US009270426B1

(12) **United States Patent**
Atlas et al.

(10) **Patent No.: US 9,270,426 B1**
(45) **Date of Patent: *Feb. 23, 2016**

(54) **CONSTRAINED MAXIMALLY REDUNDANT TREES FOR POINT-TO-MULTIPOINT LSPS**

(75) Inventors: **Alia Atlas**, Arlington, MA (US); **Maciek Konstantynowicz**, Haddenham (GB)

(73) Assignee: **Juniper Networks, Inc.**, Sunnyvale, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 425 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **13/610,520**

(22) Filed: **Sep. 11, 2012**

(51) **Int. Cl.**

H04L 12/28	(2006.01)
H04L 5/00	(2006.01)
H04W 72/04	(2009.01)

(52) **U.S. Cl.**
CPC ***H04L 5/001*** (2013.01); ***H04L 5/0044*** (2013.01); ***H04L 5/0053*** (2013.01); ***H04W 72/0466*** (2013.01)

(58) **Field of Classification Search**
CPC H04L 12/28; H04L 12/56; H04J 1/16; G06F 11/07
USPC 370/217–225, 238, 242, 255, 390
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,630,298	B2	12/2009	Shand et al.	
8,274,989	B1 *	9/2012	Allan	H04L 45/16 370/432
8,351,418	B2 *	1/2013	Zhao et al.	370/351
2003/0152024	A1 *	8/2003	Yang	H04J 3/14 370/216
2005/0135276	A1	6/2005	Bouchat et al.	
2006/0050690	A1	3/2006	Epps et al.	
2006/0087965	A1	4/2006	Shand et al.	
2008/0031130	A1 *	2/2008	Raj	H04L 45/00 370/225
2008/0123533	A1 *	5/2008	Vasseur	H04L 45/00 370/238
2008/0219272	A1 *	9/2008	Novello	H04L 47/11 370/401
2009/0010272	A1	1/2009	Wijnands et al.	
2009/0185478	A1 *	7/2009	Zhang	H04L 12/5695 370/216
2009/0219806	A1 *	9/2009	Chen	H04L 45/00 370/219
2010/0080120	A1 *	4/2010	Bejerano	H04L 12/18 370/228

(Continued)

FOREIGN PATENT DOCUMENTS

WO	2010055408	A1	5/2010

OTHER PUBLICATIONS

Office Action from U.S. Appl. No. 13/418,212, dated May 22, 2014, 14 pp.

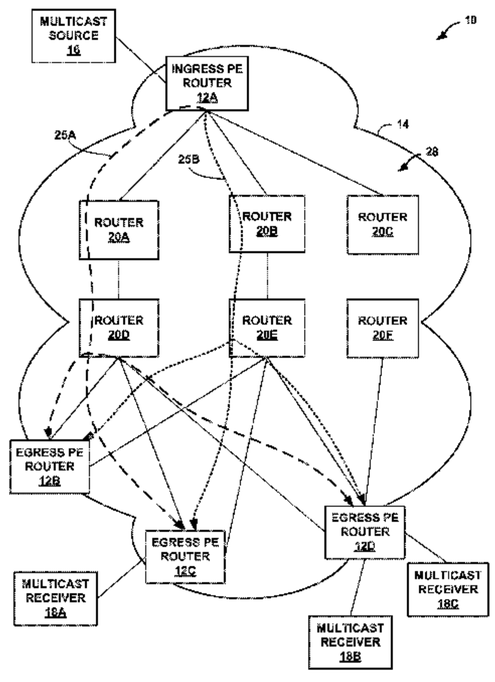
(Continued)

Primary Examiner — Khaled Kassim
Assistant Examiner — Berhanu Belete
(74) *Attorney, Agent, or Firm* — Shumaker & Sieffert, P.A.

(57) **ABSTRACT**

A network device computes a set of maximally redundant trees from an ingress network device to a plurality of egress network devices based on a network graph, in which each of the set of maximally redundant trees comprises a spanning tree to the plurality of egress network devices rooted at the network device. The network device computes the maximally redundant trees such that each of the links along each of the set of maximally redundant trees satisfies a specified traffic-engineering constraint. The ingress network device establishes a plurality of point to multipoint (P2MP) label switched paths (LSPs) from the network device as an ingress network device to the plurality of egress network devices along the set of maximally redundant trees, wherein each of the P2MP LSPs corresponds to a different one of the maximally redundant trees.

21 Claims, 6 Drawing Sheets



(56) **References Cited**

U.S. PATENT DOCUMENTS

2011/0051727	A1	3/2011	Cai et al.	
2011/0069609	A1*	3/2011	Le Roux	H04L 12/1868 370/221
2011/0158128	A1*	6/2011	Bejerano	H04L 12/1877 370/256
2011/0199891	A1*	8/2011	Chen	H04L 45/22 370/218
2011/0211445	A1*	9/2011	Chen	370/221
2011/0286336	A1*	11/2011	Vasseur	H04L 45/00 370/238
2012/0039164	A1*	2/2012	Enyedi	H04L 45/00 370/217
2012/0281524	A1*	11/2012	Farkas	H04L 45/66 370/221
2013/0077475	A1*	3/2013	Enyedi	H04L 45/128 370/225
2013/0121169	A1*	5/2013	Zhao	H04L 43/00 370/242
2013/0232259	A1*	9/2013	Csaszar	H04L 43/0817 709/224
2013/0322231	A1	12/2013	Császár et al.	
2014/0016457	A1*	1/2014	Enyedi et al.	370/225

OTHER PUBLICATIONS

Enyedi et al., "On Finding Maximally Redundant Trees in Strictly Linear Time," IEEE Symposium on Computers and Communications (ISCC), 2009, 11 pp.
Atlas et al., "An Architecture for IP/LDP Fast-Reroute Using Maximally Redundant Trees," Routing Area Working Group Internet Draft, draft-atlas-rtgwg-mrt-frr-architecture-01, Oct. 31, 2011, 27 pp.
Karan et al., "Multicast only Fast Re-Route," Network Working Group Internet Draft, draft-karan-mofrr-00, Mar. 2, 2009, 14 pp.
Wei et al., "Tunnel Based Multicast Fast Reroute (TMFRR) Extensions to PIM," Network Working Group Internet-Draft, draft-lwei-pim-tmfrr-00.txt, Oct. 16, 2009, 20 pp.
Lindem et al., "Extensions to IS-IS and OSPF for Advertising Optional Router Capabilities," Network Working Group Internet Draft, draft-raggarwa-igp-cap-01.txt, Nov. 2002, 12 pp.
Shen et al., "Discovering LDP Next-Nexthop Labels," Network Working Group Internet Draft, draft-shen-mpls-ldp-nnhop-label-02.txt, May 2005, 9 pp.
Enyedi et al., "IP Fast ReRoute: Lightweight Not-Via without Additional Addresses," Proceedings of IEEE INFOCOM, 2009, 5 pp.
Lindem et al., "Extensions to OSPF for Advertising Optional Router Capabilities," RFC 4970, Jul. 2007, 13 pp.
Boers et al., "The Protocol Independent Multicast (PIM) Join Attribute Format," RFC 5384, Nov. 2008, 11 pp.
Enyedi, "Novel Algorithms for IP Fast ReRoute," Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics Ph.D. Thesis, Feb. 2011, 114 pp.
Atlas et al., "An Architecture for Multicast Protection Using Maximally Redundant Trees," Routing Area Working Group Internet Draft, draft-atlas-rtgwg-mrt-mc-arch-00, Mar. 2, 2012, 26 pp.
Atlas et al., "Algorithms for computing Maximally Redundant Trees for IP/LDP Fast ReRoute," Routing Area Working Group Internet Draft, Nov. 28, 2011, 39 pp.
Minei et al., "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths," Network Working Group Internet Draft, draft-ietf-mpls-ldp-p2mp-15, Aug. 4, 2011, 40 pp.
Cai et al., "PIM Multi-Topology ID (MT-ID) Join Attribute," draft-ietf-pim-mtid-10.txt, Sep. 27, 2011, 14 pp.
Atlas et al., "Basic Specification for IP Fast ReRoute: Loop-Free Alternates," RFC 5286, Sep. 2008, 32 pp.
Shand et al., "IP Fast ReRoute Framework," RFC 5714, Jan. 2010, 16 pp.
Shand et al., "IP Fast ReRoute Using Not-via Addresses," Network Working Group Internet Draft, draft-ietf-rtgwg-ipfrr-notvia-addresses-07, Apr. 20, 2011, 29 pp.
Filsfils et al., "LFA applicability in SP networks," Network Working Group Internet Draft, draft-ietf-rtgwg-lfa-applicability-03, Aug. 17, 2011, 33 pp.
Francois et al., "Loop-free convergence using oFIB," Network Working Group Internet Draft, draft-ietf-rtgwg-ordered-fib-05, Apr. 20, 2011, 24 pp.
Rétvári et al., "IP Fast ReRoute: Loop Free Alternates Revisited," Proceedings of IEEE INFOCOM, 2011, 9 pp.
Retana et al., "OSPF Stub Router Advertisement," RFC 3137, Jun. 2001, 5 pp.
Bryant et al., "Remote LFA FRR," Network Working Group Internet Draft, draft-shand-remote-lfa-00, Oct. 11, 2011, 12 pp.
Kahn, "Topological sorting of large networks," Communications of the ACM, vol. 5, Issue 11, pp. 558-562, Nov. 1962.
Moy, "OSPF Version 2," RFC 2328, Apr. 1998, 204 pp.
U.S. Appl. No. 12/419,507, by Kireeti Kompella, filed Apr. 7, 2009.
Atlas et al., "An Architecture for IP/LDP Fast-Reroute Using Maximally Redundant Trees," PowerPoint presentation, IETF 81, Quebec City, Canada, Jul. 27, 2011, 28 pp.
Callon, "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments," RFC 1195, Dec. 1990, 80 pp.
Shand et al., "A Framework for Loop-Free Convergence," RFC 5715, Jan. 2010, 23 pp.
Vasseur et al., "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information," RFC 4971, Jul. 2007, 10 pp.
Atlas et al., Algorithms for computing Maximally Redundant Trees for IP/LDP Fast-Reroute, draft-enyedi-rtgwg-mrt-frr-algorithm-00, Oct. 24, 2011, 36 pp.
Atlas et al., An Architecture for IP/LDP Fast-Reroute Using Maximally Redundant Trees, draft-atlas-rtgwg-mrt-frr-architecture-00, Jul. 4, 2011, 22 pp.
Karan et al. "Multicast Only Fast Re-Route," Internet-Draft, draft-karan-mofrr-01, Mar. 13, 2011, 15 pp.
U.S. Appl. No. 13/418,152, by Alia Atlas, filed Mar. 12, 2012.
U.S. Appl. No. 13/418,180, by Alia Atlas, filed Mar. 12, 2012.
U.S. Appl. No. 13/418,212, by Alia Atlas, filed Mar. 12, 2012.
Aggarwal, "Extensions to Resource Reservation Protocol—Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)," Network Working Group, RFC 4875, 53 pp.
Awduche et al., "RSVP-TE: Extensions to RSVP for LSP Tunnels," Network Working Group, RFC 3209, 61 pp.
Andersson et al., "LDP Specification, Network Working Group," RFC 3036, 124 pp.
S. Giacalone et al., "OSPF Traffic Engineering (TE) Express Path," draft-giacalone-ospf-te-express-path-02.txt, Network Working Group, Internet draft, 15 pp.
U.S. Appl. No. 13/112,961, by David Ward, filed May 20, 2011.
Notice of Allowance from U.S. Appl. No. 13/418,212, mailed Dec. 9, 2014, 10 pp.
Office Action from U.S. Appl. No. 14/015,613, dated May 15, 2015, 20 pp.
Response filed Aug. 17, 2015 to the Office Action mailed May 15, 2015 in U.S. Appl. No. 14/015,613, 16 pgs.
Response to Office Action dated May 22, 2014, from U.S. Appl. No. 13/418,212, filed Aug. 21, 2014, 10 pp.
Final Office Action from U.S. Appl. No. 14/015,613, dated Nov. 30, 2015, 27 pp.

