Mapping sub-flows to p2mp LSPs

Fernando Solano, Ramon Fabregat
Institut d'Informàtica i Aplicacions
Universitat de Girona,
Girona, Spain.
{fsolanod,ramon}@eia.udg.es

Yezid Donoso
Computer Science Department
Universidad del Norte,
Barranquilla, Colombia.
ydonoso@uninorte.edu.co

Jose L. Marzo
Institut d'Informàtica i Aplicacions
Universitat de Girona,
Girona, Spain.
marzo@eia.udg.es

***Abstract-* In previous work we proposed a multi-objective traffic engineering scheme (MHDB-S model) using different distribution trees to multicast several flows. In this paper, we propose a heuristic algorithm to create multiple point-to-multipoint (p2mp) LSPs based on the optimum sub-flow values obtained with our MHDB-S model. Moreover, a general problem for supporting multicasting in MPLS networks is the lack of labels. To reduce the number of labels used, a label space reduction algorithm solution is also considered.**

Keywords – Subflows, MPLS, Multicast, Mathematical programming, optimization, traffic engineering.)

I. SUB-FLOW TO LSPS MAPPING PROBLEM IN P2MP

Traffic engineering is concerned with optimizing the performance of operational networks. The main objective is to reduce congestion in hot spots and to improve resource utilization. This can be achieved by setting up explicit routes through the physical network in such a way that traffic distribution is balanced across several traffic trunks [1]. Current configurations in computer networks provide an opportunity to disperse traffic over multiple paths to decrease congestion and achieve the aggregated end-to-end bandwidth requirement.

This load balancing technique can be achieved by a multicommodity network flow formulation [2], [3] and [4], which leads to the traffic being shared over multiple routes between the ingress node and the egress nodes in order to avoid link saturation and hence the possibility of congestion. Several advantages of using multipath routing are discussed in [5]: links do not get overused and therefore do not get congested, and multipath has the potential to aggregate bandwidth allowing a network to support more data transfer than it is possible with only one path, etc.

In previous work [6], [7], [8] and [9] we proposed a multi-objective traffic engineering scheme, the MHDB-S model, to multicast several flows. The aim of this model is to combine the following weighting objectives into a single aggregated metric: maximum link utilization, hop count, total bandwidth consumption, and total end-to-end delay. Moreover, our proposal solves the traffic split ratio for multiple trees. In unicast transmission, the split ratio is fed to the routers which divide the traffic from the same pair of ingress-egress nodes into multiple paths, i.e. each flow is split into multiple sub-flows. In multicast transmission, the load balancing consists of traffic being split (using the multipath approach) across multiple trees [10], depending on the solution obtained, between the ingress node and the set of egress nodes.

This work was partially supported by the MCyT under the project TIC2003-05567. The work of Yezid Donoso was supported by the Universidad del Norte (Colombia) under contract G01 02-MP03-CCBFPD-0001-2001. The work of Fernando Solano was supported by the Ministry of Universities, Research and Information Society (DURSI) of the Government of Catalonia under contract 2004FI-00693.

In this paper, we focus on the specific problem of mapping sub-flows to point-to-multipoint label switched paths (p2mp LSPs) for a multi-protocol label switched (MPLS) network implementations. The aim of this is to obtain an efficient solution to formulate p2mp LSPs given a set of optimum sub-flow values. [11] presents a sub-flow mapping solution based on a linear equation system which needs a large number of equations and variables. To solve this problem, a sub-flow mapping heuristic for creating multiple p2mp LSPs based on the optimum sub-flow values obtained with the MHDB-S model is proposed in this paper.

The rest of this paper is organized as follows. In Section 2, some related work and previously proposed MHDB-S models [6], [7], [8] and [9] are described. The sub-flow mapping problem is analyzed in Section 3. In Section 4 a heuristic algorithm that solves the previous analyzed problem is presented. In Section 5 the proposed solution is evaluated. Finally, in Section 6, we give our conclusions and suggestions for further study.

II. RELATED WORK

Various traffic engineering solutions using programming techniques for load balancing with multiple routes have been designed and analyzed in different studies (see [6] and [7] for a detailed explanation of these proposals). It should be pointed out that several proposals can be applied to MPLS networks. In [11], related works about splitting a multipath problem and support of multicasting in MPLS networks are commented. In this section, we describe the MHDB-S model and the problem related to the lack of labels in MPLS networks, since a sub-flow mapping solution could increase the number of p2mp LSPs used in the network.

