **100** and various other ALTO servers located within other autonomous systems of the network. Client interface **114** receives network and cost maps from these various other ALTO servers. Upon consolidation of these network and cost maps by map modules **104** into a master network and cost map, client interface **114** sends the master network and cost maps to the various other ALTO servers to provide a high-resolution network topology and cost information to other areas of the network.

[0142] FIG. **8** is a flowchart illustrating an example mode of operation of ALTO server **100** for dynamically generating or modifying an inter-AS network map for an autonomous system served by ALTO server **100** according to the described techniques. BGP listener **124** operates as a passive BGP listener to peer with BGP speakers of the autonomous system and receive BGP messages (**200**). BGP in this instance may refer to Interior BGP. When BGP listener **124** receives a BGP UPDATE message, BGP listener **124** installs the received route to topology information base **116** (**202**). In some instances, ALTO server **100** runs a BGP decision process over topology information base **116** to select routes for processing.

[0143] Network map module **106** of ALTO server **100** analyzes the BGP UPDATE message attributes to determine whether the route originated in the autonomous system (AS) of ALTO server **100**, that is, if the length of the AS_PATH attribute is equal to zero (**203**). If the route originated in a different AS (NO branch of **203**), PID generator **110** adds the advertised prefix to a PID of network map **120** (in this example, an inter-AS network map) that has both the same AS_PATH attribute and the same NEXT_HOP attribute as that of the route (**214**). If such a PID does not yet exist in network map **120**, PID generator **110** creates a new PID, sets the PID NEXT_HOP attribute to the NEXT_HOP value of the route, sets the PID AS_PATH attribute to the AS_PATH value of the route, and adds the prefix to the new PID in network map **120**. If some instances, network map module **106** generates an intra-AS network map that only includes prefixes served by the autonomous system of ALTO server **100**. In such instances, PID generator **110** ignores all advertised prefixes for which the AS_PATH length exceeds 0 and may therefore skip step **214**.

[0144] If the route originated in the AS of ALTO server **100** (YES branch of **203**), attribute module **108** determines whether the route includes a BGP community attribute or extended community attribute that is mapped, in provisioning policies **126**, to a particular PID (**204**). If so (YES branch of **204**), then attribute module **108** directs PID generator **110** to add the advertised prefix to the mapped PID (**206**). If necessary, PID generator **110** creates the mapped PID in network map **120**.

[0145] Attribute module **108** determines whether any of the one or more community attribute values embedded in the BGP UPDATE message are mapped within community attribute map **118** to PID attribute values (**208**). If so (YES branch of **208**), attribute module **108** directs PID generator **110** to add the advertised prefix to a PID in network map **120** that (1) has the same set of mapped PID attributes to which the community attribute values of the prefix are mapped in community attribute map **118**, and (2) also has the same NEXT_HOP attribute as the advertised prefix (**210**). In other words, PID generator **110** aggregates prefixes tagged with the same set of mapped community attribute values and the same NEXT_HOP into a single PID. If necessary, PID generator **110** creates such a PID in network map **120**.

[0146] If the advertised prefix is not mapped in either community attribute map **118** or provisioning policies **126**, PID generator **110** adds the prefix to a PID in network map **120** that has no mapped PID attribute value and the yet has the same NEXT_HOP attribute as the advertised prefix (**212**). If necessary, PID generator **110** creates such a PID in network map **120**. After dynamically generating or modifying network map **120**, client interface **114** provides the updated network map to ALTO clients of ALTO server **100** in update message **132** (**216**).

[0147] FIGS. **9A-9D** present a flowchart that illustrates an example operation of ALTO server **100** to dynamically generate or modify an inter-AS cost map for an autonomous system served by ALTO server **100** according to the techniques described herein. IGP listener **130** executes a routing protocol to receive routing information in routing protocol

messages from internal routers of the autonomous system that includes ALTO server **100** (**230**). When IGP listener **130** receives routing protocol information (**232**), such as link-state advertisements, IPG listener **130** installs the routing information to a link-state database of topology information base **116** (**234**). In accordance with the routing protocol, IGP listener **130** executes a shortest-path first algorithm over the LSDB to compute IGP metric between routers of the autonomous system that includes ALTO server **100** (**236**). IGP listener **130** may store the IGP metrics to topology information base **116**.