* cited by examiner

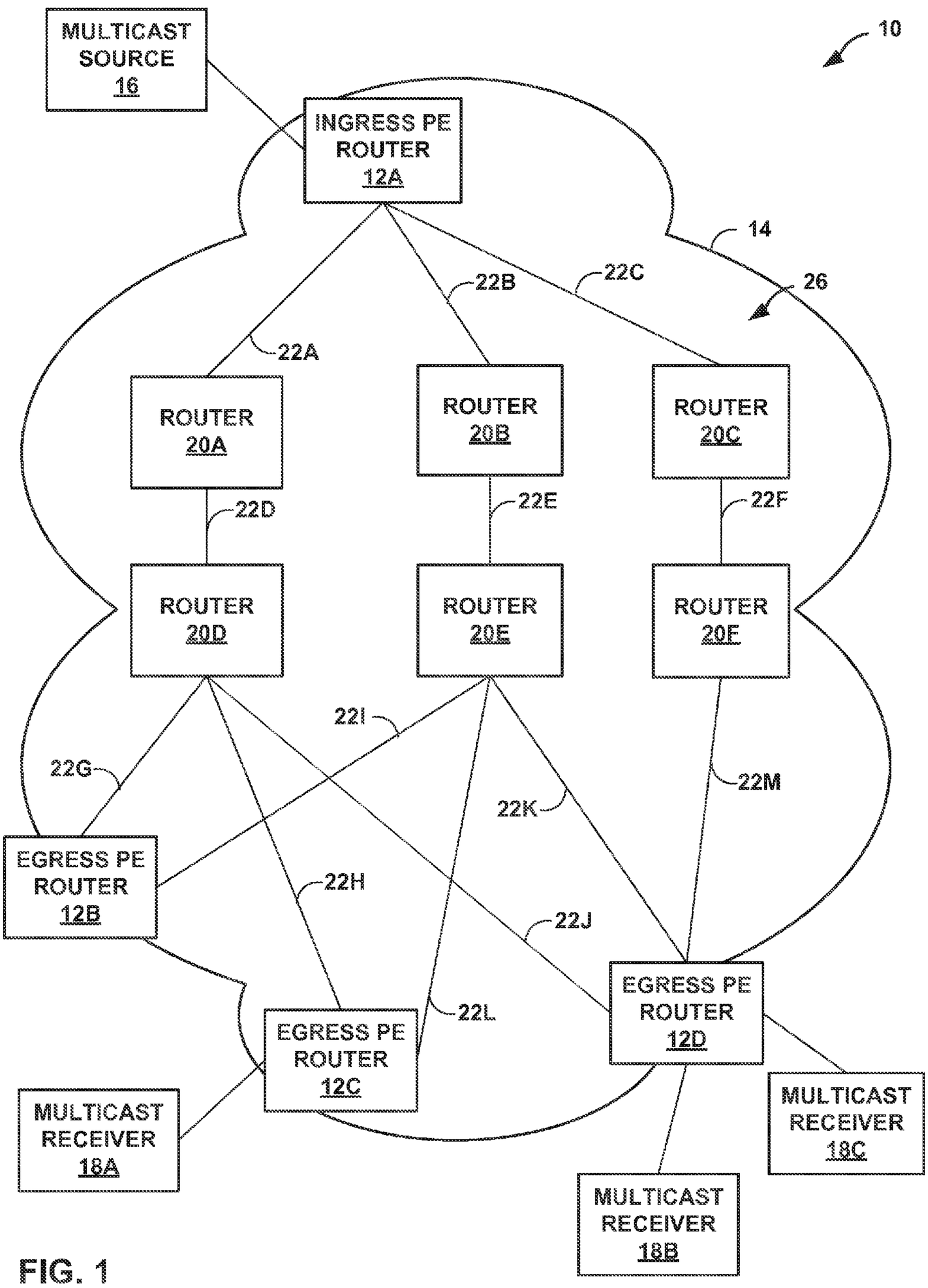

MULTICAST SOURCE 16
INGRESS PE ROUTER 12A
10
14
26
22A
22B
22C
ROUTER 20A
ROUTER 20B
ROUTER 20C
22D
22E
22F
ROUTER 20D
ROUTER 20E
ROUTER 20F
22G
22I
22K
22M
EGRESS PE ROUTER 12B
22H
22J
22L
EGRESS PE ROUTER 12C
EGRESS PE ROUTER 12D
MULTICAST RECEIVER 18A
MULTICAST RECEIVER 18C
MULTICAST RECEIVER 18B

FIG. 1

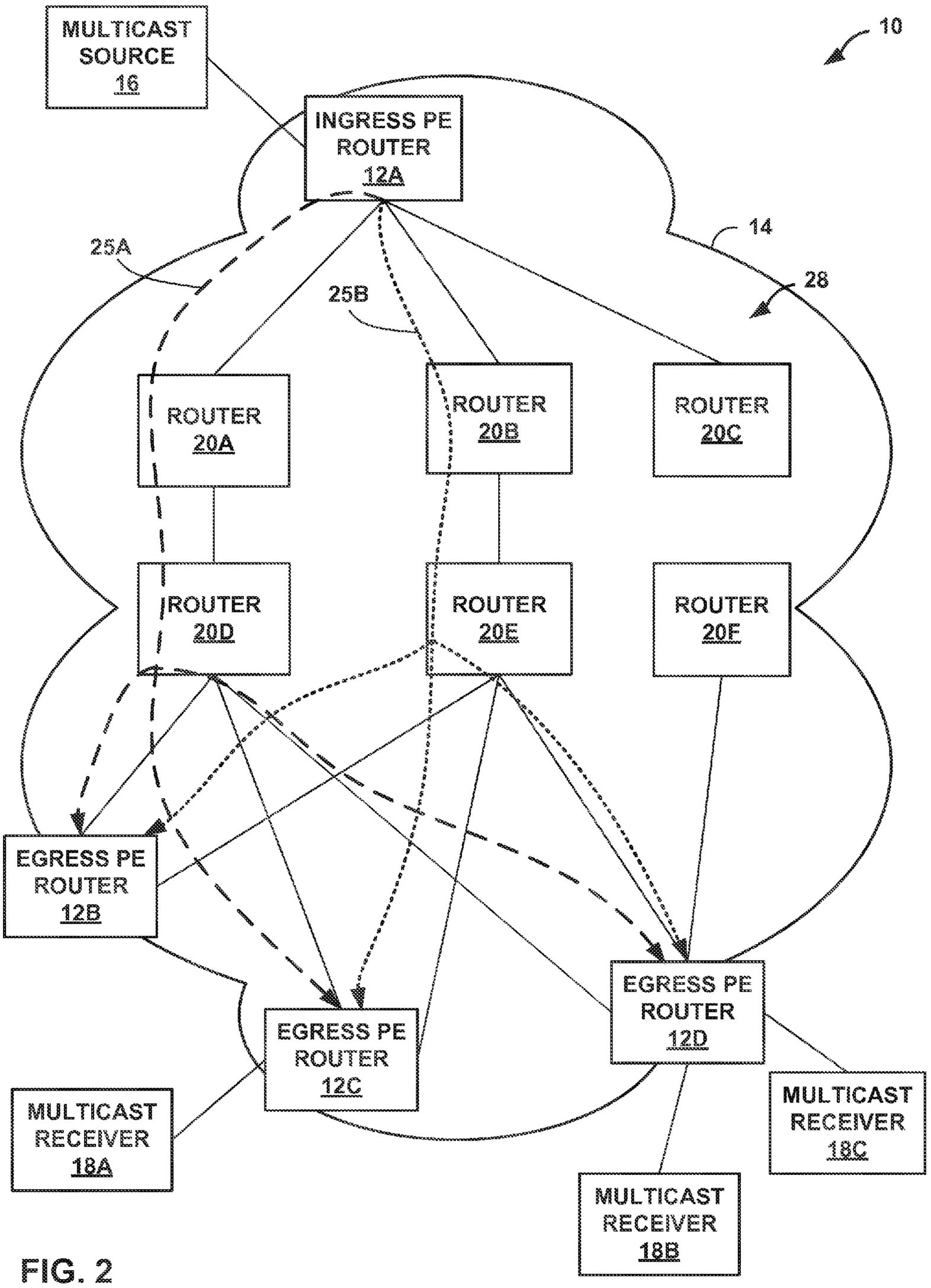

MULTICAST SOURCE 16
INGRESS PE ROUTER 12A
10
14
25A
25B
28
ROUTER 20A
ROUTER 20B
ROUTER 20C
ROUTER 20D
ROUTER 20E
ROUTER 20F
EGRESS PE ROUTER 12B
EGRESS PE ROUTER 12C
EGRESS PE ROUTER 12D
MULTICAST RECEIVER 18A
MULTICAST RECEIVER 18B
MULTICAST RECEIVER 18C