A. Lack of label problem

A general problem of supporting multicasting in MPLS networks is the lack of labels. The MPLS architecture allows aggregation in p2p LSPs. Aggregation reduces the number of

0-7803-8836-4/04/$20.00 ©2004 IEEE.

labels that are needed to handle a particular set of flows, and may also reduce the amount of label distribution control traffic needed [12]. The addition of new LSPs increases the label space and hence the lookup delay. So, reducing the number of labels used is a desirable characteristic for any algorithm that maps flows to LSPs.

As pointed out in [12], the label based forwarding mechanism of MPLS can also be used to route along multi-point to point (mp2p) LSPs. In [13] and [14], aggregation algorithms that merge p2p LSPs into a minimal number of mp2p LSPs are considered. In this case, labels assigned to different incoming links are merged into one label assigned to an outgoing link. If two p2p LSPs follow the same path from an intermediate node to the egress node, these aggregation algorithms allocate the same label to the two p2p LSPs and thus reduce the number of used labels. In [15], an algorithm reducing the number of MPLS labels to N (number of nodes) + M (number of links) without increasing any link load is presented. For differentiated services with K traffic classes with different load constraints, their limit increases to K(N+M). Their stack-depth is only one, justifying implementations of MPLS with limited stack-depth. To reduce the number of used labels for multicast traffic, another label aggregation algorithm is presented in [16]. In this case, if two p2mp LSPs follow the same tree from an ingress node to the egress node set, the aggregation algorithm allocates the same labels to the two p2mp LSPs. Ingress nodes have a new table (named the Tree Node Table) saving node information from the p2mp LSP and label allocation is executed by using this table.

The label stack was introduced into the MPLS framework to allow multiple LSPs to be aggregated into a single LSP tunnel [12]. In [17], a comprehensive study of label size versus stack depth trade-off for MPLS routing protocols on lines and trees is undertaken. They show that, in addition to LSP tunneling, label stacks can also be used to dramatically reduce the number of labels required for setting up LSPs in a network. Their protocols have numerous practical applications that include implementation of multicast trees, and virtual private networks using MPLS as the underlying signaling mechanism.

B. *MHDB-S model*

The following model is a summary of that presented in [6], [7], [8] and [9]. In [6], we show that the multi-objective model produces a better result than various mono-objective models. In [8], we present an enhanced model (MHDB-D) for multicasting dynamic groups, and in [7] and [9] we present two heuristic algorithms to solve the previous models. The network is modeled as a directed graph $G=(N,E)$, where *N* is the set of nodes and *E* is the set of links. The set of links is $E \subseteq N \times N$. We use *n* to denote the number of network nodes, i.e. $n=|N|$. Among the nodes, we have a source $s \in N$ (ingress node) and some destinations T (the set of egress nodes). Let $t \in T$ be any egress node. Let $(i,j) \in E$ be the link from node *i* to node *j*. Let $f \in F$ be any multicast flow, where *F* is the flow set and T_f is the egress node subset to the multicast flow *f*. We use $|F|$ to denote the number of flows.

Let X_{ij}^{tf} be the fraction of flow *f* to egress node *t* assigned to link *(i,j)*; note that these variables include the egress node *t*. Including the egress node variables allows us to control the bandwidth consumption in each link with the destination of the set of egress nodes. Therefore, it is possible to maintain the exact constraint of flow equilibrium to the intermediate nodes. The problem solution, X_{ij}^{tf} variables, provides optimum flow values.

Let c_{ij} be the capacity of each link *(i,j)*. Let bw_f be the traffic demand of a flow *f* from the ingress node *s* to T_f. The binary variables Y_{ij}^{tf} represent whether link *(i,j)* is used (1) or not (0) for the multicast tree rooted at the ingress node *s* and reaching the egress node subset T_f. Let v_{ij} be the propagation delay of link *(i,j)*. Let *m* be the number of variables in the multi-objective function. Let $connection_{ij}$ be the indicator of whether there is a link between nodes *i* and *j*.