[0148] Cost map module **112** of ALTO server **100** computes and sets ALTO costs for each PID pair generated by network map module **106** in cost map **122**. In this instance, the cost map is an inter-AS cost map. In some instances, however, cost map module **112** generates an intra-AS in addition to, or instead of, the inter-AS cost map as cost map **122**.

[0149] Cost map module **112** computes ALTO costs differently for different combinations of PID pair member PID types. For each PID pair in network map **120** (**238**), cost map module **112** determines whether both member PIDs of the PID pair include prefixes that are local to the autonomous system that includes ALTO server **100** (i.e., whether the member PIDs are local PIDs) (**240**). If so (YES branch of **240**), cost map module **112** determines the IGP metric, from topology information base **116**, between the respective next hop routers identified in the NEXT_HOP attribute of the member PIDs of the PID pair (**250**). Cost map module **112** applies one or more policies of provisioning policies **126** to the determined IGP metric to compute an ALTO cost for the PID pair (**252**), then sets this ALTO cost for the PID pair in cost map **122** (**254**).

[0150] If one member PID of the PID pair is local and the other member PID is non-local (i.e., remote) (NO branch of **240**), then cost map **112** determines whether the remote member PID is located in a neighboring autonomous system, as identified for example by the remote member PID AS identifier attribute (**242**). If located in a neighboring AS (YES branch of **242**), client interface **114** receives a partial cost map from an ALTO server of the neighboring autonomous system that serves the prefixes of the remote PID (**260**). The partial cost map includes ALTO Network map module **106** runs a BGP decision process with topology information of topology information base **116** that accords with the perspective of the local PID member of the PID pair in order to identify the likely traffic path from the local PID to the remote PID (**262**).

[0151] The ALTO cost for the PID pair is the sum of the intra-AS and inter-AS ALTO costs for the identified traffic path. Cost map module **112** queries topology information base **116** to obtain the IGP metric between the next hop router identified in the NEXT_HOP attribute of the local member PID and the next hop router identified in the NEXT_HOP attribute of the remote member PID (**264**). Cost map module **112** applies one or more policies of provisioning policies **126** to the determined IGP metric to compute an intra-AS ALTO cost (**266**). Cost map module **112** an inter-AS ALTO cost between the border routers that connect the AS and the neighboring AS along the identified traffic path (**268**). This inter-AS ALTO may be configured in topology information base **116**. Cost map module **112** queries the partial cost map received by client interface **114** to obtain a remote ALTO cost between the border router of the neighboring AS that lies along the identified traffic path and prefixes of the remote PID (**270**). Cost map module **112** computes the sum of the intra-AS cost, the remote cost, and the inter-AS cost (**272**) and sets the sum as the ALTO cost for the PID pair in cost map **122** (**274**).

[0152] If the PID pair includes both a local PID and a non-neighbor remote PID (YES branch of **244**), the ALTO cost for the PID pair is set as the sum of the intra-AS cost and the inter-AS cost from the autonomous system that includes ALTO server **100** to the remote AS that serves the prefixes of the remote PID. Cost map module **112** queries topology information base **116** to obtain the IGP metric between the next hop router identified in the NEXT_HOP attribute of the local member PID and the next hop router identified in the NEXT_HOP attribute of the remote member PID (**280**). Cost map module **112** applies one or more policies of provisioning policies **126** to the determined IGP metric to compute an intra-AS ALTO cost (**282**).

[0153] In this example, a default inter-AS ALTO cost is configured in topology information base **116**. Cost map module **112** next determines the AS_PATH length from the AS_PATH attribute of the remote PID and then computes the inter-AS ALTO cost as the product of the AS_PATH length and the default inter-AS ALTO cost (**284**). Cost map module **112** sums the inter-AS ALTO cost with the intra-AS ALTO cost (**286**) and sets the sum as the ALTO cost for the PID pair in cost map **122** (**288**). In some instances, administrator **138** configures topology information base **116** with inter-AS costs between autonomous system identified by AS ID. In such instances, cost map module **116** may query topology information base **116** to obtain a more accurate inter-AS costs to compute a total inter-AS cost from the local PID to the remote PID.

[0154] If both members of the PID pair are remote PIDs (NO branch of **244**), the cost map module **112** sets the ALTO cost for the PID pair as proportional to the number of inter-AS links between the remote PIDs (**246**). In one example, cost map module **112** may determine the number of inter-AS links by analyzing the AS_PATH attributes of the remote PIDs to find the AS_PATH value that includes each of the respective AS IDs of the respective ASes that serve the remote PIDs. Cost map module **112** determines the number of inter-AS links from the determined value and calculates the ALTO cost as the product of the number of inter-AS links and the default inter-AS ALTO cost. If neither PID has such an AS_PATH value, cost map module **112** may be unable to determine the inter-AS ALTO cost and sets the ALTO cost for the PID pair to infinity. After dynamically generating or modifying cost map **122**, client interface **114** provides the updated cost to ALTO clients of ALTO server **100** in update message **132** (**248**).