FIG. 2

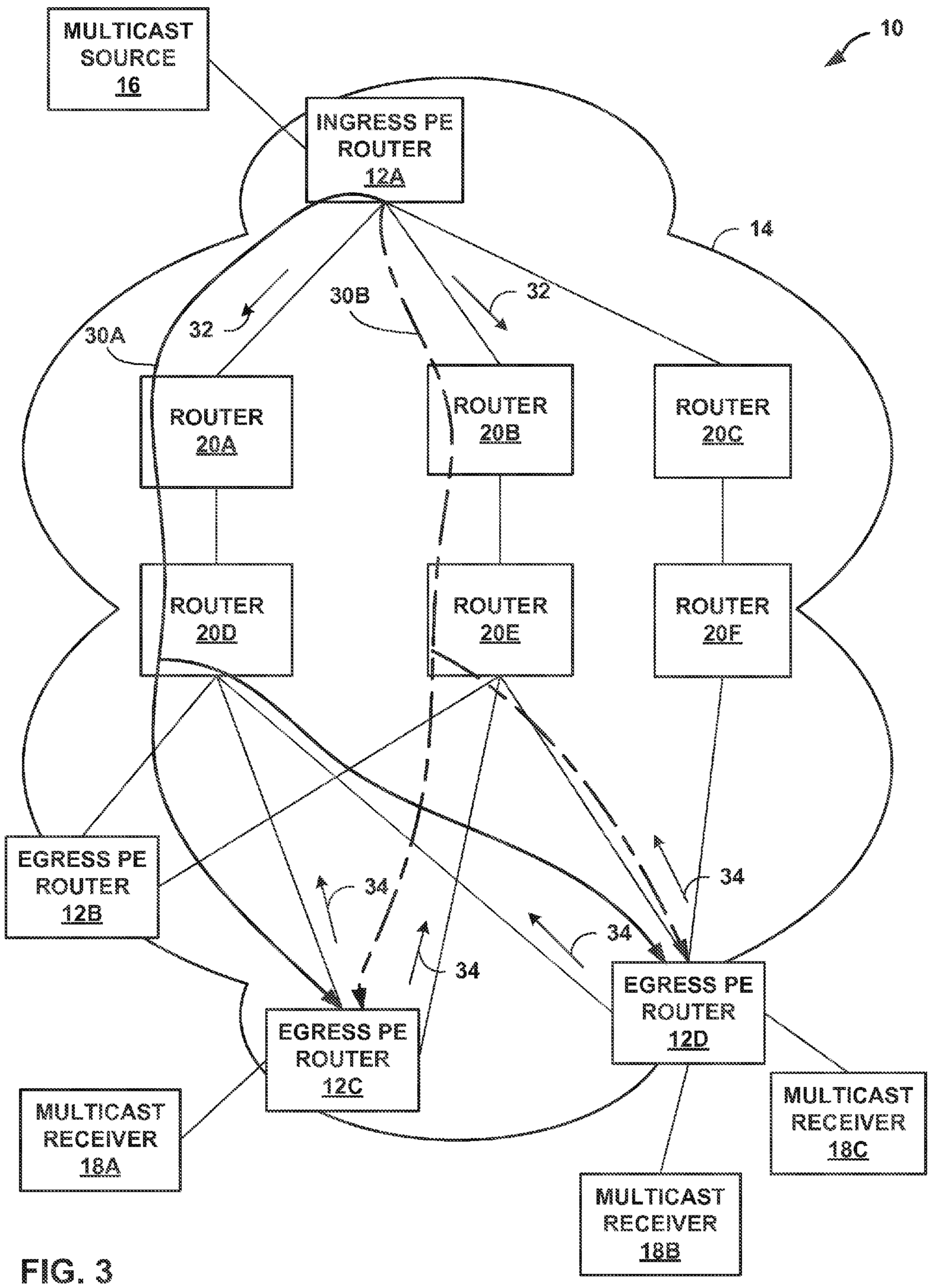

MULTICAST SOURCE 16
INGRESS PE ROUTER 12A
10
14
30A
32
30B
32
ROUTER 20A
ROUTER 20B
ROUTER 20C
ROUTER 20D
ROUTER 20E
ROUTER 20F
EGRESS PE ROUTER 12B
34
34
34
34
EGRESS PE ROUTER 12C
EGRESS PE ROUTER 12D
MULTICAST RECEIVER 18A
MULTICAST RECEIVER 18B
MULTICAST RECEIVER 18C

FIG. 3

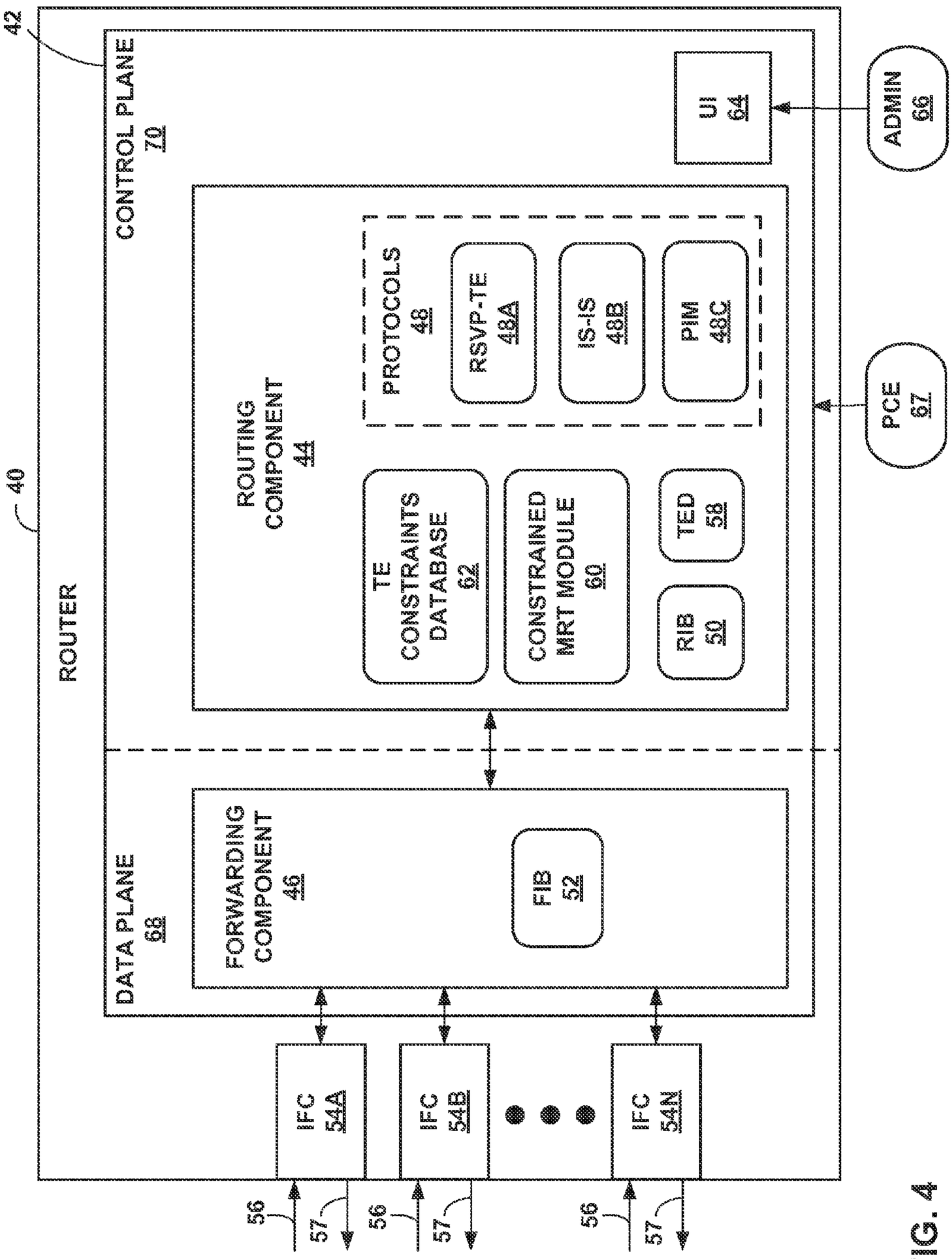

42
CONTROL PLANE
70
UI
64
ADMIN
66
40
ROUTER
ROUTING
COMPONENT
44
PROTOCOLS
48
RSVP-TE
48A
IS-IS
48B
PIM
48C
TE
CONSTRAINTS
DATABASE
62
CONSTRAINED
MRT MODULE
60
RIB
50
TED
58
PCE
67
DATA PLANE
68
FORWARDING
COMPONENT
46
FIB
52
IFC
54A
IFC
54B
IFC
54N
56
57
56
57
56
57

FIG. 4

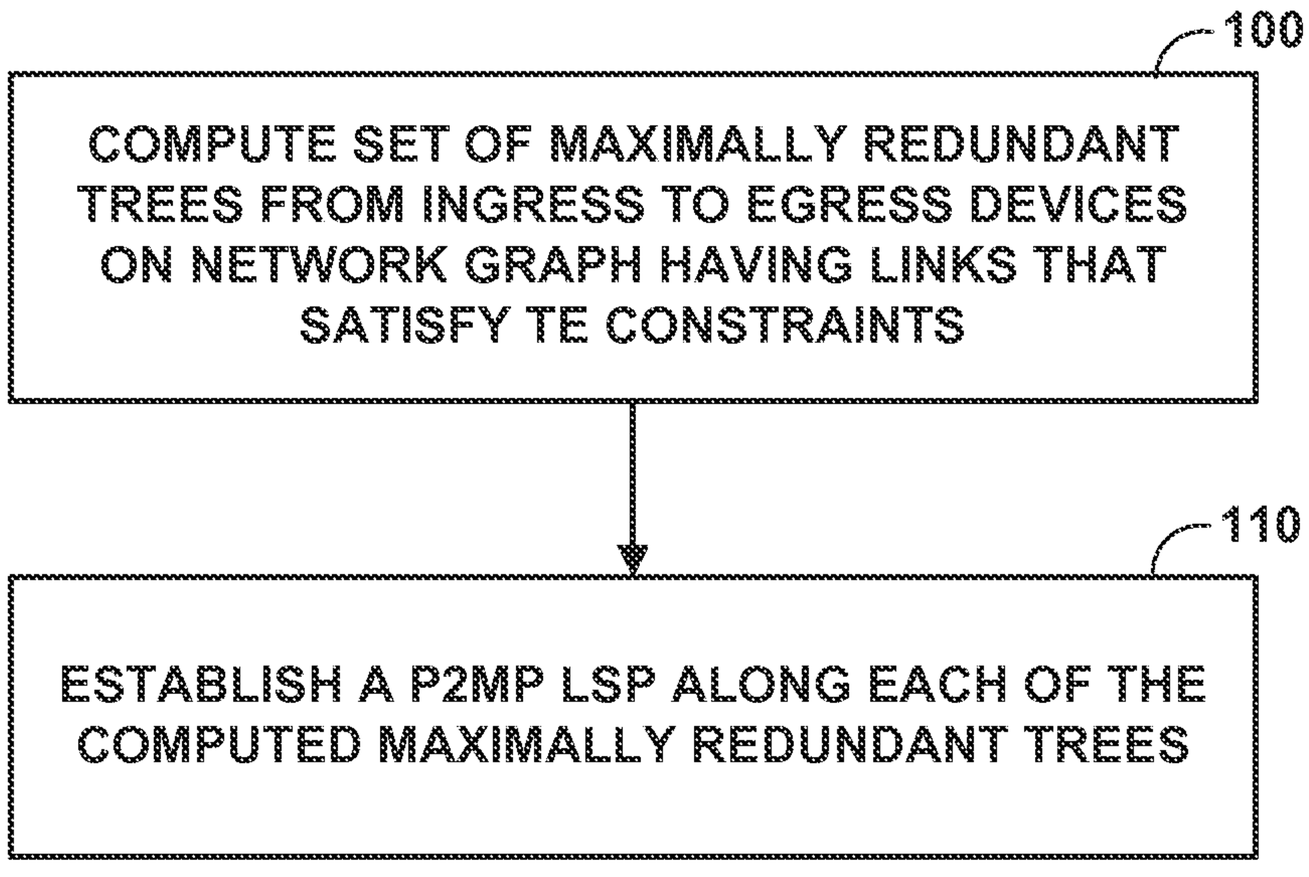

100
COMPUTE SET OF MAXIMALLY REDUNDANT TREES FROM INGRESS TO EGRESS DEVICES ON NETWORK GRAPH HAVING LINKS THAT SATISFY TE CONSTRAINTS
110
ESTABLISH A P2MP LSP ALONG EACH OF THE COMPUTED MAXIMALLY REDUNDANT TREES

