



- (51) International Patent Classification:  
H04L 12/761 (2013.01) H04L 12/803 (2013.01)  
H04L 12/721 (2013.01)
- (21) International Application Number:  
PCT/US2016/057383
- (22) International Filing Date:  
17 October 2016 (17.10.2016)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
62/250,820 4 November 2015 (04.11.2015) US  
15/187,415 20 June 2016 (20.06.2016) US
- (71) Applicant (for all designated States except DJ, KW, US):  
**HUAWEI TECHNOLOGIES CO., LTD.** [CN/CN];  
Huawei Administration Building, Bantian, Longgang Dis-  
trict, Guangdong, Shenzhen, Guangdong 518129 (CN).
- (71) Applicant (for US only): **FUTUREWEI TECHNOLO-  
GIES, INC.** [US/US]; 5340 Legacy Drive, Suite 175, Pla-  
no, Texas 75024 (US).
- (72) Inventors: **MCCORMICK, William Carson**; 6 Redenda  
Cr, Ottawa, Ontario K2G 0N6 (CA). **ASH-  
WOOD-SMITH, Peter**; 20 Des Genevriers, Gatineau,  
Québec J9A 2V8 (CA). **KELLY, Francis Patrick**; 23  
Lyndewode Road, Cambridge Cambridgeshire CB1 2HN  
(GB).

(74) Agent: **WIEBUSCH, Landon E.**; Slater Matsil, LLP,  
17950 Preston Road, Suite 1000, Dallas, Texas 75252  
(US).

(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,  
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,  
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,  
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,  
KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,  
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,  
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,  
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,  
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,  
ZW.

(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,  
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,  
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,  
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,  
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,  
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the  
claims and to be republished in the event of receipt of  
amendments (Rule 48.2(h))

(54) Title: APPARATUS AND METHOD FOR NETWORK FLOW SCHEDULING

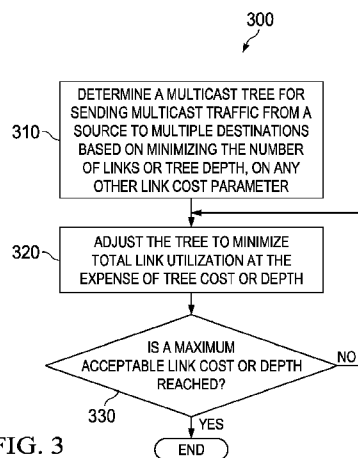


FIG. 3

(57) Abstract: Embodiments are provided for path flow scheduling of multicast traffic through a network. The paths for traffic flow are determined to optimize link utilization in terms of bandwidth and link capacity, and limit link cost. In an embodiment, a method is implemented for network flow scheduling. The method includes establishing, by a controller of a network, a multicast tree which includes a plurality of links for sending multicast traffic from a source to multiple destinations. The tree is established based on minimizing a number of links in the multicast tree. The tree is then adjusted by replacing one or more of the plurality of links to reduce the link utilization. The tree adjustment is repeated by further replacing one or more links in the multicast tree to further reduce the link utilization.

WO 2017/078922 A1

## Apparatus and Method for Network Flow Scheduling

This patent application claims priority to U.S. Non-Provisional Patent Application No. 15/187,415, filed on June 20, 2016 and entitled "Apparatus and Method for Network Flow Scheduling," which claims priority to U.S. Provisional Application No. 62/250,820  
5 filed on November 4, 2015 and entitled "System and Method for Network Flow Scheduling," which are hereby incorporated herein by reference as if reproduced in its entirety.

### **TECHNICAL FIELD**

The present invention relates to networking and telecommunications, and, in  
10 particular embodiments, to an apparatus and method for network flow scheduling.

### **BACKGROUND**

Generally, wired and wireless internet communications are provided by connections via one or more links between network nodes. These connections can be wired or wireless, and can include both access network and backhaul connections. The  
15 nodes connected can include wireless access points such as base stations, user equipment, routers, optical devices and other core network nodes. The links are used to provide an end-to-end communications connection. A mechanism such as multiprotocol label switching (MPLS) may be used to direct data from one network node to the next based on short path labels that identify virtual paths between nodes, rather than the endpoints.  
20 Shortest path flow scheduling, however, tends to congest the shortest path links. Therefore it may reduce network throughput for best effort traffic, and it may create congested network spots that require operator attention. It also may block new flow establishment. An efficient network flow scheduling scheme, which could avoid or limit such unwanted situations and could improve network link utilization, is thus needed.

### **SUMMARY OF THE INVENTION**

In accordance with an embodiment, a method for network flow scheduling is provided. The method includes establishing a multicast tree that includes a plurality of links for sending multicast traffic from a source to multiple destinations based on minimizing a number of links in the multicast tree. The multicast tree has a link  
30 utilization and a tree cost for the plurality of links. The method further includes evaluating the tree cost of the multicast tree in accordance with an acceptable tree cost requirement, and adjusting the multicast tree by replacing one or more of the plurality of links in accordance with a link utilization objective. The method further includes re-

evaluating the tree cost of the adjusted multicast tree, and re-adjusting the evaluated multicast tree until the acceptable tree cost requirement is satisfied. An apparatus for performing this method is also provided.

5 In accordance with another embodiment, another method for network flow scheduling is provided. The method includes establishing a multicast tree that includes a plurality of links for sending multicast traffic from a source to multiple destinations based on minimizing a tree depth of the multicast tree and according to a link capacity constraint, and replacing one or more of the links to reduce link utilization on the multicast tree until reaching or exceeding a maximum tree depth allowed for the multicast  
10 tree. The link utilization represents a total bandwidth used for sending the multicast traffic on the links to a total capacity of the links. An apparatus for performing this method is also provided.