[0155] FIG. **10** is a flowchart illustrating an example operation of ALTO server **100** to generate a master ALTO network map according to the techniques set forth herein. In this example operation, client interface **114** receives first and second inter-AS network maps from respective ALTO servers operating within first and second autonomous systems (**300**). In some examples, ALTO server **100** may comprise one of these respective ALTO servers and in such examples ALTO server **100** uses inter-AS network and cost map that it has generated to generate master network and cost maps.

[0156] The inter-AS network maps each include local PIDs that network map module **106** "carries over" to network map **120** (in this operation, network map **120** is a

master network map for the network) (**302**). That is, network map module **106** copies the prefixes, attributes, and other values of the local PIDs into corresponding PIDs for network map **120**.

[0157] For each of the first and second inter-AS network maps (**304**), network map module **106** compares the remote PIDs of the inter-AS network map to local PIDs carried over to the master network map **120** from the other inter-AS network map (**306**). In particular, network map module **106** splits PIDs of the master network map **120** when such PIDs, corresponding to local PIDs of the other inter-AS network map, have prefixes that are not advertised to the AS of the inter-AS network map, as evidenced by prefixes that are not included in the remote PIDs (**308**).

[0158] In addition, network map module **106** splits PIDs of the master network map **120** when such PIDs, corresponding to local PIDs of the other inter-AS network map, have divergent AS_PATH lengths according to the remote PIDs (**310**). That is, when the remote PIDs separate prefixes to a single local PID of the other inter-AS network map according to different AS_PATH lengths to the prefixes, network map module **106** splits the corresponding PID for the local PID in master network map **120** according to the separation defined by the remote PIDs. Network map module **106** additionally splits PIDs of the master network map **120** when such PIDs, corresponding to local PIDs of the other inter-AS network map, have divergent NEXT_HOP attribute values according to the remote PIDs (**312**). That is, when the remote PIDs separate prefixes to a single local PID of the other inter-AS network map according to different NEXT_HOP values for the prefixes, network map module **106** splits the corresponding PID for the local PID in master network map **120** according to the separation defined by the remote PIDs. After dynamically generating or modifying master network map **120**, client interface **114** provides the updated network map in update message **132** to ALTO clients of ALTO server **100** or to other ALTO servers in various ASes of the network (**314**).

[0159] FIG. **11** is a flowchart illustrating an example operation of ALTO server **100** to generate master ALTO network and cost maps according to the techniques set forth herein. In this example operation, client interface **114** receives first and second inter-AS network and cost maps from respective ALTO servers operating within first and second autonomous systems (**326**). In some examples, ALTO server **100** may comprise one of these respective ALTO servers and in such examples ALTO server **100** uses inter-AS network and cost map that it has generated to generate master network and cost maps. Network map module **106** generates a master network map **120** using the first and second inter-AS network maps (**328**). While illustrated and described sequentially, the operations of generating a master network map **120** and master cost map **122** may occur concurrently.

[0160] Cost map module **112** then assembles corresponding master cost map **122** for master network map **120**. In this example, for each PID pair in master network map **120** (**330**), cost map module **112** determines whether both member PIDs are carried over from the same inter-AS network map, i.e., whether the PIDs are served by the same autonomous system (**332**). If so (YES branch of **332**), cost map module **112** sets the ALTO cost for the PID pair to the ALTO cost for the corresponding PIDs as specified in the inter-AS cost map for the autonomous system that serves the PIDs (**334**). If the PID members represent a "split" PID that network map module **106** created to account for divergence in the first and second inter-AS network maps, cost map module **112** sets the ALTO cost for the PID pair to zero to reflect that the prefixes of these PID are aggregated into the same PID by the ALTO server of the autonomous system that serves these PIDs (**338**).