FIG. 5

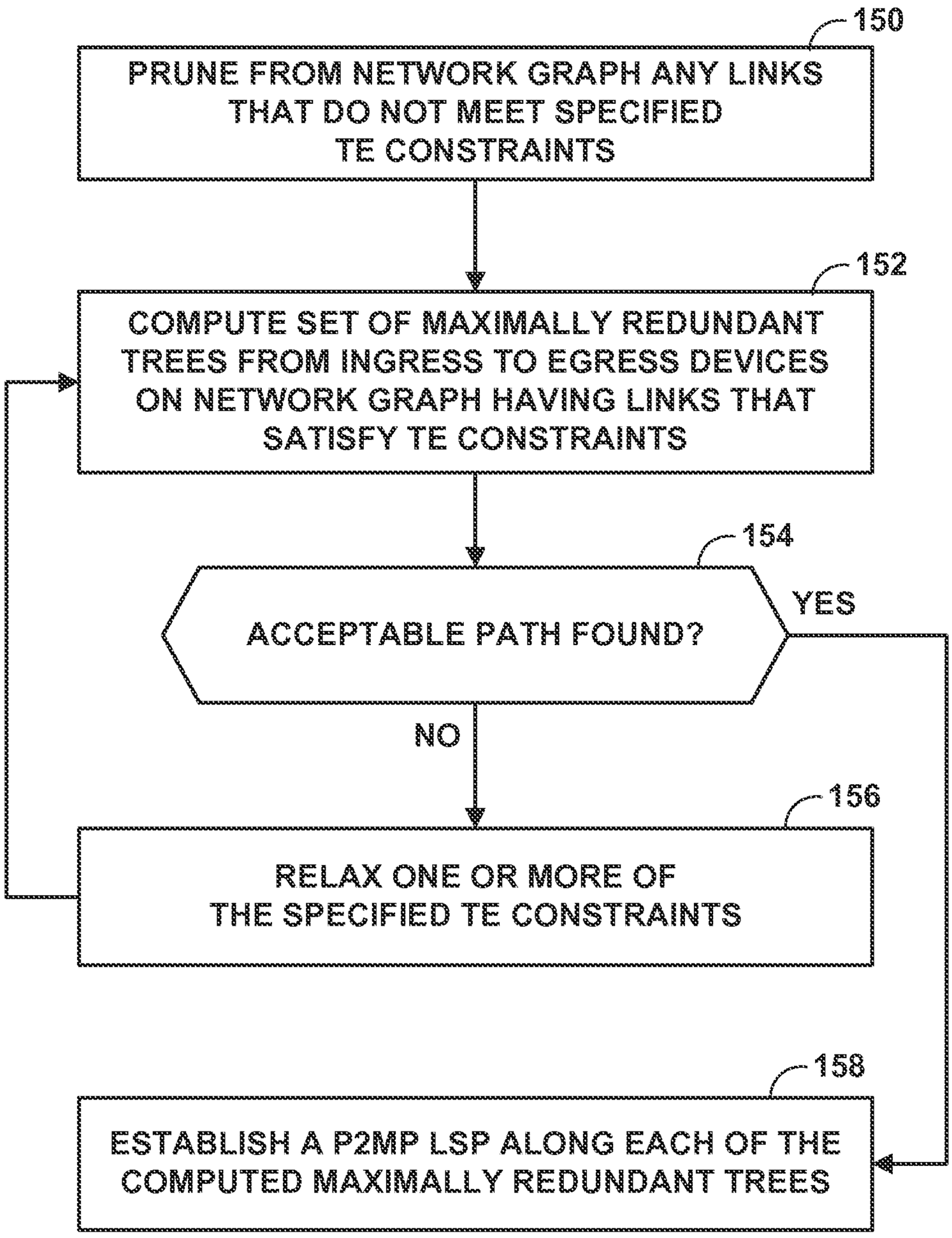

150
PRUNE FROM NETWORK GRAPH ANY LINKS THAT DO NOT MEET SPECIFIED TE CONSTRAINTS
152
COMPUTE SET OF MAXIMALLY REDUNDANT TREES FROM INGRESS TO EGRESS DEVICES ON NETWORK GRAPH HAVING LINKS THAT SATISFY TE CONSTRAINTS
154
ACCEPTABLE PATH FOUND?
YES
NO
156
RELAX ONE OR MORE OF THE SPECIFIED TE CONSTRAINTS
158
ESTABLISH A P2MP LSP ALONG EACH OF THE COMPUTED MAXIMALLY REDUNDANT TREES

FIG. 6

CONSTRAINED MAXIMALLY REDUNDANT TREES FOR POINT-TO-MULTIPOINT LSPS

TECHNICAL FIELD

The disclosure relates to computer networks and, more particularly, to forwarding network traffic within computer networks.

BACKGROUND

The term "link" is often used to refer to the connection between two devices on a computer network. The link may be a physical medium, such as a copper wire, a coaxial cable, any of a host of different fiber optic lines or a wireless connection. In addition, network devices may define "virtual" or "logical" links, and map the virtual links to the physical links. As networks grow in size and complexity, the traffic on any given link may approach a maximum bandwidth capacity for the link, thereby leading to congestion and loss.

Multi-protocol Label Switching (MPLS) is a mechanism used to engineer traffic patterns within Internet Protocol (IP) networks. By utilizing MPLS, a source device can request a path through a network, i.e., a Label Switched Path (LSP). An LSP defines a distinct path through the network to carry packets from the source device to a destination device. A short label associated with a particular LSP is affixed to packets that travel through the network via the LSP. Routers along the path cooperatively perform MPLS operations to forward the MPLS packets along the established path. LSPs may be used for a variety of traffic engineering purposes including bandwidth management and quality of service (QoS).

A variety of protocols exist for establishing LSPs. For example, one such protocol is the label distribution protocol (LDP). Procedures for LDP by which label switching routers (LSRs) distribute labels to support MPLS forwarding along normally routed paths are described in L. Anderson, "LDP Specification," RFC 3036, Internet Engineering Task Force (IETF), January 2001, the entire contents of which are incorporated by reference herein. Another type of protocol is a resource reservation protocol, such as the Resource ReserVation Protocol with Traffic Engineering extensions (RSVP-TE). RSVP-TE uses constraint information, such as bandwidth availability, to compute and establish LSPs within a network. RSVP-TE may use bandwidth availability information accumulated by a link-state interior routing protocol, such as the Intermediate System-Intermediate System (IS-IS) protocol or the Open Shortest Path First (OSPF) protocol. RSVP-TE establishes LSPs that follow a single path from an ingress device to an egress device, and all network traffic sent on the LSP must follow exactly that single path. The use of RSVP-TE, including extensions to establish LSPs in MPLS, are described in D. Awduche, "RSVP-TE: Extensions to RSVP for LSP Tunnels," RFC 3209, IETF, December 2001, the entire contents of which are incorporated by reference herein.

In some cases, RSVP-TE can be used to establish a point-to-multipoint (P2MP) LSP that can be used for sending traffic across a network from a single ingress to multiple egress routers. The use of RSVP-TE for establishing P2MP LSPs is described in R. Aggarwal, "Extensions to RSVP-TE for P2MP TE LSPs," RFC 4875, May 2007, the entire contents of which are incorporated by reference herein. A P2MP LSP is comprised of multiple source-to-leaf (S2L) sub-LSPs. These S2L sub-LSPs are set up between the ingress and egress LSRs and are appropriately combined by the branch LSRs using RSVP semantics to result in a P2MP TE LSP.

SUMMARY

In general, techniques are described for using maximally redundant trees (MRTs) to establish a set of P2MP LSPs for delivering redundant multicast streams across a network from an ingress device to multiple egress devices of the network. MRTs are a set of trees where, for each of the MRTS, a set of paths from a root node of the MRT to one or more leaf nodes share a minimum number of nodes and a minimum number of links. Techniques are described herein for a network device of a network to compute a set of MRTs that traverse links that satisfy certain constraints. For example, a network device, such as a router, can compute a set of constrained MRTs on a network graph that includes only links that satisfy one or more configured traffic-engineering (TE) constraints, such as bandwidth, link color, and the like. The network device can then use a resource reservation protocol (e.g., RSVP-TE) for establishing P2MP LSPs along the paths of the constrained MRTs to several egress network devices, yielding multiple P2MP LSPs that traverse different maximally redundant tree paths from the same ingress to egresses.

Using MRTs for computing the spanning trees for the P2MP LSPs generally provides link and node disjointness of the spanning trees to the extent physically feasible, regardless of topology, based on the topology information distributed by a link-state Interior Gateway Protocol (IGP). The techniques set forth herein provide mechanisms for handling real networks, which may not be fully 2-connected, due to previous failure or design. A 2-connected graph is a graph that requires two nodes to be removed before the network is partitioned.

The techniques may provide one or more advantages. For example, the techniques can allow for dynamically adapting in network environments in which the same traffic is sent on two or more diverse paths, such as in multicast live-live. Multicast live-live functionality can be used to reduce packet loss due to network failures on any one of the paths. The techniques of this disclosure can provide for dynamically adapting to network changes in the context of multicast live-live for RSVP-TE P2MP. The techniques can provide a mechanism that is responsive to changes in network topology without requiring manual configuration of explicit route objects or heuristic algorithms. The techniques of this disclosure do not require operator involvement to recalculate the P2MP LSPs in the case of network topology changes. The use of MRTs in this manner can provide multicast live-live functionality, and provide a mechanism for sending live-live multicast streams across an arbitrary network topology so that the disjoint trees can be dynamically recalculated by the ingress device as the network topology changes.

In one aspect, a method includes with a network device, computing a set of maximally redundant trees from an ingress network device to a plurality of egress network devices based on a network graph, in which each of the set of maximally redundant trees comprises a spanning tree to the plurality of egress network devices rooted at the ingress network device, wherein the maximally redundant trees are computed such that each of the links along each of the set of maximally redundant trees satisfies a specified traffic-engineering constraint, and with the ingress network device, establishing a plurality of point to multipoint (P2MP) label switched paths (LSPs) from the ingress network device to the plurality of egress network devices along the set of maximally redundant trees, wherein each of the P2MP LSPs corresponds to a different one of the maximally redundant trees.

In another aspect, a network device includes a constrained maximally redundant tree module to compute a set of maximally redundant trees from the network device to a plurality of egress network devices based on a network graph, in which each of the set of maximally redundant trees comprises a spanning tree to the plurality of egress network devices rooted at the network device, wherein the maximally redundant trees are computed such that each of the links along each of the set of maximally redundant trees satisfies a specified traffic-engineering constraint. The network device also includes a resource reservation protocol module to establish a plurality of P2MP LSPs from the network device as an ingress network device to the plurality of egress network devices along the set of maximally redundant trees, wherein each of the P2MP LSPs corresponds to a different one of the maximally redundant trees.

In another aspect, a computer-readable storage medium includes instructions. The instructions cause a programmable processor to compute a set of maximally redundant trees from an ingress network device to a plurality of egress network devices based on a network graph, in which each of the set of maximally redundant trees comprises a spanning tree to the plurality of egress network devices rooted at the ingress network device, wherein the maximally redundant trees are computed such that each of the links along each of the set of maximally redundant trees satisfies a specified traffic-engineering constraint, and establish a plurality of point to multipoint (P2MP) label switched paths (LSPs) from the ingress network device to the plurality of egress network devices along the set of maximally redundant trees, wherein each of the P2MP LSPs corresponds to a different one of the maximally redundant trees.