The foregoing has outlined rather broadly the features of an embodiment of the present invention in order that the detailed description of the invention that follows may  
15 be better understood. Additional features and advantages of embodiments of the invention will be described hereinafter, which form the subject of the claims of the invention. It should be appreciated by those skilled in the art that the conception and specific embodiments disclosed may be readily utilized as a basis for modifying or designing other structures or processes for carrying out the same purposes of the present  
20 invention. It should also be realized by those skilled in the art that such equivalent constructions do not depart from the spirit and scope of the invention as set forth in the appended claims.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

For a more complete understanding of the present invention, and the advantages  
25 thereof, reference is now made to the following descriptions taken in conjunction with the accompanying drawing, in which:

Figures 1A, 1B, and 1C show an embodiment of implementing network flow scheduling to improve link utilization;

Figure 2 shows a cumulative distribution function comparison for network flow  
30 scheduling;

Figure 3 shows an embodiment method for implementing network flow scheduling to improve link utilization; and

Figure 4 is a diagram of a processing system that can be used to implement various embodiments.

Corresponding numerals and symbols in the different figures generally refer to corresponding parts unless otherwise indicated. The figures are drawn to clearly illustrate the relevant aspects of the embodiments and are not necessarily drawn to scale.

#### **DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS**

5           The making and using of the presently preferred embodiments are discussed in detail below. It should be appreciated, however, that the present invention provides many applicable inventive concepts that can be embodied in a wide variety of specific contexts. The specific embodiments discussed are merely illustrative of specific ways to make and use the invention, and do not limit the scope of the invention.

10           Various Internet based applications require flow and path scheduling throughout a network infrastructure. For example, in a network supporting Internet Protocol Television (IPTV) for digital terrestrial television (DTTV), ad hoc services and on demand multicast, and best effort services, information is delivered to transmit stations or to homes in the form of packets through a network. Typically, the packet traffic is sent from a source,  
15 e.g., a server where the programming data is stored, to multiple end or destination points. Due to the nature of the networks in question, different flows may be placed on different paths. The multiple end points may correspond to end users (e.g., at homes or businesses) or may correspond to wireless transmit stations.

          System and method embodiments are provided herein for path flow scheduling of  
20 multicast traffic through a network. The paths for traffic flow are determined to optimize link utilization in terms of bandwidth and link capacity, and limit link cost. This scheme provides improved link utilization compared to other schemes for multicast traffic flow scheduling, such as shortest path flow scheduling, which are based primarily on link cost, e.g., the total number of links or the depth of the multicast tree. In an embodiment, the  
25 multicast tree paths allocated for traffic flow are determined by minimizing a function of link utilization and tree depth. The link utilization can be measured as a ratio of the bandwidth needed for the traffic to the available link capacity, as described further below. This scheme can be used to rout IPTV packets for DTTV, ad hoc traffic for video applications, or best effort traffic, more efficiently to avoid congestion spots or bottleneck  
30 links, to avoid reserved or selected links, or both.

          Figures 1A, 1B and 1C illustrate steps for network flow scheduling to improve (minimize) link utilization in a network. A multipath tree is built to send traffic form a source point (the tree root) to multiple points (the tree leaf, or terminal, nodes) with the objective of minimizing link utilization while limiting tree depth. The link utilization is

defined as the percentage or ratio of bandwidth used for directing the traffic on selected links to the total capacity on the selected links. This is achieved by first building a minimum depth tree, as shown in Figure 1A. The minimum tree depth in this example is equal to 2 where each one of the tree leaf nodes 121, 122, and 123 is separated from the root 101 by one of the intermediate nodes 111, 112, and 113. The resulting optimal depth tree contains two hot links: a first hot link 131 between root 101 and intermediate node 113; and a second hot link 132 between intermediate node 112 and leaf node 122. A hot link is a link of which bandwidth usage has exceeded a determined threshold of its total capacity. Therefore, the utilization of such links is high.

Next, the utilization of the hot links for traffic flow is minimized at the expense of relaxing the minimum tree depth. As shown in Figure 1B, traffic is routed around the hot link 131 from root 101 to intermediate node 113 through the links 141 and 142. This increases the tree depth to 3 but avoids overburdening links (which may be described as a reduction in total link utilization). Figure 1C shows further improved link utilization by rerouting the traffic around hot link 132 from intermediate node 112 to leaf node 122 through the links 141, 142 and 143. Thus, link cost in term of depth is increased to 4.

The scheme above is used to optimize link utilization by compromising minimum link depth. Increasing the link depth of the distribution tree may increase total link length for routing traffic. However, the resulting minimum link utilization tree avoids bottlenecks or hot links which can compensate for delay created with increasing the total link length. The same scheme can also be used to avoid selected links, which are herein referred to as cold links. A cold link is a link that at the time at which the multipath tree is computed carries a volume of data below a threshold. In some networks, a link may be based on leased capacity. If no traffic is transmitted on the link, there may be no fee for usage. In other examples, charging for the leased link may be done based on volume tiers such that the rates charged for traffic are associated with thresholds such that the cost of transmitting a block of data increases as the traffic levels on the link increase. In these examples, in addition to (or in place of) avoiding placing traffic on hot links, it may be beneficial to avoid putting traffic on cold links (or at least avoid putting certain volumes of traffic on cold links). For example, the tree can be used to multicast best effort traffic, avoiding cold links that if used would result in an increase in the cost of operating the network. Thus, the risk of new flow blocking is reduced. In another embodiment, the multicast tree is initially built to minimize a link cost rather than tree depth. For example, each link is assigned a cost value in terms of charge of usage, bandwidth or any other cost

target. Subsequently, link utilization is minimized by compromising the minimum link cost. It should be noted that it is also possible to create a multipath tree that takes into account link capacity that has been committed to other traffic flows.

In an embodiment, the algorithm and objective (optimization) function for  
 5 determining the multicast tree for path flow scheduling can be established using a plurality of variables. The variables can include:  $f$ , the bandwidth requested for the multicast;  $l$ , the index for each considered link in the network;  $k$ , the index for each destination node in the multicast tree;  $C_l$ , the capacity of each considered link;  $c_l$ , the cost of each considered link;  $x_{lk}$ , the indicator variable that indicates if a link  $l$  is on the path  
 10 to destination node  $k$ ;  $y_l$ , the indicator variable that indicates if a link  $l$  is used in the multicast tree;  $d$ , the integer variable that tracks the depth of the tree;  $u$ , the real variable that tracks the maximum link utilization in the tree; and  $r_l$ , the indicator variable used to track candidates for second highest link utilization.