[0161] If the PID members include prefixes from both the first and second inter-AS network maps (NO branch of **336**), cost map module **112** sets the unidirectional ALTO costs between the PID members using costs from both the first and second inter-AS cost maps. Specifically, cost map module **112** sets the ALTO cost from the first PID member to the second PID member in master cost map **122** as the cost specified by the inter-AS cost map provided by the autonomous system that serves the first PID (**340**). Cost map module **112** sets the ALTO cost from the second PID member to the first PID member in master cost map **122** as the cost specified by the inter-AS cost map provided by the autonomous system that serves the second PID (**342**). After generating the master network and cost maps, client interface **114** provides the updated maps in update message **132** to ALTO clients of ALTO server **100** or to other ALTO servers in various ASes of the network (**344**).

[0162] The techniques described in this disclosure may be implemented, at least in part, in hardware, software, firmware or any combination thereof. For example, various aspects of the described techniques may be implemented within one or more processors, including one or more microprocessors, digital signal processors (DSPs), application specific integrated circuits (ASICs), field programmable gate arrays (FPGAs), or any other equivalent integrated or discrete logic circuitry, as well as any combinations of such components. The term "processor" or "processing circuitry" may generally refer to any of the foregoing logic circuitry, alone or in combination with other logic circuitry, or any other equivalent circuitry. A control unit comprising hardware may also perform one or more of the techniques of this disclosure.

[0163] Such hardware, software, and firmware may be implemented within the same device or within separate devices to support the various operations and functions described in this disclosure. In addition, any of the described units, modules or components may be implemented together or separately as discrete but interoperable logic devices. Depiction of different features as modules or units is intended to highlight different functional aspects and does not necessarily imply that such modules or units must be realized by separate hardware or software components. Rather, functionality associated with one or more modules or units may be performed by separate hardware or software components, or integrated within common or separate hardware or software components.

[0164] The techniques described in this disclosure may also be embodied or encoded in a computer-readable medium, such as a non-transitory computer-readable medium or computer-readable storage medium, containing instructions. Instructions embedded or encoded in a computer-readable medium may cause a programmable processor, or other processor, to perform the method, e.g., when the instructions are executed. Computer readable storage media may include random access memory (RAM), read only memory (ROM), programmable read only memory (PROM), erasable programmable read only memory

(EPROM), electronically erasable programmable read only memory (EEPROM), flash memory, a hard disk, a CD-ROM, a floppy disk, a cassette, magnetic media, optical media, or other computer-readable storage media. It should be understood that the term "computer-readable storage media" refers to physical storage media, and not signals or carrier waves, although the term "computer-readable media" may include transient media such as signals, in addition to physical storage media.

[0165] Various embodiments of the invention have been described. These and other embodiments are within the scope of the following claims.

1-20: (canceled)

21: A method comprising:

storing, by an application-layer traffic optimization (ALTO) server, layer 3 topology information that specifies one or more endpoints, a next hop attribute that indicates a next hop for the one or more endpoints, an autonomous system path attribute that indicates an autonomous system path for the one or more endpoints, and a communities path attribute that indicates the one or more endpoints share a common property;

generating, by the ALTO server based at least on the community value for the one or more endpoints, a provider-defined identifier (PID) that includes the one or more endpoints; and

providing, by the ALTO server, an ALTO service to a client device in accordance with an ALTO network map that includes the PID.

22: The method of claim **21**, further comprising:

receiving, by the ALTO server, a layer 3 topology information advertisement that includes the layer 3 topology information.

23: The method of claim **21**, further comprising:

receiving, by the ALTO server, a Border Gateway Protocol (BGP) UPDATE message that includes the layer 3 topology information,

wherein Network Layer Reachability Information of the BGP UPDATE message specifies the one or more endpoints,

wherein a NEXT_HOP path attribute of the BGP UPDATE message indicates the next hop for the one or more endpoints,

wherein an AS_PATH path attribute of the BGP UPDATE message indicates the autonomous system path for the one or more endpoints, and

wherein a BGP COMMUNITIES path attribute indicates the community value for the one or more endpoints.

24: The method of claim **21**, wherein the communities path attribute comprises one of a Border Gateway Protocol (BGP) COMMUNITIES path attribute, a BGP EXTENDED COMMUNITY path attribute, and a combination of a BGP COMMUNITIES path attribute and a BGP EXTENDED COMMUNITY path attribute.

25: The method of claim **21**, further comprising:

applying, by the ALTO server, a community attribute map to the attribute value to map the attribute value for the one or more endpoints to a PID identifier that identifies the PID,

wherein generating the PID comprises generating, by the ALTO server, the PID based at least on the PID identifier.