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

BRIEF DESCRIPTION OF DRAWINGS

FIG. **1** is a block diagram illustrating an example network in which one or more network devices employ the techniques of this closure.

FIG. **2** is a block diagram illustrating a modified network graph used by an ingress router of the network of FIG. **1** after pruning a link from the network graph of the network.

FIG. **3** is a block diagram illustrating example point-to-multipoint (P2MP) label switched paths (LSPs) that are established from an ingress PE router to egress PE routers in accordance with the techniques of this disclosure.

FIG. **4** is a block diagram illustrating an example router that operates in accordance with the techniques of this disclosure.

FIG. **5** is a flowchart illustrating exemplary operation of a network device, such as a router, in accordance with the techniques of this disclosure.

FIG. **6** is a flowchart illustrating exemplary operation of a network device, such as a router, in accordance with the techniques of this disclosure.

DETAILED DESCRIPTION

FIG. **1** is a block diagram illustrating an example system **10** in which a network **14** includes one or more network devices that employ the techniques of this closure. In this example, network **14** includes provider edge (PE) routers **12**A-**12**D ("PE routers **12**"), including ingress PE router **12**A and egress PE routers **12**B, **12**C, and **12**D. Network **14** also includes routers **20**A-**20**F ("routers **20**"), which may be referred to as intermediate routers.

PE routers **12** and routers **20** are coupled by links **22**A-**22**M ("links **22**"). Links **22** may be a number of physical and logical communication links that interconnect routers **12**, **20** of network **14** to facilitate control and data communication between the routers. Physical links of network **14** may include, for example, Ethernet PHY, Synchronous Optical Networking (SONET)/Synchronous Digital Hierarchy (SDH), Lambda, or other Layer 2 data links that include packet transport capability. Logical links of network **14** may include, for example, an Ethernet Virtual local area network (LAN), an MPLS LSP, or an MPLS-TE LSP.

System **10** also includes multicast source device **16** sends multicast traffic into network **14** via ingress PE router **12**A, and multicast receiver devices **18**A-**18**C ("multicast receivers **18**") that receive multicast traffic from egress PE router **12**C and egress PE router **12**D, respectively. The multicast traffic may be, for example, multicast video or multimedia traffic. For distribution of multicast traffic, including time-sensitive or critical multicast traffic, it can be desirable for routers **12**, **20** of network **14** to employ multicast live-live techniques, in which the same traffic is sent on two or more diverse paths. Multicast live-live functionality can be used to reduce packet loss due to network failures on any one of the paths. As explained in further detail below with respect to FIGS. **2-3**, ingress PE router **12**A computes a set of spanning trees that are maximally redundant trees from ingress PE router **12**A to egress PE routers **12**B, **12**C, and **12**D. Ingress PE router **12**A establishes a set of P2MP LSPs for concurrently sending the same multicast traffic from multicast source **16** to multicast receivers **18**. This provides a mechanism for sending live-live multicast streams across network **14** so that the maximally redundant trees can be dynamically recalculated by ingress PE router **12**A as the topology of network **14** changes.

Changes in the network topology may be communicated among routers **12**, **20** in various ways, for example, by using a link-state protocol, such as interior gateway protocols (IGPs) like the Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS) protocols. That is, routers **12**, **20** can use the IGPs to learn link states and link metrics for communication links **22** within the interior of network **14**. If a communication link fails or a cost value associated with a network node changes, after the change in the network's state is detected by one of the routers **12**, **20**, that router may flood an IGP Advertisement communicating the change to the other routers in the network. In other examples, routers **12**, **20** can communicate the network topology using other network protocols, such as an interior Border Gateway Protocol (iBGP), e.g., BGP-Link State (BGP-LS). In this manner, each of the routers eventually "converges" to an identical view of the network topology.

For example, ingress PE router **12**A may use OSPF or IS-IS to exchange routing information with routers **12**, **20**. Ingress PE router **12**A stores the routing information to a routing information base that ingress PE router **12**A uses to compute optimal routes to destination addresses advertised within network **14**. In addition, ingress PE router **12**A can store to a traffic engineering database (TED) any traffic engineering (TE) metrics or constraints received via the IGPs advertisements.

In accordance with the techniques of this disclosure, ingress PE router **12**A uses a method of computing the maximally redundant trees that also considers traffic-engineering constraints, such as bandwidth, link color, priority, and class type, for example. Ingress PE router **12**A creates multicast

trees by computing the entire trees as MRTs and signaling each branch of a P2MP LSP (e.g., using a resource reservation protocol such as RSVP-TE). As a further example, a path computation element (PCE) may alternatively or additionally provide configuration information to router **12**A, e.g., may compute the MRTs and provide them to ingress router **12**A. For example, network **14** may include a PCE that can learn the topology from IGP, BGP, or another mechanism and then perform a constrained MRT computation and provide the result to ingress PE router **12**A.

In accordance with one example aspect of this disclosure, ingress PE router **12**A computes a set of spanning trees that are maximally redundant trees (MRTs) over a network graph that represents at least a portion of links **22** and nodes (routers **12**, **20**) in network **14**. For example, ingress PE router **12**A can execute a shortest-path first (SPF) algorithm over its routing information base to compute forwarding paths through network **14** to egress PE routers **12**B, **12**C, and **12**D. In some instances, ingress PE router **12**A may execute a constrained SPF (CSPF) algorithm over its routing information base and its traffic engineering database to compute paths for P2MP LSPs subject to various constraints, such as link attribute requirements, input to the CSPF algorithm. For example, source router **212** may execute a CSPF algorithm subject to a bandwidth constraint that requires each link of a computed path from ingress PE router **12**A to egress PE routers **12**B, **12**C, and **12**D to have at least a specified amount of maximum link bandwidth, residual bandwidth, or available bandwidth.

Prior to computing the set of MRTs, ingress PE router **12**A may prune from the network graph any links **22** that do not satisfy one or more specified TE constraints. In some examples, ingress PE router **12**A may obtain the network graph having links that each satisfy the TE constraints by starting with an initial network graph based on stored network topology information, and pruning links of the initial network graph to remove any network links that do not satisfy the TE constraints, resulting in a modified network graph. In this example, ingress PE router **12**A uses the modified network graph for computing the set of MRTs. In this manner, ingress PE router **12**A computes the set of MRTs over a network graph in which all links satisfy the specified constraints, resulting in a set of "constrained MRTs." In some examples, ingress PE router **12**A may prune the links from the network graph as part of the CSPF computation.

In some aspects, ingress PE router **12**A may compute the MRTs in response to receiving a request to traffic-engineer a diverse set of P2MP LSPs to the plurality of egress PE routers **12**B-**12**D, such as to be used for multicast live-live redundancy in forwarding multicast content. For example, a network administrator may configure ingress PE router **12**A with the request. The request may specify that the P2MP LSPs satisfy certain constraints, including, for example, one or more of bandwidth, link color, Shared Risk Link Group (SRLG), priority, class type, and the like.

For example, suppose that ingress PE router **12**A is configured to compute a set of MRTs and establish a set of corresponding P2MP LSPs along trees in which all links have an available bandwidth of 50 megabytes per second (mBps). Assume that all of the links **22** of the initial network graph **26** of FIG. **1** satisfy this constraint, except for link **22**F between router **20**C and router **20**F, which has only 10 mBps of available bandwidth and therefore does not satisfy the specified constraint. Prior to computing the set of MRTs, ingress PE router **12**A prunes link **22**F from its network graph **26** to be used for computing the MRTs.

FIG. **2** is a block diagram illustrating the modified network graph **28** used by ingress PE router **12**A of network **14** after pruning link **22**F from the network graph **26** (FIG. **1**) of network **14**. FIG. **2** includes the components of FIG. **1** (with link reference numerals omitted for simplification). After pruning from the network graph **26** any links that do not satisfy the specified TE constraints, ingress PE router **12**A then computes a set of MRTs **25**A-**25**B ("MRTs **25**") on the modified network graph **28**, which results in constrained MRTs **25** that include only paths on links that each satisfy the TE constraints. The constrained MRTs **25** are spanning trees to reach all routers in the network graph, rooted at ingress PE router **12**A. Each of MRTs **25** is computed from the ingress PE router **12**A to egress PE routers **12**B, **12**C, **12**D. Ingress PE router **12**A computes MRTs **25** as a pair of MRTs that traverse maximally disjoint paths from the ingress PE router **12**A to egress PE routers **12**B, **12**C, **12**D. The pair of MRTs **25** may sometimes be referred to as the Blue MRT and the Red MRT.

The following terminology is used herein. A network graph is a graph that reflects the network topology where all links connect exactly two nodes and broadcast links have been transformed into the standard pseudo-node representation. The term "2-connected," as used herein, refers to a graph that has no cut-vertices, i.e., a graph that requires two nodes to be removed before the network is partitioned. A "cut-vertex" is a vertex whose removal partitions the network. A "cut-link" is a link whose removal partitions the network. A cut-link by definition must be connected between two cut-vertices. If there are multiple parallel links, then they are referred to as cut-links in this document if removing the set of parallel links would partition the network.

A "2-connected cluster" is a maximal set of nodes that are 2-connected. The term "2-edge-connected" refers to a network graph where at least two links must be removed to partition the network. The term "block" refers to either a 2-connected cluster, a cut-edge, or an isolated vertex. A Directed Acyclic Graph (DAG) is a graph where all links are directed and there are no cycles in it. An Almost Directed Acyclic Graph (ADAG) is a graph that, if all links incoming to the root were removed, would be a DAG. A Generalized ADAG (GADAG) is a graph that is the combination of the ADAGs of all blocks. Further information on MRTs may be found at A. Atlas, "An Architecture for IP/LDP Fast-Reroute Using Maximally Redundant Trees," Internet-Draft, draft-atlas-rtgwg-mrt-frr-architecture-01, October, 2011; A. Atlas, "Algorithms for Computing Maximally Redundant Trees for IP/LDP Fast-Reroute, Internet-Draft, draft-enyedi-rtgwg-mrt-frr-algorithm-01, November, 2011; A. Atlas, "An Architecture for Multicast Protection Using Maximally Redundant Trees," Internet-Draft, draft-atlas-rtgwg-mrt-mc-arch-00, March 2012; the entire contents of each of which are incorporated by reference herein.

Redundant trees are directed spanning trees that provide disjoint paths towards their common root. These redundant trees only exist and provide link protection if the network graph is 2-edge-connected and node protection if the network graph is 2-connected. Such connectiveness may not be the case in real networks, either due to architecture or due to a previous failure. Maximally redundant trees are useful in a real network because they may be computable regardless of network topology. Maximally Redundant Trees (MRT) are a set of trees where the path from any node X to the root R along one tree and the path from the same node X to the root along any other tree of the set of trees share the minimum number of nodes and the minimum number of links. Each such shared node is a cut-vertex. Any shared links are cut-links. That is, the maximally redundant trees are computed so that only the

cut-edges or cut-vertices are shared between the multiple trees. In any non-2-connected graph, only the cut-vertices and cut-edges can be contained by both of the paths. That is, a pair of MRTs, such as MRT **25**A and MRT **25**B, are a pair of trees that share a least number of links possible and share a least number of nodes possible. Any RT is an MRT but many MRTs are not RTs. MRTs are practical to maintain redundancy even after a single link or node failure. If a pair of MRTs is computed rooted at each destination, all the destinations remain reachable along one of the MRTs in the case of a single link or node failure. The MRTs of a pair of MRTs may be individually referred to as a Red MRT and a Blue MRT.