The core optimization algorithm can be embodied in a program (or a set of  
 15 programs) that is built around a set of constraints which force the algorithm solver to construct a multicast tree. These constraints can include flow conservation constraints, link usage constraints and link capacity constraints.

Flow conservation constraints force the optimizer to build a path from the source node to each destination node in the network. Each destination node can have a set of  
 20 constraints associated with it, including the following. Each egress node must have at least 1 incoming link and 0 outgoing links. Each ingress node must have at least 1 outgoing link and 0 incoming links. This constraint may be redundant with the above constraint. If constraints are determined to be redundant with other constraints, they may not need to be implemented. In some embodiments, nodes other than the ingress and  
 25 egress nodes have the same number of outgoing and incoming links. Additionally, the number of incoming/outgoing links is either 0 or 1.

Link usage constraints allow the solver to connect all the individual paths into a tree. For every link, the link usage constraint indicates whether the link is part of the  
 30 multicast tree or not. The constraint can be formulated as follows:  $y_l \geq x_{lk}, \forall k, \forall l$ . For this constraint to function properly, the objective function should be trying to minimize  $L(y)$  (where  $L(\bullet)$  is a linear function and  $y = [y_l]$ ). ]

Link capacity constraints can be used to ensure that the solver does not overcommit any links in the network. In this case, these constraints just ensure that each link has enough capacity to support the multicast.

In order to build a minimum depth tree, a variable  $d$  is used to track the depth of the tree. The tree depth is based on the path length to any of the destination nodes, so the

following constraint is also included:  $d \geq \sum_l x_{lk}, \forall k$ . The objective function attempts to minimize  $d$  for this constraint to function properly. This can be achieved by weighing the depth parameter more heavily than the total cost of links. In order to build a minimum depth tree, an objective function in the following form can be used: minimize

$vd + \sum_l c_l y_l$ . In this objective function, the variable  $d$  is scaled up in value by a weight

parameter,  $v$ , to permit minimizing the cost of all the links in the tree  $(\sum_l c_l y_l)$  as a secondary objective. The depth parameter should be scaled large enough to emphasize the minimization of  $d$  as objective over the total link cost. Different scaling factors can be used for different purposes depending on the application or network scenario. In another embodiment, the tree is built by minimizing individual metrics using a unicast to each egress node, and then combining the resulting unicasts into a multicast tree.

The second optimization objective to solve is to minimize the maximum link utilization of all the links used in the tree, given a maximum allowed tree depth. This is intended to route the tree around congested links when possible. In order to implement this objective, the utilization of the links in the tree is tracked by combining the  $y_l$  variables with the bandwidth and link capacity:  $u \geq y_l \frac{f}{C_l}, \forall l$ .

$$(1)$$

The objective function to minimize the maximum link utilization can be formulated as follows: minimize  $wu + \sum_l c_l y_l$ . In the second objective function, the variable  $u$  is scaled up in value by a weight parameter,  $w$ , to permit minimizing the total link cost in the tree as a secondary objective. The link utilization parameter should be scaled large enough to emphasize the minimization of  $u$  as objective over the total link cost. The link utilization in equation (1) is an approximation of the link utilization. The actual link capacity may be  $C$ , but the link capacity available to the solver is  $C' = C - b$ , where  $b$  is some part of the link capacity that has already been committed. The real link

utilization would be  $(f + b)/C$ . The constraint may then be better represented as:

$$u \geq y_l \frac{f + b}{C}, \forall l.$$

The above describes a method to route traffic around congested links by using an objective function that minimizes the total utilization of all the links used in the tree (subject to given constraints). An alternative method is to construct an objective function that minimizes the second highest link utilization in the following manner. A variable  $u_2$  is defined as the second highest link utilization of the links used by the flow. A new set of binary variables  $\{r_l\}$  - one per link - is defined. The variables in the set are used to count the number of links where  $u_2$  is greater than the link utilization. The following constraint is used to establish a value for  $r_l$ :  $r_l > u_2 - \frac{f + b}{C}$ . Because  $r_l$  is either 0 or 1, if  $u_2 > r_l \frac{f + b}{C}$ ,  $r_l$  will be forced to 1. If  $u_2 \leq r_l \frac{f + b}{C}$ , then  $r_l$  will be forced to 0. If  $u_2 - \frac{f + b}{C}$  is greater than 1, then solving the problem may become infeasible. To avoid this situation, the problem should be scaled properly to constrain the link utilization to the interval  $[0, 1]$ .

The final constraint is to constrain the number of non-zero instances of  $r_l$ :

$$\sum_l r_l = \sum_l y_l - 1.$$

This constraint should force  $u_2$  to be equal to the second highest link

utilization, provided that the objective function is trying to minimize  $u_2$  in some manner. This approach may be extended to compute the third, fourth, and other higher order highest link utilizations if necessary.

The above description maintains the status quo for all of the flows except the next one, which reduces link congestion. An alternative approach is to route all flows, including the next one, using this technique. By leaving existing flows with their status quo assignments, an overall suboptimal distribution pattern may be arrived at, but the operational difficulties associated with rerouting already placed flows are avoided.

In an embodiment, the flow scheduling scheme for minimizing link utilization can be implemented as an algorithm using mixed integer linear programming (MILP). Rescheduling existing flows that are currently causing (or are subject to) congestion on a given link may be performed using techniques such as a multiprotocol label switching (MPLS) hitless reroute. Those skilled in the art will appreciate that Software Defined Networking (SDN) controllers could be used in another embodiment to provide the rerouting, and can further be used in a hitless manner as well.



In further embodiments, the algorithm can be adjusted to predict and avoid links that are expected to be congested in the future based on recurrence patterns (e.g., every Sunday), based on calendared events (e.g., a football schedule where a football game is expected to cause congestion on certain links), or statistics patterns of previous similar events. For instance, machine learning techniques can be used to take a recurrence pattern/calendar as input and generate link congestion data as output. Historical data can be recorded so that it may subsequently be used to estimate link utilization at a specific time. Routing may then be based on anticipation that certain links are going to be congested.