The problem of minimizing $|F|$ multicast flows from ingress node *s* to the egress nodes of each subset T_f is formulated as follows (MHDB-S model):

Minimize

$$r_1.\alpha + r_2 \sum_{f\in F}\sum_{t\in T_f}\sum_{(i,j)\in E} Y_{ij}^{tf} +$$

$$r_3 \sum_{f\in F}\sum_{(i,j)\in E} bw_f \max_{t\in T_f}\left(X_{ij}^{tf}\right) + r_4 \sum_{f\in F}\sum_{t\in T_f}\sum_{(i,j)\in E} v_{ij} Y_{ij}^{tf} \quad (1)$$

Subject to

$$\sum_{(i,j)\in E} X_{ij}^{tf} - \sum_{(j,i)\in E} X_{ji}^{tf} = 1 \quad ,t\in T_f, f\in F, i=s \quad (2)$$

$$\sum_{(i,j)\in E} X_{ij}^{tf} - \sum_{(j,i)\in E} X_{ji}^{tf} = -1 \quad ,i,t\in T_f, f\in F \quad (3)$$

$$\sum_{(i,j)\in E} X_{ij}^{tf} - \sum_{(j,i)\in E} X_{ji}^{tf} = 0 \quad ,t\in T_f, f\in F, i\neq s, i\notin T_f \quad (4)$$

$$\sum_{f\in F} bw_f . \max_{t\in T_f}(X_{ij}^{tf}) \le c_{ij}.\alpha \quad ,\alpha \ge 0, (i,j)\in E \quad (5)$$

$$\sum_{j\in N}\sum_{t\in T_f} Y_{ij}^{tf} \le \left\lceil \frac{bw_f}{\left\lfloor \sum_{j\in N} c_{ij} \Big/ \sum_{j\in N} connection_{ij} \right\rfloor} \right\rceil \quad ,i\in N, f\in F \quad (6)$$

where

$$X_{ij}^{tf} \in \Re, 0 \le X_{ij}^{tf} \le 1 \quad (7)$$

$$Y_{ij}^{tf} = \left\lceil X_{ij}^{tf} \right\rceil = \begin{cases} 0 & ,X_{ij}^{tf}=0 \\ 1 & ,0 < X_{ij}^{tf} \le 1 \end{cases} \quad (8)$$

$$\sum_{i=1}^{m} r_i = 1, \quad r_i \in \Re, \quad r_i \ge 0, m > 0 \quad (9)$$

The multi-objective function (MHDB model) (1) defines a function and generates a single aggregated metric through a combination of weighting objectives, r_i. The main objective consists in minimizing the maximum link utilization (MLU), which is represented as α in equation (1). In this case, the solution obtained may report long routes. In order to eliminate these routes and to minimize the hop count (HC), the term $\sum_{f\in F}\sum_{t\in T_f}\sum_{(i,j)\in E} Y_{ij}^{tf}$ is added. In order to minimize the total bandwidth consumption (BC) over all links, the term $\sum_{f\in F}\sum_{t(i,j)\in E} bw_f \max_{t\in T_f}\left(X_{ij}^{tf}\right)$ is also added. This is included so that, if there is more than one solution with the best maximum link utilization, the solution with the minimum resource utilization is chosen. Though several sub-flows of the flow *f* in the link *(i,j)* with destinations to different egress nodes are sent, in multicast IP specification just one sub-flow will be sent, that is, only the maximum value of X_{ij}^{tf} for $t \in T_f$ needs to be considered. Furthermore, in order to minimize the total end-to-end propagation delay (DL) over all the links, the term $\sum_{f\in F}\sum_{t\in T_f}\sum_{(i,j)\in E} v_{ij}\, Y_{ij}^{tf}$ is also added.

Constraints (2), (3) and (4) are flow conservation constraints. Constraint (2) ensures that the total flow emerging from ingress node to any egress node *t* at flow *f* is 1. Constraint (3) ensures that the total flow coming from an egress node *t* at flow *f* is 1. Constraint (4) ensures that for any intermediate node different from the ingress node ($i \neq s$) and egress nodes ($i \notin T_f$, $\forall\, f\in F$), the sum of their output flows to the egress node *t* minus the input flows with destination egress node *t* at flow *f* is 0.

Constraint (5) is the maximum link utilization constraint. In a unicast connection, the total amount of bandwidth consumed by all the flows with the destination of egress node *t* must not exceed the maximum utilization (α) per link capacity c_{ij}, that is, $\sum_{f\in F} bw_f \sum_{t\in T} X_{ij}^{tf} \le c_{ij}\mathrm{A}, (i,j)\in E$. Nevertheless, in constraint (5) only the maximum value of X_{ij}^{tf} for $t \in T_f$ needs to be considered.