Computationally practical algorithms for computing MRTs may be based on a common network topology database. A variety of algorithms may be used to calculate MRTs for any network topology. These may result in trade-offs between computation speed and path length. Many algorithms are designed to work in real networks. For example, just as with SPF, an algorithm is based on a common network topology database, with no messaging required. In one example aspect, MRT computation for multicast Live-Live may use a path-optimized algorithm based on heuristics. Some example algorithms for computing MRTs can be found in U.S. patent application Ser. No. 13/418,212, entitled "Fast Reroute for Multicast Using Maximally Redundant Trees," filed on Mar. 12, 2012, the entire contents of which are incorporated by reference herein.

In the example of FIG. **2**, for example, ingress PE router **12**A computes MRTs **25**A and **25**B from ingress PE router **12**A to egress PE routers **12**B, **12**C, and **12**D. MRT **25**A includes a path from router **12**A to router **20**A to router **20**D. At router **20**D, MRT **25**A branches off to three branches, with a first branch from router **20**D to egress router **12**B, a second branch from router **20**D to egress router **12**B, and a third branch from router **20**D to egress PE router **12**D. MRT **25**B includes a path from router **12**A to router **20**B to router **20**E. At router **20**D, MRT **25**B branches off to three branches, with a first branch from router **20**E to egress router **12**B, a second branch from router **20**E to egress router **12**B, and a third branch from router **20**E to egress PE router **12**D. Although described for purposes of example in terms of a set of MRTs being a pair of MRTs, in other examples ingress PE router **12**A may compute a set of constrained MRTs that includes more than two MRTs, where each MRT of the set traverses a different path, where each path is as diverse as possible from each other path. In such examples, more complex algorithms may be needed to compute a set of MRTs that includes more than two MRTs.

In one example aspect, for constraints based upon path, as compared to link attributes, the SPF-based MRT algorithm used by ingress PE router **12**A can keep track of the constraint at each node, for each direction. The algorithm would then not attach to a node in the GADAG if that node would cause too large a value. If the first try does not find an acceptable tree to egress PE routers **12**B, **12**C, and **12**D, ingress PE router **12**A can be configured to relax one or more of the constraints and try again to find an acceptable tree. That is, if the re-computed pair of maximally redundant trees MRTs is not 2-connected, ingress PE router **12**A may repeat the steps of modifying the specified traffic-engineering constraint, modifying the network graph, and re-computing the pair of maximally redundant trees until re-computing the pair of maximally redundant trees yields a pair of maximally redundant trees that are 2-connected.

For example, if the computed pair of maximally redundant trees MRTs is not 2-connected, ingress PE router **12**A can modify the specified traffic-engineering constraint to have a less restrictive value, and modify the network graph to add back links to the network graph that satisfy the modified traffic-engineering constraint to obtain a modified network graph. Ingress PE router **12**A can then re-compute the pair of maximally redundant trees based on the modified network graph, to attempt to obtain a pair of 2-connected MRTs.

There may be a trade-off between how much the network graph is pruned based on constraints, and whether this will result in a 2-connected network after the pruning. The service provider may prefer to scale back or add costs in some fashion. How this is done depends on whether it is more important to meet the constraints or more important to have path diversity. These preferences may be a configurable option on ingress PE router **12**A.

So for example, ingress PE router **12**A may do an initial computation based on initial specified constraints. If the resulting trees are not 2-connected, ingress PE router **12**A identifies the cut nodes and cut links, looks at the links that were removed from the topology that were connected to that cut-node and cut-links, and makes a selection among the removed links based on preferences, such as cost of the removed links, or preference as to which constraint to drop off first.

In some aspects, ingress PE router **12**A would not necessarily do a whole re-pruning of the entire tree, but instead might be targeted to a certain part of the tree where it is not 2-connected. This is because it may be better to only add back in additional links when needed in order to get path diversity, and the algorithm would otherwise respect the initial pruning that was done.

FIG. **3** is a block diagram illustrating example point-to-multipoint (P2MP) label switched paths (LSPs) **30**A-**30**B ("P2MP LSPs **30**") that are established from ingress PE router **12**A to egress PE routers **12**C and **12**D, in accordance with the techniques of this disclosure. In some aspects, routers **12**, **20** of FIGS. **1-3** may be Internet Protocol (IP) routers that implement Multi-Protocol Label Switching (MPLS) techniques and operate as label switching routers (LSRs). For example, ingress PE router **12**A can assign a label to each incoming packet based on its forwarding equivalence class before forwarding the packet to a next-hop router **20**. Each router **20** makes a forwarding selection and determines a new substitute label by using the label found in the incoming packet as a reference to a label forwarding table that includes this information.

The paths taken by packets that traverse the network in this manner are referred to as LSPs, and in the current example may be P2MP LSPs. To establish a traffic-engineered P2MP LSP, ingress PE router **12**A computes a P2MP path, and initiates signaling along the P2MP path. LSRs along the P2MP path modify their forwarding tables based on the signaling. LSRs use MPLS-TE techniques to establish LSPs that have guaranteed bandwidth under certain conditions. For example, the TE-LSPs may be signaled through the use of the RSVP protocol and, in particular, by way of RSVP-TE signaling messages sent between routers **12**, **20**.

In the example of FIG. **3**, after computing the set of MRTs **25**, ingress PE router **12**A establishes a pair of P2MP LSPs **30**A-**30**B along each of the MRTs **25**. In some cases, ingress PE router **12**A may not establish branches of the P2MP LSPs along every one of the branches of the corresponding MRTs **25**. For example, in FIG. **3**, egress PE router **12**B is not associated with any multicast receiver devices that request to receive multicast traffic from multicast source **16**. Thus, as shown in FIG. **3**, MRTs **25** each have a plurality of branches, and ingress PE router **12**A establishes P2MP LSPs **30** along only the subset of the possible branches of the MRTs **25**, to

only a subset of the possible egress PE routers **12** that are of interest to ingress PE router **12**A.

Generally stated, ingress PE router **12**A establishes the first P2MP LSP **30**A along the first maximally redundant tree **25**A by sending resource reservation requests **32** to routers **20** (LSRs) along the first maximally redundant tree **25**A. The resource reservation requests **32** can each include an identifier associating the requests with the first maximally redundant tree. In addition, ingress PE router **12**A receives resource reservation messages **34** in response to the resource reservation requests **32**, where the resource reservation messages **34** specify reserved resources and labels allocated to the P2MP LSP to be used for forwarding network traffic to corresponding next hops along the sub-paths of the P2MP LSP, wherein the resource reservation messages each include an identifier associating the messages with the same P2MP LSP. The same process is used for establishing the second P2MP LSP **30**B along the MRT **25**B.

The example of FIG. **3** is now described with reference to the specific example of RSVP-TE. For example, in accordance with RSVP-TE, to establish the P2MP LSPs **30**A-**30**B LSP between ingress PE router **12**A and egress PE routers **12**C and **12**D, ingress PE router **12**A may send multiple RSVP-TE Path messages **32** downstream hop-by-hop along the MRTs **25**A, **25**B to the egress PE routers **12**C and **12**D to identify the sender and indicate TE constraints (e.g., bandwidth) needed to accommodate the data flow, along with other attributes of the P2MP LSP. The Path messages **32** may contain various information about the TE-LSP including, e.g., various characteristics of the TE-LSP. For example, the Path messages **32** sent by ingress PE router **12**A may specify the hop-by-hop path to be established by way of an Explicit Route Object (ERO) contained in the Path message **32**. The ERO can define the trees corresponding to the MRTs.

A P2MP LSP is comprised of multiple source-to-leaf (S2L) sub-LSPs. These S2L sub-LSPs are set up between the ingress and egress LSRs and are appropriately combined by the branch LSRs using RSVP semantics to result in a P2MP TE LSP. One Path message may signal one or multiple S2L sub-LSPs for a single P2MP LSP. Hence the S2L sub-LSPs belonging to a P2MP LSP can be signaled using one Path message or split across multiple Path messages. Details of the RSVP-TE signaling semantics for establishing P2MP LSPs are described in R. Aggarwal, "Extensions to RSVP-TE for P2MP TE LSPs," RFC 4875, May 2007, the entire contents of which are incorporated by reference herein.

After receiving a Path message **32**, routers **20** may update their forwarding tables, allocate an MPLS label, and forward the Path message **32** to the next hop along the path specified in the Path message. To establish the P2MP LSP (data flow) between the receiver and the sender, egress PE routers **12**D may return RSVP-TE Reserve (Resv) messages **34** upstream along the paths of the MRTs to ingress PE router **12**A to confirm the attributes of the P2MP LSPs, and provide respective LSP labels.

The P2MP LSPs **30**A and **30**B of FIG. **3** may be used by network **14** for a variety of applications, such as Layer 2 Multicast over P2MP Multi-Protocol Label Switching (MPLS) TE, Internet Protocol (IP) Multicast over P2MP MPLS TE, Multicast VPNs (MVPNs) over P2MP MPLS TE, and Virtual Private Local Area Network Service (VPLS) Multicast over P2MP MPLS TE, for example.

After establishing the P2MP LSPs **30**, ingress PE router **12**A may receive multicast data traffic from multicast source device **16**, and ingress PE router **12**A can forward the multicast data traffic along both of P2MP LSPs **30**A and **30**B. That is, ingress PE router **12**A concurrently sends multicast traffic received from the multicast source device **16** to the plurality of multicast receivers **18** on both of the first P2MP LSP **30**A and the second P2MP LSP **30**B. In this manner, ingress PE router **12**A sends redundant multicast data traffic along both of P2MP LSPs **30**A and **30**B, which provides multicast live-live service.

The techniques of this disclosure allow for dynamically adapting multicast live-live for RSVP-TE P2MP, and provides a mechanism that is responsive to changes in network topology without requiring manual configuration of explicit route objects or heuristic algorithms. Operator involvement is not needed to recalculate the P2MP LSPs in the case of network topology changes.

If ingress PE router **12**A detects that changes have occurred to the topology of system **10**, ingress PE router **12**A may re-compute MRTs **25**, or portions of MRTs **25**, to determine whether changes are needed to P2MP LSPs **30**. For example, if ingress PE router **12**A detects that that a new multicast receiver is added to system **10** coupled to egress PE router **12**B, ingress PE router **12**A can send an updated Path message to add a branch to each of P2MP LSPs **30**, e.g., from router **20**D to egress PE router **12**B and from router **20**E to egress PE router **12**B, without needing to re-signal the entire P2MP LSPs **30**.

For multicast live-live to provide the desired protection, a safe way is needed of transitioning multicast traffic from being sent on the pair of MRTs computed on the old topology to the pair of MRTs computed on the new topology. A make-before-break process may be used. For example, as the topology of network **14** changes, ingress PE router **12**A can compute and signal a new set of MRTs (not shown), and once state for those new MRTs is successfully created and usable, ingress PE router **12**A can easily transition from sending on the old MRTs **25** to sending multicast traffic on the new MRTs. Then ingress PE router **12**A can then tear down the old MRTs **25**.