Figure 2 shows a graph for a cumulative distribution function comparison for network flow scheduling. The values shown are calculated for a study case of a multicast tree scenario. The resulting data is shown to illustrate a pattern and advantage of the network flow scheduling scheme described above. The vertical axis in the graph represents the normalized link count, while the horizontal axis represents link utilization in the interval  $[0, 1]$ . Curve 210 shows the cumulative distribution of data scheduled in a multicast tree based on minimum tree depth. Curve 220 shows the cumulative distribution of the data after rescheduling to minimize link utilization at the expense of tree depth, as described above. The cumulative distribution reflects the percentage of links (normalized link count) at or below a normalized utilization value. As can be seen, curve 220 has lower link utilization on the more congested links, i.e., has a lower link count at high utilization values.

Figure 3 shows an embodiment method 300 for implementing network flow scheduling to improve link utilization. The method 300 may be implemented by a scheduler, controller or network management entity in the form of algorithm steps using MILP or any suitable machine programming technique. As described above, the method 300 can be used for routing multicast traffic of high priority such as traffic for ad hoc events or best effort traffic. The method 300 can also be implemented for any suitable networks including MPLS and SDN networks.

At block 310, a multicast tree is determined for sending multicast traffic from a source to multiple destinations based on minimizing the number of links or tree depth, on any other link cost parameter. The tree determination could be subject to constraints such as the conservation of all considered flows, link capacity, and link utilization. At block 320, the tree is adjusted to minimize total link utilization at the expense of tree cost or depth. The tree cost or depth is relaxed by a determined amount, for instance an integer

value of 1, to allow further reduction of the link utilization. This step allows hot links avoidance where link utilization due to congestion is typically high. Similarly, cold links can be avoided by assigning high utilization values to such links. At step 330, the method determines whether a maximum acceptable link cost or depth is reached. If this condition  
5 is true, then the method ends and the resulting adjusted tree is used for scheduling the traffic. Otherwise, step 320 is repeated to further reduce total link utilization at the expense of increase in link cost or depth.

Figure 4 is a block diagram of a processing system 400 that can be used to implement various embodiments including the methods above. For instance, the  
10 processing system 400 can be part of a solver entity, e.g., a network controller or router, configured to solve the network flow scheduling problem to determine the multicast tree. Specific devices may utilize all of the components shown, or only a subset of the components, and levels of integration may vary from device to device. Furthermore, a device may contain multiple instances of a component, such as multiple processing units,  
15 processors, memories, transmitters, receivers, etc. The processing system 400 may comprise a processing unit 401 equipped with one or more input/output devices, such as a speaker, microphone, mouse, touchscreen, keypad, keyboard, printer, display, and the like. The input/output devices also include a camera for capturing digital picture or video. The processing unit 401 may include a central processing unit (CPU) 410, a memory 420,  
20 a mass storage device 430, a video adapter 440, and an I/O interface 460 connected to a bus. The bus may be one or more of any type of several bus architectures including a memory bus or memory controller, a peripheral bus, a video bus, or the like.

The CPU 410 may comprise any type of electronic data processor. The memory  
25 420 may comprise any type of system memory such as static random access memory (SRAM), dynamic random access memory (DRAM), synchronous DRAM (SDRAM), read-only memory (ROM), a combination thereof, or the like. In an embodiment, the memory 420 may include ROM for use at boot-up, and DRAM for program and data storage for use while executing programs. In embodiments, the memory 420 is non-transitory. The mass storage device 430 may comprise any type of storage device  
30 configured to store data, programs, and other information and to make the data, programs, and other information accessible via the bus. The mass storage device 430 may comprise, for example, one or more of a solid state drive, hard disk drive, a magnetic disk drive, an optical disk drive, or the like.

The video adapter 440 and the I/O interface 460 provide interfaces to couple external input and output devices to the processing unit. As illustrated, examples of input and output devices include a display or touchscreen 490 coupled to the video adapter 440 and any combination of camera/keyboard/mouse 470 coupled to the I/O interface 460.

5 Other devices may be coupled to the processing unit 401, and additional or fewer interface cards may be utilized. For example, a serial interface card (not shown) may be used to provide a serial interface for a printer.

The processing unit 401 also includes one or more network interfaces 450, which may comprise wired links, such as an Ethernet cable or the like, and/or wireless links to  
10 access nodes or one or more networks 480. The network interface 450 allows the processing unit 401 to communicate with remote units via the networks 480. For example, the network interface 450 may provide wireless communication via one or more transmitters/transmit antennas and one or more receivers/receive antennas. In an embodiment, the processing unit 401 is coupled to a local-area network or a wide-area  
15 network for data processing and communications with remote devices, such as other processing units, the Internet, remote storage facilities, or the like.