Constraint (6) limits the maximum number of sub-flows (MSF) in each node by means of the capacity of each link and the traffic demand. This formulation represents the amount of necessary links for a particular traffic demand. Without this constraint, the model could suffer from scalability problems, i.e. the label space used by LSPs would be too high.

Expression (7) shows that the X_{ij}^{tf} variables must be real numbers between 0 and 1. These variables form multiple tree transport multicast flow. The demand between the ingress node and the egress node *t* may be split over multiple routes. When the problem is solved without load balancing, this variable will only be able to take values 0 and 1, which will show, respectively, whether or not the link *(i,j)* is used to carry information to egress node *t*.

Expression (8) calculates Y_{ij}^{tf} as a function of X_{ij}^{tf}.

Finally, expression (9) shows that the weighting coefficients, r_i, assigned to the objectives are normalized. These values are calculated by solving the optimization problem.

III. SUB-FLOW TO LSPS MAPPING PROBLEM IN P2MP

In this section, we detail the problem of mapping multiple p2mp LSPs based on the optimum sub-flow values X_{ij}^{tf} obtained with MHDB-S model (1). However, this mapping is difficult using the MHDB-S model because there is no index for identifying sub-flows [11]. Remember that X_{ij}^{tf} is the fraction of flow *f* to destination node *t* assigned to link *(i,j)*. As the presented algorithm applies only to one flow *f*, the index *f* will be omitted when it does not cause confusion.

To explain the problem, the MHDB-S models have been applied to the topology of Figure 1, with a single flow *f*, where s=N1 and T={ N5, N6}. In this case, a possible sub-flow solution (X_{ij}^{t}) obtained is shown in Figure 2.

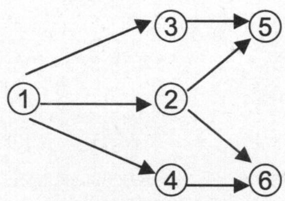


Figure 1. Physical network topology.

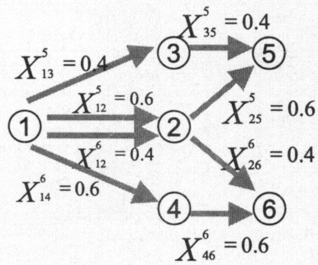


Figure 2. MDDB-S solution

The simplest solution (Fig. 3) to create LSPs based on the optimum sub-flow values is to send each sub-flow (0.4 and 0.6 fractions) to the group separately, and in this case each sub-flow is mapped to one p2mp LPS.

In Figure 3, each packet represents a 0.2 fraction of the flow. With this mapping, sub-flows X_{12}^5 and X_{12}^6 are different and the maximum link utilization constraint (5) could be violated. Moreover, the network is inefficiently used because multicast node capabilities are not considered. Only ingress node multicast capabilities are considered when applying the multipath approach, which permits the flow to be balanced across several links.

A second approach considers that one sub-flow is included in the other, i.e. min (X_{12}^5, X_{12}^6) $\subseteq$ max (X_{12}^5, X_{12}^6), in the example $X_{12}^5 \subseteq X_{12}^6$. If both sub-flows X_{12}^5 and X_{12}^6 are sent over the link (1,2) to each member of the group separately (Fig. 4), a part of the same flow is being transmitted over the same link and the network is also inefficiently used.

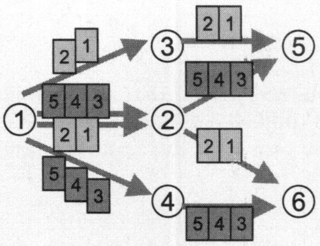


Figure 3. Simplest p2mp mapping.

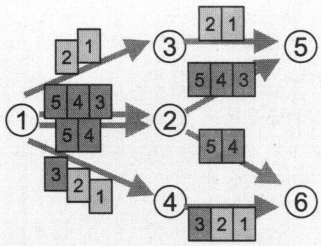


Figure 4. p2mp assignment: unicast transmission

Moreover, if a node has multicast capabilities, it is not necessary to transmit all the sub-flows over the link. In particular, if N2 has multicast capabilities, only the max (X_{12}^5, X_{12}^6) must be transmitted over link (1,2) (Fig. 5).

However, this solution presents a problem in the forwarding mechanism. Some incoming packets at node 2 must be forwarded only once (packet 3), but other packets of the same sub-flow must be forwarded by different output links (packets 4 and 5). To solve this, the ingress node must split this sub-flow in several LSPs (Fig. 6).