In some aspects, ingress PE router **12**A may be configured to run a periodic re-optimization of the MRT computation, which may result in a similar transition from old MRTs to new MRTs. For example, ingress PE router **12**A can periodically rerun the CSPF computation to recompute the pair of MRTs, to determine whether a more optimal pair of MRTs exists on the network graph.

Router **12**A can use any of a variety of advertised link bandwidths as bandwidth information for pruning a network graph to remove links that do not satisfy specified TE constraints. As one example, the advertised link bandwidth may be a "maximum link bandwidth." The maximum link bandwidth defines a maximum amount of bandwidth capacity associated with a network link. As another example, the advertised link bandwidth may be a "residual bandwidth," i.e., the maximum link bandwidth less the bandwidth currently reserved by operation of a resource reservation protocol, such as being reserved to RSVP-TE LSPs. This is the bandwidth available on the link for non-RSVP traffic. Residual bandwidth changes based on control-plane reservations.

As a further example, the advertised link bandwidth may be an "available bandwidth" (also referred to herein as "currently available bandwidth"). The available bandwidth is the residual bandwidth less measured bandwidth used to forward non-RSVP-TE packets. In other words, the available bandwidth defines an amount of bandwidth capacity for the network link that is neither reserved by operation of a resource reservation protocol nor currently being used by the first router to forward traffic using unreserved resources. The

amount of available bandwidth on a link may change as a result of SPF, but may be a rolling average with bounded advertising frequency.

As one example, the computing router (e.g., routers **12**, **20**) can determine an amount of bandwidth capacity for a network link that is reserved by operation of a resource reservation protocol, and can determine an amount of bandwidth capacity that is currently being used by the router to forward traffic using unreserved resources, by monitoring traffic through the data plane of the computing router, and may calculate the amount of residual bandwidth and/or available bandwidth based on the monitored traffic.

In the example of currently available bandwidth, routers **12**, **20** may advertise currently available bandwidth for the links **22** of network **16**, which takes into account traffic that may otherwise be unaccounted for. That is, routers **12**, **20** can monitor and advertise currently available bandwidth for a link, expressed as a rate (e.g., mBps), that takes into account bandwidth that is neither reserved via RSVP-TE nor currently in use to transport Internet Protocol (IP) packets or LDP packets over the link, where an LDP packet is a packet having an attached label distributed by LDP. Currently available bandwidth for a link is therefore neither reserved nor being used to transport traffic using unreserved resources. Routers **12**, **20** can measure the amount of bandwidth in use to transport IP and LDP packets over outbound links and compute currently available bandwidth as a difference between the link capacity and the sum of reserved bandwidth and measured IP/LDP packet bandwidth.

Routers **12**, **20** can exchange computed available bandwidth information for their respective outbound links as link attributes in extended link-state advertisements of a link-state interior gateway protocol and store received link attributes to a respective Traffic Engineering Database (TED) that is distinct from the generalized routing information base (including, e.g., the IGP link-state database). The computing device, such as ingress PE router **12**A, may execute an IGP-TE protocol, such as OSPF-TE or IS-IS-TE that has been extended to advertise link bandwidth information. For example, routers **12**, **20** may advertise a maximum link bandwidth, a residual bandwidth, or a currently available bandwidth for the links **22** of network **14** using a type-length-value (TLV) field of a link-state advertisement. As another example, an OSPF-TE protocol may be extended to include an OSPF-TE Express Path that advertises maximum link bandwidth, residual bandwidth and/or available bandwidth. Details on OSPF-TE Express Path can be found in S. Giacalone, "OSPF Traffic Engineering (TE) Express Path," Network Working Group, Internet Draft, September 2011, the entire contents of which are incorporated by reference herein.

FIG. **4** is a block diagram illustrating an example router **40** that operates in accordance with the techniques of this disclosure. Router **40** may correspond to any of routers **12**, **20** of FIGS. **1-3**. Router **40** includes interface cards **54**A-**54**N ("IFCs **54**") for receiving packets via input links **56**A-**56**N ("input links **56**") and sending packets via output links **57**A-**57**N ("output links **57**"). IFCs **54** are interconnected by a high-speed switch (not shown) and links **56**, **57**. In one example, switch **40** comprises switch fabric, switchgear, a configurable network switch or hub, and the like. Links **56**, **57** comprise any form of communication path, such as electrical paths within an integrated circuit, external data busses, optical links, network connections, wireless connections, or other type of communication path. IFCs **54** are coupled to input links **56** and output links **57** via a number of interface ports (not shown).

When router **40** receives a packet via one of input links **56**, control unit **42** determines via which of output links **57** to send the packet. Control unit **42** includes routing component **44** and forwarding component **46**. Routing component **44** determines one or more routes through a network, e.g., through interconnected devices such as other routers. Control unit **42** provides an operating environment for protocols **48**, which are typically implemented as executable software instructions. As illustrated, protocols **48** include RSVP-TE **48**A and intermediate system to intermediate system (IS-IS) **48**B. Router **40** uses RSVP-TE **48**A to set up LSPs. As described herein, RSVP-TE **48**A is programmatically extended to allow for establishment of LSPs that include a plurality of sub-paths on which traffic is load balanced between the ingress router and the egress router of the LSPs. Protocols **48** also include Protocol Independent Multicast **48**C, which can be used by router **40** for transmitting multicast traffic. Protocols **48** may include other routing protocols in addition to or instead of RSVP-TE **48**A and IS-IS **48**B, such as other Multi-protocol Label Switching (MPLS) protocols including LDP; or routing protocols, such as Internet Protocol (IP), the open shortest path first (OSPF), routing information protocol (RIP), border gateway protocol (BGP), interior routing protocols, other multicast protocols, or other network protocols.

By executing the routing protocols, routing component **44** identifies existing routes through the network and determines new routes through the network. Routing component **44** stores routing information in a routing information base (RIB) **50** that includes, for example, known routes through the network. RIB **50** may simultaneously include routes and associated next-hops for multiple topologies, such as the Blue MRT topology (e.g., MRT **25**A) and the Red MRT topology (e.g., MRT **25**B).

Forwarding component **46** stores forwarding information base (FIB) **52** that includes destinations of output links **57**. FIB **52** may be generated in accordance with RIB **50**. FIB **52** may be a radix tree programmed into dedicated forwarding chips, a series of tables, a complex database, a link list, a radix tree, a database, a flat file, or various other data structures. FIB **52** may include MPLS labels, such as for RSVP-TE LSPs. FIB **52** may simultaneously include labels and forwarding next-hops for multiple topologies, such as the Blue MRT topology MRT **25**A and the Red MRT topology MRT **25**B.

A system administrator ("ADMIN 66") may provide configuration information to router **40** via user interface **64** ("UI **64**") included within control unit **42**. For example, the system administrator **66** may configure router **40** or install software to provide constrained MRT functionality as described herein. As another example, the system administrator **66** may configure RSVP-TE **48**A with a request to traffic-engineer a set of P2MP LSPs from an ingress router to a plurality of egress routers. As a further example, a path computation element (PCE) **67** may alternatively or additionally provide configuration information to router **40**, e.g., may compute the set of MRTs and provide them to router **40**.

Router **40** includes a data plane **68** that includes forwarding component **46**. In some aspects, IFCs **54** may be considered part of data plane **68**. Router **40** also includes control plane **70**. Control plane **234** includes routing component **44** and user interface (UI) **64**. Although described for purposes of example in terms of a router, router **40** may be, in some examples, any network device capable of performing the techniques of this disclosure, including, for example, a network device that includes routing functionality and other functionality.

As shown in FIG. **4**, control plane **70** of router **40** has a modified CSPF module, referred to as constrained MRT module **60** that computes the trees using an MRT algorithm. In some aspects, constrained MRT module **60** may compute the MRTs in response to receiving a request to traffic-engineer a diverse set of P2MP LSPs to a plurality of egress routers, such as to be used for multicast live-live redundancy in forwarding multicast content. For example, administrator **66** may configure router **40** with the request via UI **64**. The request may specify that the P2MP LSPs satisfy certain constraints. TE constraints specified by the request may include, for example, bandwidth, link color, Shared Risk Link Group (SRLG), and the like. Router **40** may store the specified TE constraints to TE constraints database **62**.

Constrained MRT module **60** computes a set of MRTs from router **40** as the ingress device, to a plurality of egress devices. Router **40** computes the set of MRTs on a network graph having links that each satisfy stored traffic engineering (TE) constraints obtained from TE constraints database **62** in the control plane **70** of router **40**. In some examples, constrained MRT module **60** may obtain the network graph having links that each satisfy the TE constraints by starting with an initial network graph based on network topology information obtained from TED **58**, and pruning links of the initial network graph to remove any network links that do not satisfy the TE constraints, resulting in a modified network graph. In this example, constrained MRT module **60** uses the modified network graph for computing the set of MRTs.

After computing the set of MRTs, router **40** establishes multiple P2MP LSPs from router **40** to the egress network devices, such as by using a resource reservation protocol (e.g., RSVP-TE) to send Path messages that specify a constrained path for setting up the P2MP LSPs. For example, constrained MRT module **60** invokes RSVP-TE **48**A to carry out the signaling of P2MP LSPs along the trees, and each of the LSRs along the signaled tree installs the necessary forwarding state based on the signaling. For example, constrained MRT module **60** can communicate with RSVP-TE **48**A to provide RSVP-TE module **48**A with the computed set of MRTs to be used for signaling the P2MP LSPs. Router **40** can establish a different P2MP LSP for each MRT of the set of MRTs, such as shown in FIG. **3**. An LSP ID field in the Path messages may be used to identify each P2MP LSP. This can distinguish between an old Red MRT and a new Red MRT, for example, when transitioning from the old to new Red MRT after recomputing the MRT.

In the example of FIG. **4**, IS-IS **48**B of router **40** can receive advertisements from other routers **212** of system **200**, formed in accordance with traffic engineering extensions to include available, unreserved bandwidth for advertised links. IS-IS **48**B of router **40** can also send such advertisements to other routers advertising available bandwidth. IS-IS **48**B stores advertised available bandwidth values for advertised links in traffic engineering database (TED) **58**. Router **40** may use the available bandwidth information from TED **58** when computing the MRTs **25**. In addition to available bandwidth, the TED **58** may store costs of each advertised link, such as latency, metric, number of hops, link color, and Shared Risk Link Group (SRLG), geographic location, or other characteristics that may be used as traffic-engineering constraints. Router **40** and other LSRs may determine available bandwidth of its associated links, as described in U.S. Ser. No. 13/112,961, entitled "Weighted Equal-Cost Multipath," filed May 20, 2011, the entire contents of which are incorporated by reference herein.

FIG. **5** is a flowchart illustrating exemplary operation of a network device, such as a router, in accordance with the techniques of this disclosure. For purposes of example, FIG. **5** will be explained with reference to ingress network device **12**A of FIGS. **1-3** and router **40** of FIG. **4**. In the example of FIG. **5**, constrained MRT module **60** of router **40** computes a set of maximally redundant trees (MRTs) from router **40** as the ingress device, to a plurality of egress devices. Constrained MRT module **60** computes the set of MRTs on a network graph having links that each satisfy traffic engineering (TE) constraints (**100**).