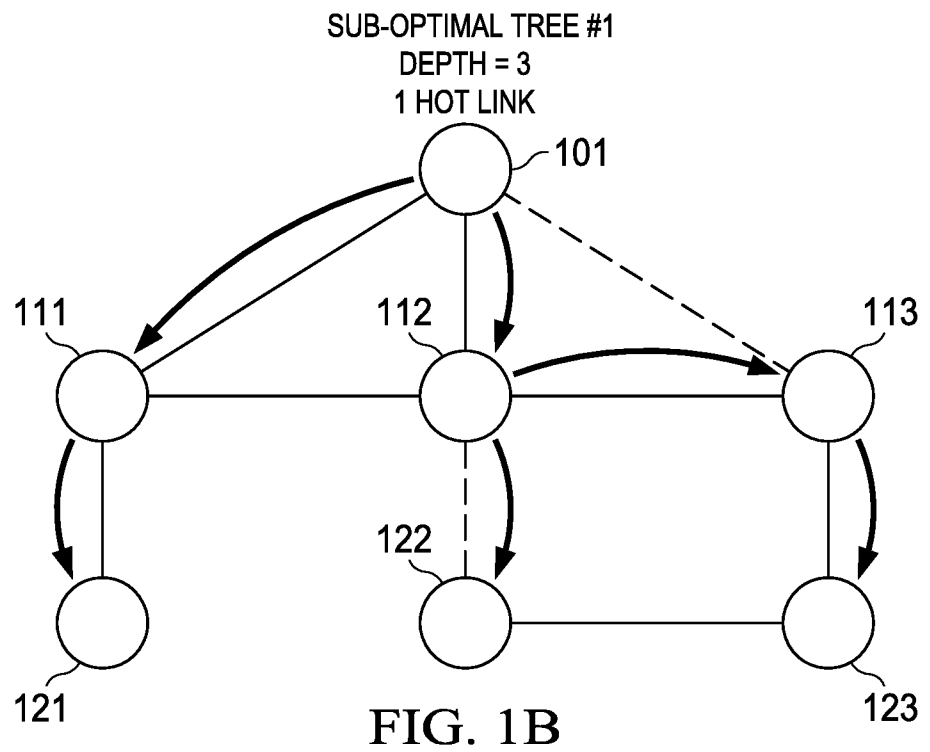
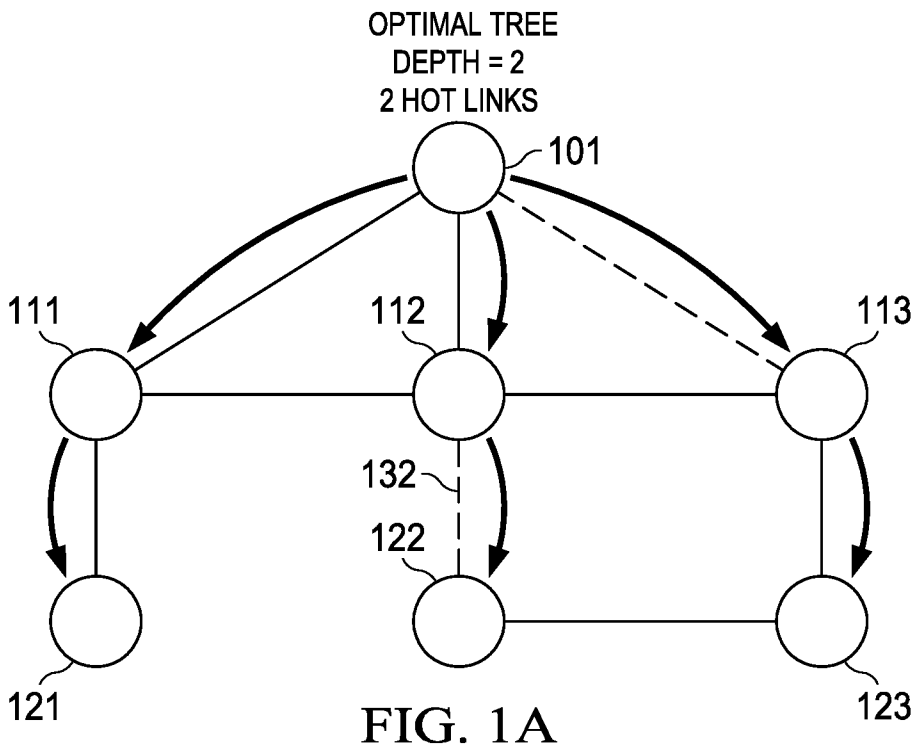
While several embodiments have been provided in the present disclosure, it should be understood that the disclosed systems and methods might be embodied in many other specific forms without departing from the spirit or scope of the present disclosure.  
20 The present examples are to be considered as illustrative and not restrictive, and the intention is not to be limited to the details given herein. For example, the various elements or components may be combined or integrated in another system or certain features may be omitted, or not implemented.

In addition, techniques, systems, subsystems, and methods described and  
25 illustrated in the various embodiments as discrete or separate may be combined or integrated with other systems, modules, techniques, or methods without departing from the scope of the present disclosure. Other items shown or discussed as coupled or directly coupled or communicating with each other may be indirectly coupled or communicating through some interface, device, or intermediate component whether electrically,  
30 mechanically, or otherwise. Other examples of changes, substitutions, and alterations are ascertainable by one skilled in the art and could be made without departing from the spirit and scope disclosed herein.

**WHAT IS CLAIMED IS:**

1. A method for network flow scheduling, the method comprising:
  - establishing, by a controller of a network, a multicast tree, including a plurality of links for sending multicast traffic from a source to multiple destinations, based on minimizing a number of links in the multicast tree, the multicast tree having a link utilization and a tree cost for the plurality of links;
  - evaluating the tree cost of the multicast tree in accordance with an acceptable tree cost requirement;
  - adjusting the multicast tree by replacing one or more of the plurality of links in accordance with a link utilization objective; and
  - re-evaluating the tree cost of the adjusted multicast tree, and re-adjusting the evaluated multicast tree until the acceptable tree cost requirement is satisfied.
2. The method of claim 1, wherein step of adjusting the multicast tree includes replacing a link in the plurality of links to reduce the link utilization.
3. The method of claim 1, wherein the step of adjusting the multicast tree includes replacing a link in the plurality of links to avoid placing traffic on a hot link.
4. The method of claim 1 wherein the step of adjusting the multicast tree includes replacing a link in the plurality of links to avoid placing traffic on a cold link.
5. The method of claim 1, wherein the tree cost is a measure of tree depth or a number of links in the multicast tree.
6. The method of claim 1, wherein the link utilization is a ratio of bandwidth for sending the multicast traffic on the links to a total capacity for routing traffic on the links.
7. The method of claim 1, wherein the multicast tree is established and adjusted using mixed integer linear programming (MILP).
8. The method of claim 1, wherein establishing the multicast tree includes:
  - building a plurality of unicast paths from the source to the multiple destinations in accordance with minimizing the number of links in the multicast tree; and
  - combining the unicast paths into a multipath tree.
9. The method of claim 1, wherein adjusting the multicast tree increases the tree cost of the multicast tree.
10. A method for network flow scheduling, the method comprising:
  - establishing, by a controller of a network, a multicast tree including a plurality of links for sending multicast traffic from a source to multiple destinations based on minimizing a tree depth of the multicast tree and according to a link capacity constraint; and
  - icast tree, until reaching or exceeding a maximum tree depth allowed for the multicast tree, the link utilization representing a total bandwidth used for sending the multicast traffic on the links to a total capacity of the links.