26: The method of claim **21**, wherein the client device comprises one of a peer-to-peer device, a content request router, and another ALTO server.

27: The method of claim **21**, wherein providing the ALTO service comprises one of sending the ALTO network map to the client device and selecting, based at least on the ALTO network map, one of the one or more endpoints of the PID for a content request received by the ALTO server.

28: The method of claim **21**, wherein the common property shared by the one or more endpoints affects an application layer cost for application layer traffic exchanged between the one or more endpoints of the PID and one or more endpoints of at least one other PID of the ALTO network map.

29: The method of claim **21**, further comprising:

computing, by the ALTO server based at least on the attribute value, an application layer cost between the PID and at least one other PID of a plurality of PIDs of the ALTO network map,

wherein providing the ALTO service comprises providing the ALTO service based at least on the application layer cost between the PID and the at least one other PID.

30: An application-layer traffic optimization (ALTO) server comprising:

at least one processor operably coupled to a memory;

a database for a routing protocol, the database configured to store layer 3 topology information that specifies one or more endpoints, a next hop attribute that indicates a next hop for the one or more endpoints, an autonomous system path attribute that indicates an autonomous system path for the one or more endpoints, and a communities path attribute that indicates the one or more endpoints share a common property;

a network map module configured for execution by the at least one processor to generate, based at least on the community value for the one or more endpoints, a provider-defined identifier (PID) that includes the one or more endpoints;

a client interface configured for execution by the at least one processor to provide an ALTO service to a client device in accordance with an ALTO network map that includes the PID.

31: The ALTO server of claim **30**, further comprising:

a protocol listener configured for execution by the at least one processor to receive a layer 3 topology information advertisement that includes the layer 3 topology information.

32: The ALTO server of claim **30**, further comprising:

a protocol listener configured for execution by the at least one processor to receive a Border Gateway Protocol (BGP) UPDATE message that includes the layer 3 topology information,

wherein Network Layer Reachability Information of the BGP UPDATE message specifies the one or more endpoints,

wherein a NEXT_HOP path attribute of the BGP UPDATE message indicates the next hop for the one or more endpoints,

wherein an AS_PATH path attribute of the BGP UPDATE message indicates the autonomous system path for the one or more endpoints, and

wherein a BGP COMMUNITIES path attribute indicates the community value for the one or more endpoints.

33: The ALTO server of claim **30**, wherein the communities path attribute comprises one of a Border Gateway Protocol (BGP) COMMUNITIES path attribute, a BGP EXTENDED COMMUNITY path attribute, and a combination of a BGP COMMUNITIES path attribute and a BGP EXTENDED COMMUNITY path attribute.

34: The ALTO server of claim **30**,

wherein the network map module is further configured to apply a community attribute map to the attribute value to map the attribute value for the one or more endpoints to a PID identifier that identifies the PID, and

wherein generating the PID comprises generating, by the ALTO server, the PID based at least on the PID identifier.

35: The ALTO server of claim **30**, wherein the client device comprises one of a peer-to-peer device, a content request router, and another ALTO server.

36: The ALTO server of claim **30**, wherein the client interface is further configured to provide the ALTO service by one of sending the ALTO network map to the client device and selecting, based at least on the ALTO network map, one of the one or more endpoints of the PID for a content request received by the ALTO server.

37: The ALTO server of claim **30**, wherein the common property shared by the one or more endpoints affects an application layer cost for application layer traffic exchanged between the one or more endpoints of the PID and one or more endpoints of at least one other PID of the ALTO network map.

38: The ALTO server of claim **30**, further comprising:

an ALTO cost map module configured for execution by the at least one processor to compute, based at least on the attribute value, an application layer cost between the PID and at least one other PID of a plurality of PIDs of the ALTO network map,

wherein the client interface is further configured to provide the ALTO service based at least on the application layer cost between the PID and the at least one other PID.

39: A non-transitory computer-readable medium comprising instructions that, when executed, cause one or more processors of an application-layer traffic optimization (ALTO) server to:

store layer 3 topology information that specifies one or more endpoints, a next hop attribute that indicates a next hop for the one or more endpoints, an autonomous system path attribute that indicates an autonomous system path for the one or more endpoints, and a communities path attribute that indicates the one or more endpoints share a common property;

generate, based at least on the community value for the one or more endpoints, a provider-defined identifier (PID) that includes the one or more endpoints; and

provide an ALTO service to a client device in accordance with an ALTO network map that includes the PID.

* * * * *