After computing the set of MRTs, router **40** uses RSVP-TE **48**A to establish multiple P2MP LSPs from router **40** to the egress network devices (**110**), such as by using a resource reservation protocol (e.g., RSVP-TE) to send Path messages that specify a constrained path for setting up the P2MP LSPs. Router **40** can establish a different P2MP LSP for each MRT of the set of MRTs. In some aspects, router **40** may establish a P2MP LSP along only a subset of branches/paths of an MRT, when there are egress devices included as leaf nodes of the MRT that router **40** does not need for sending multicast traffic to any receiver devices. In this case, router **40** may establish the P2MP LSP along a subset of the computed MRT.

FIG. **6** is a flowchart illustrating exemplary operation of a network device, such as a router, in accordance with the techniques of this disclosure. For purposes of example, FIG. **6** will be explained with reference to ingress network device **12**A of FIGS. **1-3** and router **40** of FIG. **4**.

Ingress PE router **12**A can prune from the network graph of network **14** any links that do not meet specified TE constraints (**150**). Ingress PE router **12**A computes a set of maximally redundant trees from ingress PE router **12**A to the egress devices **12**B-**12**D on the network graph **28** (FIG. **2**) having links **22** that satisfy the TE constraints (**152**).

If ingress PE router **12**A finds an acceptable set of MRTs (YES branch of **154**), then ingress PE router **12**A can establish a P2MP LSP along each of the computed MRTs to the egress routers of interest, as described above (**158**).

If ingress PE router **12**A does not find an acceptable set of MRTs (NO branch of **154**), e.g., perhaps the pair of MRTs are not 2-connected, then ingress PE router **12**A may be configured to relax one or more of the specified TE constraints (**156**) to obtain a new modified network graph having links that satisfy the relaxed constraints, and re-compute the set of MRTs having links that satisfy the relaxed TE constraints (**152**). Ingress PE router **12**A may relax the TE constraints more than once, or may first relax a first specified TE constraint (e.g., link color), followed by relaxing a second specified TE constraint (e.g., bandwidth) if relaxing the link color constraint does not yield a pair of 2-connected MRT paths.

In some aspects, ingress PE router **12**A would not necessarily do a whole re-pruning of the entire tree, but instead might be targeted to a certain part of the tree where it is not 2-connected. This is because it may be better to only add back in additional links when needed in order to get path diversity, and the algorithm would otherwise respect the initial pruning that was done.

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.

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.

The techniques described in this disclosure may also be embodied or encoded in a computer-readable medium, such as a 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 media may include non-transitory computer-readable storage media and transient communication media. Computer readable storage media, which is tangible and non-transitory, 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, carrier waves, or other transient media.

Various aspects of this disclosure have been described. These and other aspects are within the scope of the following claims.

The invention claimed is:

1. A method comprising:

by a network device, calculating a plurality of maximally redundant trees from an ingress network device to a plurality of egress network devices based on a network graph, in which each of the plurality of maximally redundant trees comprises a spanning tree to the plurality of egress network devices rooted at the ingress network device, wherein each of the maximally redundant trees is calculated to comprise a point to multipoint (P2MP) path from the ingress network device to the plurality of egress network devices that is as disjoint as possible from a respective P2MP path from the ingress network device to the plurality of egress network devices for each other one of the plurality of maximally redundant trees, and wherein the maximally redundant trees are calculated such that each link along each of the plurality of maximally redundant trees satisfies a specified traffic-engineering constraint;

in response to determining, by the network device, that the plurality of maximally redundant trees include at least one node whose removal partitions a network represented by the network graph:

modifying, by the network device, the specified traffic-engineering constraint to have a less restrictive value;

modifying the network graph to add links to the network graph that satisfy the modified traffic-engineering constraint to obtain a modified network graph; and

re-calculating, by the network device, at least a portion of the plurality of maximally redundant trees based on the modified network graph to obtain a plurality of maximally redundant trees in which at least two nodes must be removed before the network is partitioned; and

with the ingress network device, establishing a plurality of P2MP label switched paths (LSPs) from the ingress network device to the plurality of egress network devices along each of the plurality of maximally redundant trees in which at least two nodes must be removed before the network is partitioned, wherein each of the P2MP LSPs corresponds to a different one of the plurality of maximally redundant trees in which at least two nodes must be removed before the network is partitioned.

2. The method of claim **1**, further comprising:

concurrently sending multicast traffic from a multicast source device to a plurality of destination devices on each P2MP LSP of the plurality of P2MP LSPs.

3. The method of claim **1**, wherein the traffic-engineering constraint comprises one or more of an amount of bandwidth, link color, priority, and class type.

4. The method of claim **1**, wherein the plurality of maximally redundant trees comprises a pair of spanning trees that share a least number of links possible and share a least number of nodes possible.

5. The method of claim **4**, wherein the plurality of maximally redundant trees comprises a pair of spanning trees that are link disjoint and node disjoint.

6. The method of claim **1**, wherein a maximally redundant tree of the plurality of maximally redundant trees comprises a plurality of branches, and wherein establishing the plurality of P2MP LSPs along the maximally redundant trees comprises establishing the plurality of P2MP LSPs along a subset of the plurality of branches of the maximally redundant tree.

7. The method of claim **1**, further comprising, by the ingress network device, periodically re-calculating the plurality of maximally redundant trees to determine whether a more optimal plurality of maximally redundant trees exists on the network graph.

8. The method of claim **1**, further comprising:

detecting a change to a network topology yielding a modified network graph; and

with the ingress network device, automatically re-calculating the plurality of maximally redundant trees based on the modified network graph.

9. The method of claim **1**, further comprising:

prior to calculating the plurality of maximally redundant trees on the network graph, modifying the network graph to remove links from the network graph that do not satisfy the traffic-engineering constraint to obtain a modified network graph, wherein calculating the plurality of maximally redundant trees based on the network graph comprises calculating the plurality of maximally redundant trees based on the modified network graph.

10. The method of claim **1**, further comprising using a resource reservation protocol to establish the P2MP LSPs.

11. The method of claim **10**, wherein the resource reservation protocol comprises the Resource Reservation Protocol with Traffic Engineering extensions (RSVP-TE).

12. The method of claim **1**, wherein establishing the plurality of P2MP LSPs comprises:

sending resource reservation requests to label-switching routers (LSRs) along each of the maximally redundant trees, wherein the resource reservation requests each include an identifier associating the requests with the respective maximally redundant tree; and

receiving resource reservation messages in response to the resource reservation requests that specify reserved resources and labels allocated to the respective one of the plurality of P2MP LSPs to be used for forwarding network traffic to corresponding next hops, wherein the resource reservation messages each include an identifier associating the messages with the respective maximally redundant tree.

13. The method of claim **1**, wherein the traffic engineering constraint comprises one of:

a maximum link bandwidth for each of the one or more network links, wherein the maximum link bandwidth defines a maximum amount of bandwidth capacity associated with a network link;

a residual bandwidth for each of the one or more network links, wherein the residual bandwidth defines an amount of bandwidth capacity for a network link that is a maximum link bandwidth less a bandwidth of the network link reserved by operation of a resource reservation protocol; and

an available bandwidth for each of the one or more network links, wherein the available bandwidth defines an amount of bandwidth capacity for the network link that is neither reserved by operation of a resource reservation protocol nor currently being used by the first router to forward traffic using unreserved resources.

14. The method of claim **1**, wherein the ingress network device comprises the network device that computes the plurality of maximally redundant trees.

15. A network device comprising:

a processor;

a constrained maximally redundant tree module configured for execution by the processor to calculate a plurality of maximally redundant trees from the network device to a plurality of egress network devices based on a network graph, in which each of the plurality of maximally redundant trees comprises a spanning tree to the plurality of egress network devices rooted at the network device, wherein each of the maximally redundant trees is calculated to comprise a point to multipoint (P2MP) path from the network device to the plurality of egress network devices that is as disjoint as possible from a respective P2MP path from the network device to the plurality of egress network devices for each other one of the plurality of maximally redundant trees, and wherein the maximally redundant trees are calculated such that each link along each of the plurality of maximally redundant trees satisfies a specified traffic-engineering constraint,

wherein the constrained maximally redundant tree module is configured to, in response to determining that the plurality of maximally redundant trees includes at least one node whose removal partitions a network represented by the network graph:

modify the specified traffic-engineering constraint to have a less restrictive value;

modify the network graph to add links to the network graph that satisfy the modified traffic-engineering constraint to obtain a modified network graph; and

re-calculate at least a portion of the plurality of maximally redundant trees based on the modified network graph to obtain a plurality of maximally redundant trees in which at least two nodes must be removed before the network is partitioned; and

a resource reservation protocol module configured for execution by the processor to establish a plurality of P2MP label switched paths (LSPs) from the network device as an ingress network device to the plurality of egress network devices along each of the plurality of maximally redundant trees in which at least two nodes must be removed before the network is partitioned, wherein each of the P2MP LSPs corresponds to a different one of the plurality of maximally redundant trees in which at least two nodes must be removed before the network is partitioned.

16. The network device of claim **15**, further comprising a forwarding component that concurrently sends multicast traffic from a multicast source device to a plurality of destination devices on each P2MP LSP of the plurality of P2MP LSPs.

17. The network device of claim **15**, wherein the plurality of maximally redundant trees comprises a pair of spanning trees that share a least number of links possible and share a least number of nodes possible.

18. The network device of claim **15**, wherein the constrained maximally redundant tree module automatically recalculates the plurality of maximally redundant trees based on the modified network graph upon the network device detecting a change to a network topology yielding a modified network graph.

19. A non-transitory computer-readable storage medium comprising instructions for causing a programmable processor to:

calculate a plurality of maximally redundant trees from an ingress network device to a plurality of egress network devices based on a network graph, in which each of the plurality of maximally redundant trees comprises a spanning tree to the plurality of egress network devices rooted at the ingress network device, wherein each of the maximally redundant trees is calculated to comprise a point to multipoint (P2MP) path from the ingress network device to the plurality of egress network devices that is as disjoint as possible from a respective P2MP path from the ingress network device to the plurality of egress network devices for each other one of the plurality of maximally redundant trees, and wherein the maximally redundant trees are calculated such that each link along each of the plurality of maximally redundant trees satisfies a specified traffic-engineering constraint;

in response to determining that the plurality of maximally redundant trees includes at least one node whose removal partitions a network represented by the network graph:

modify the specified traffic-engineering constraint to have a less restrictive value;

modify the network graph to add links to the network graph that satisfy the modified traffic-engineering constraint to obtain a modified network graph; and

re-calculate at least a portion of the plurality of maximally redundant trees based on the modified network graph to obtain a plurality of maximally redundant trees in which at least two nodes must be removed before the network is partitioned; and

establish a plurality of P2MP label switched paths (LSPs) from the ingress network device to the plurality of egress network devices along each of the plurality of maximally redundant trees, wherein each of the P2MP LSPs corresponds to a different one of the plurality of maximally redundant trees.

20. The method of claim **1**, further comprising repeatedly modifying the specified traffic-engineering constraint, modifying the network graph, and re-calculating at least a portion of the plurality of maximally redundant trees until obtaining the plurality of maximally redundant trees in which at least two nodes must be removed before the network is partitioned.

21. The method of claim 1, wherein modifying the specified traffic-engineering constraint to have the less restrictive value comprises modifying the specified traffic-engineering constraint for only a certain part of the maximally redundant trees that includes the at least one node whose removal partitions the network.

* * * * *