11. The method of claim 10, wherein the link capacity constraint ensures that each link in the multicast tree has enough capacity to support the multicast traffic.
12. The method of claim 10, wherein the network is a multiprotocol label switching (MPLS) network.
13. The method of claim 10, wherein the network is a Software Defined Networking (SDN) network.
14. The method of claim 10, wherein the multicast traffic includes Internet Protocol Television (IPTV) packets for digital terrestrial television (DTTV), ad hoc traffic for video applications, or best effort traffic.
15. A network entity for scheduling traffic flow, the network entity comprising:
  - at least one processor coupled to a memory; and
  - a non-transitory computer readable storage medium storing programming for execution by the at least one processor, the programming including instructions to:
    - establish a multicast tree, including a plurality of links for sending multicast traffic from a source to multiple destinations, based on minimizing a number of links in the multicast tree, the multicast tree having a link utilization and a tree cost for the plurality of links;
    - evaluate the tree cost of the multicast tree in accordance with an acceptable tree cost requirement;
    - adjust the multicast tree by replacing one or more of the plurality of links in accordance with a link utilization objective; and
    - re-evaluate the tree cost of the adjusted multicast tree, and re-adjust the evaluated multicast tree until the acceptable tree cost requirement is satisfied.
16. The network entity of claim 15, wherein the tree cost is a measure of tree depth or a number of links in the multicast tree.
17. The network entity of claim 15, wherein the link utilization is a ratio of bandwidth for sending the multicast traffic on the links to a total capacity for routing traffic on the links.
18. The network entity of claim 15, wherein the programming is mixed integer linear programming (MILP).
19. The network entity of claim 15, wherein the network entity is a Software Defined Networking (SDN) controller.
20. The network entity of claim 15, wherein the multiple destinations are transmit stations for wireless communications.



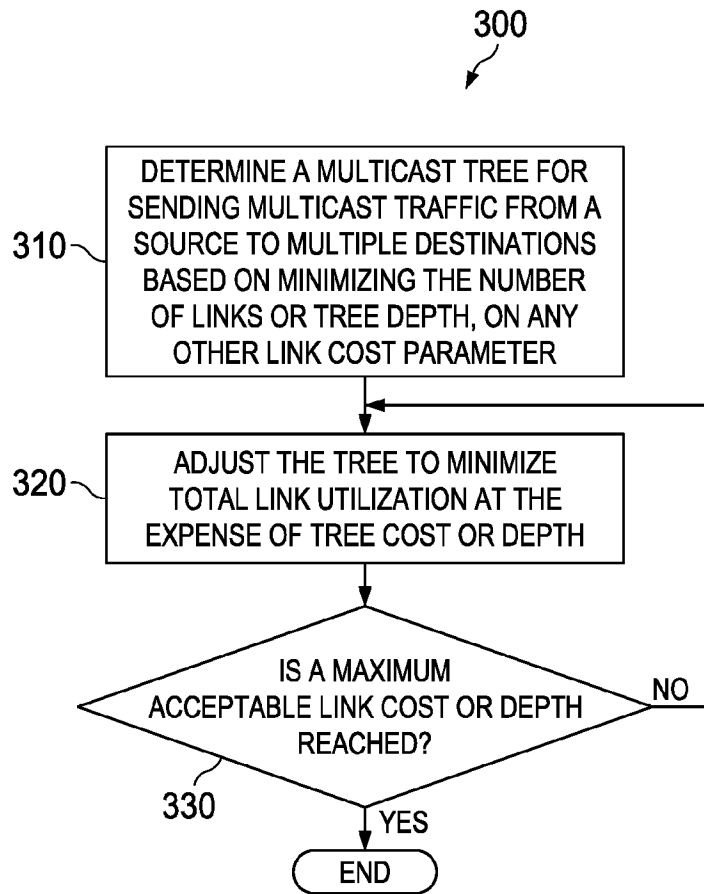
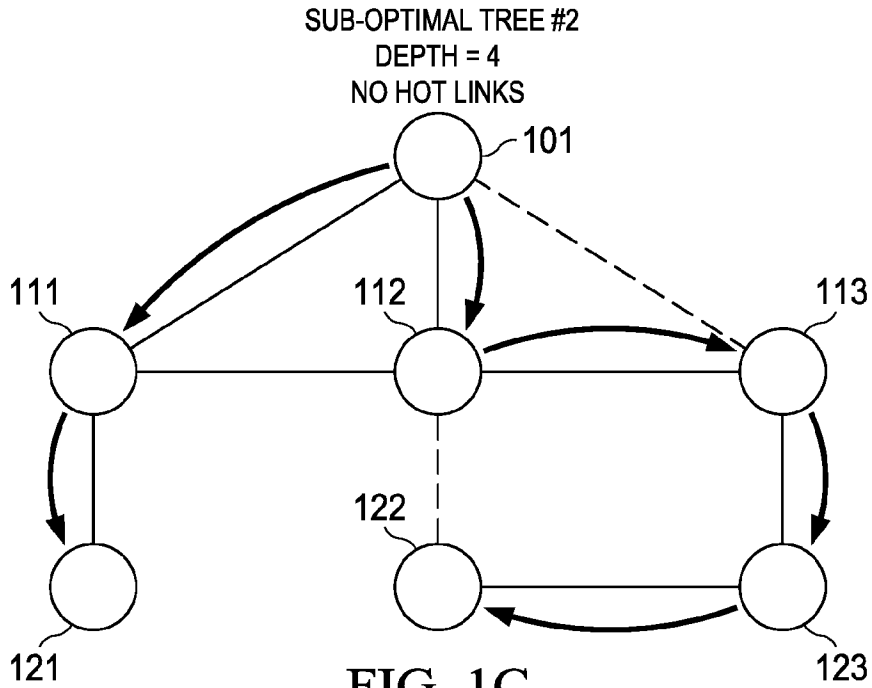
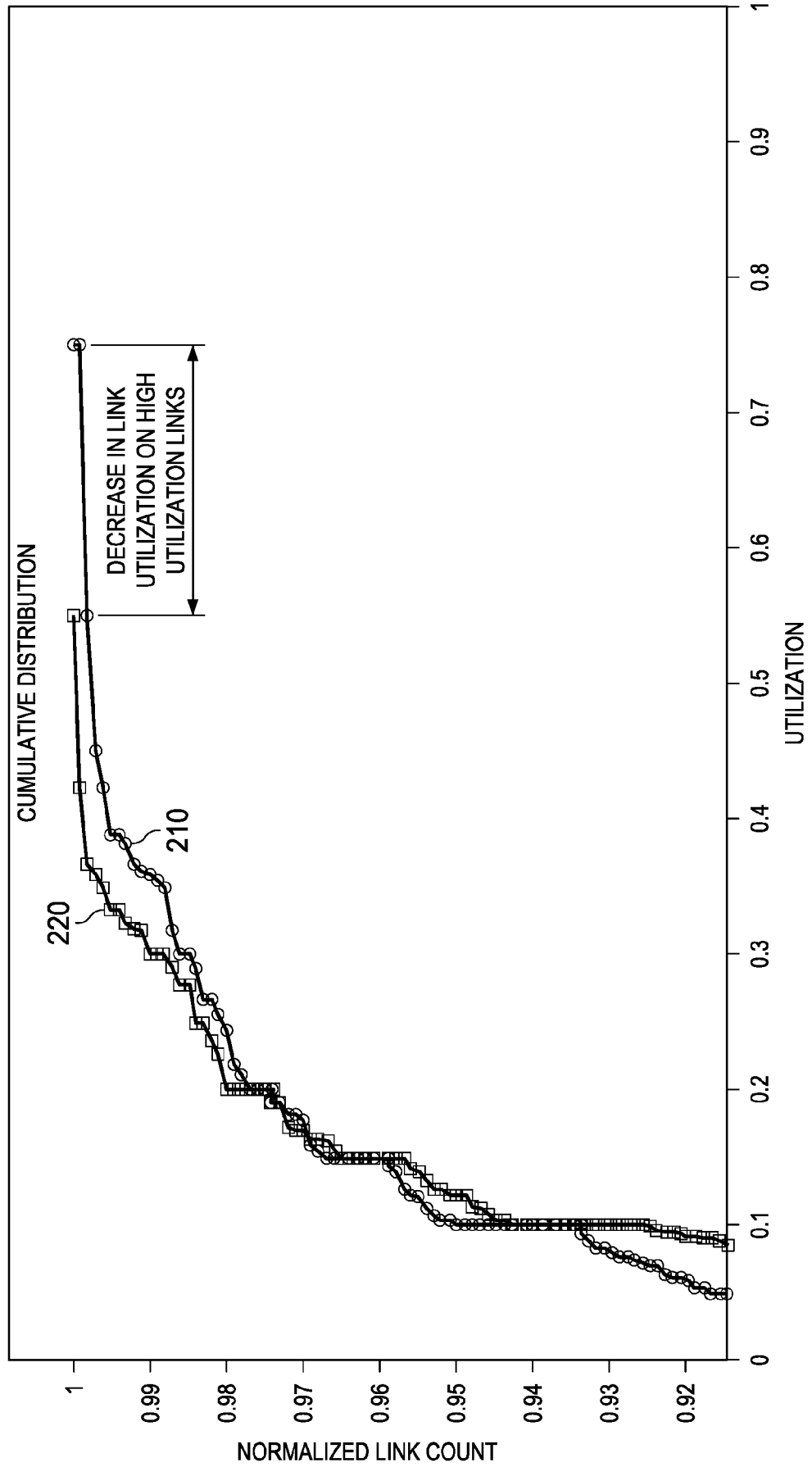


FIG. 3

FIG. 2





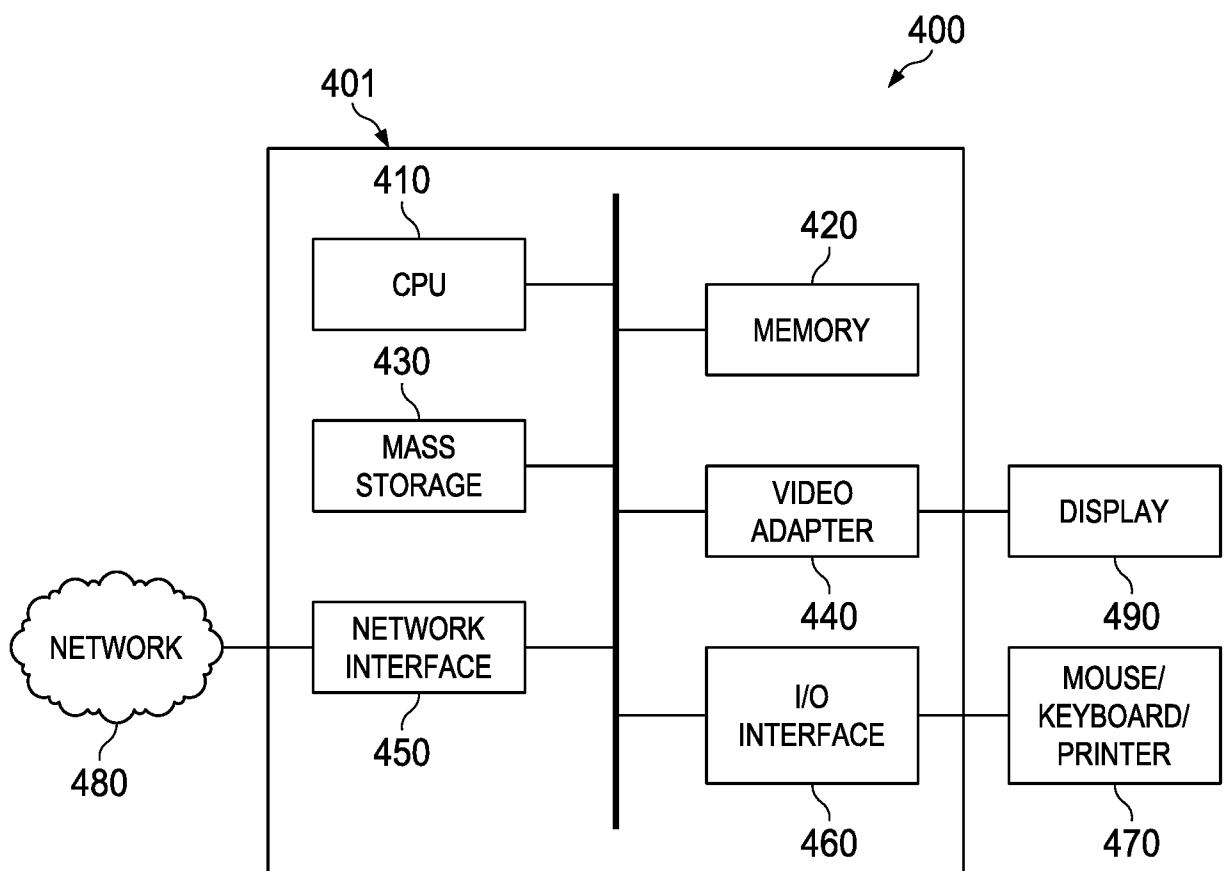


FIG. 4

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/57383

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC(8) - H04L 12/761, 12/721, 12/803 (2017.01) CPC - H04L 45/16, 45/20, 45/48, 12/18, 47/15; H04W 28/02 According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC(8): H04L 12/761, 12/721, 12/803 (2017.01) CPC: H04L 45/16, 45/20, 45/48, 12/18, 47/15; H04W 28/02 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); Google Scholar; EBSCO; IEEE Keywords: Multicast Multipoint Tree Matrix Mesh Cost Expense Price Link Connection Hop Target Acceptable Objective Budget Minimum Lowest Iterate Repeat Loop Replace Modify		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2010/0208733 A1 (ZHAO, Q et al.) 19 August 19 2010, abstract, figures 1, 6, 7, 8, paragraphs [0004], [0005], [0033], [0043], [0061], [0065]-[0072], [0101].	1, 15
--		-----
Y		2-3, 5-6, 9, 16-17, 19-20
--		-----
A		4, 8
Y	US 2015/0062285 A1 (FUTUREWEI TECHNOLOIES INC.) 5 March 2015, abstract, paragraphs [0003], [0008], [0064].	2, 3, 19
Y	US 2014/0147120 A1 (NEC LABORATORIES AMERICA, INC.) 29 May 2014, paragraph [0068].	5, 16
Y	US 2013/0191829 A1 (SHIMOKAWA, I et al.) 25 July 2013, paragraphs [0136], [0138].	6, 17
Y	US 2011/0273990 A1 (RAJAGOPALAN, S et al.) 10 November 2011, figures 3, 4, paragraphs [0035], [0036].	9
Y	US 2008/0198865 A1 (RUDNICK, W et al.) 21 August 2008, paragraph [0003].	20
A	US 2014/0133302 A1 (FU, S et al.) 15 May 2014, entire document.	1-9, 15-20
A	US 2011/0211517 A1 to (MOSCIBRODA, T et al.) 1 September 2011, entire document.	1-9, 15-20
A	US 6,456,597 B1 (BARE, B) 24 September 2002, entire document.	1-9, 15-20
A	US 2015/0244647 A1 (INFINERA CORPORATION) 27 August 2015, entire document.	1-9, 15-20
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
14 February 2017 (14.02.2017)		10 MAR 2017
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300		Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/57383

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3.  Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:  
See extra sheet.

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2.  As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-9, 15-20

- Remark on Protest**
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
  - The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
  - No protest accompanied the payment of additional search fees.

-\*\*\*-Continued from Box No. III Observations where unity of invention is lacking -\*\*\*-

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fee must be paid.

Group I: Claims 1-9 and 15-20 appear to be directed towards adjusting a multicast tree by replacing links within the tree to satisfy an acceptable tree cost requirement.

Group II: Claims 10-14 appear to be directed towards adjusting a multicast tree by replacing links within the tree to reach an allowed tree depth and bandwidth utilization of the plurality of links.

The inventions listed as Groups I-II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical features of Group I include at least, establishing a multicast tree based on minimizing a number of links in the multicast tree, the multicast tree having a link utilization and a tree cost for the plurality of links; evaluating the tree cost of the multicast tree in accordance with an acceptable tree cost requirement; adjusting the multicast tree by replacing one or more of the plurality of links in accordance with a link utilization objective; and re-evaluating the tree cost of the adjusted multicast tree, and re-adjusting the evaluated multicast tree until the acceptable tree cost requirement is satisfied, which are not present in Group II.

The special technical features of Group II include at least, establishing a multicast tree based on minimizing a tree depth of the multicast tree and according to a link capacity constraint; and replacing one or more of the links to reduce link utilization on the multicast tree, until reaching or exceeding a maximum tree depth allowed for the multicast tree, the link utilization representing a total bandwidth used for sending the multicast traffic on the links to a total capacity of the links, which are not present in Group I.

The common technical features shared by Groups I-II are establishing by a controller a multicast tree in a network including a plurality of links for sending multicast traffic between a source and multiple destinations; and replacing one or more links.

However, these common features are previously disclosed by US 2010/0208733 A1 to ZHAO, Q et al. (hereinafter "Zhao"). Zhao discloses establishing by a controller a multicast tree in a network including a plurality of links for sending multicast traffic between a source and multiple destinations (a path computation element, PCE 130 (controller) computes (establishes) a core tree (multicast tree) to allow a source node 302 to communicate with a plurality of (multicast traffic) destinations 308 in a network using multiple sub-paths (plurality of links); abstract, figures 6, 7, paragraphs [0033], [0061]); and replacing one or more of the links (merging (replacing) the plurality of sub-paths to minimize number of hops (link utilization objective); paragraphs [0065], [0068]).

Since the common technical features are previously disclosed by the Zhao reference, these common features are not special and so Groups I-II lack unity.