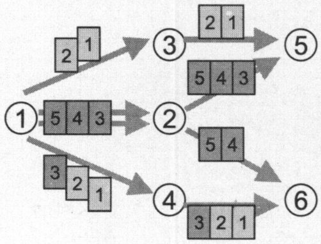


Figure 5. p2mp assignment: multicast transmission

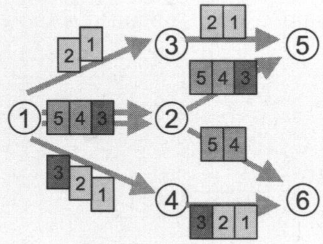


Figure 6. p2mp assignment: sub-flow mapping

IV. SUB-FLOW MAPPING HEURISTIC

In more complex networks the linear system equation previously presented [11] needs a large number of equations and variables. Therefore, in this section, a sub-flow mapping heuristic to map sub-flow values into LSPs is proposed.

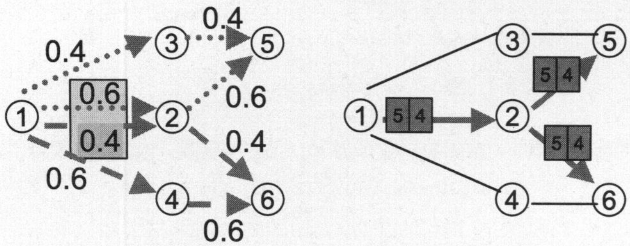
The method presented in this section returns a set of p2mp that will transmit 100% of a given flow *f* from the source node *s* to the destinations node set T_f. In order to explain the procedure, we will start presenting the complementary notation used. Let $\le_{s,d}$ be a total ordered relationship defined over pair of links, i.e. $E \times E$, in which: $(i,j) \le (k,m)$ if, and only if, *j*=*k*; this means that there exists a node, *j* or *k*, that can forward packets incoming from node *i* to node *j* directly; $\forall (i,j) \in E, (s,x) \le (i,j)$; and $\forall (i,j) \in E, (i,j) \le (x,d)$. Let $(P \subseteq E, \le_{s,d})$ be the set that represents a single unicast path in graph *G* starting at node *s* and ending at node *d*, briefly resumed as $P_{s,d}$. By aggregating a bandwidth restriction to the set of links, we can restrict the set $P_{s,d}$ to $P_{s,d}^{t,c}$ in the following way: $P_{s,d}^{t,c} = \{(i,j) \mid (i,j) \in P_{s,d}, X_{ij}^t \ge c\}$ and $P_{s,d}^{t,c}$ fits in $\le_{s,d}$. This means that $P_{s,d}^{t,c}$ is a path (found in X_{ij}^t) that could transmit *c*-percent of the total flow from *s* to *d*. Similar to $P_{s,d}^{t,c}$, a tree is modeled as a set of links following a partial order relationship over the links analogous to $\le_{s,d}$. Let *p2mp* be a set, initially empty, in which each element of the set is a tuple of a tree, usually denoted as *M*, and a real number *c* between 0 and 1 indicating the percentage of the flow been transmitted using that tree.

1 Sub-flow mapping (p2mp, X_{ij}^t, s, T_f)
2 Parameter: X_{ij}^t s, T_f
3 Return: p2mp
4 Begin
5 $p2mp = \phi$
6 While $\sum X_{ij}^t > 0$
7 $L = \left\{ (i,j) \mid \sum_{t \in T_f} Y_{ij}^t = \sup_{(i,j) \in E} \left(\sum_{t \in T_f} Y_{ij}^t \right) \right\}$
//** Computes the set of network links (L) that transmit a maximum number of sub-flows.
8 $(\alpha,\beta) \in \left\{ (i,j) \mid \inf_{(i,j) \in L} \left(X_{ij}^t \right) \right\}$
//** Choose link (α,β) in L that transmits the less fraction.
9 $c = \inf_{(\alpha,\beta)} \left(X_{\alpha\beta}^t \right) \in \Re, 0 < c \le 1$
10 For each $t \in T_f$ do
11 If $P_{\beta,t}^{t,c}$ exists and $X_{\alpha,\beta}^t \ge c$ then
12 $M_t = P_{s,\alpha}^{t,c} \cup \{(\alpha,\beta)\} \cup P_{\beta,t}^{t,c}$
13 Else
14 $M_t = P_{s,t}^{t,c}$
15 End if
16 $\forall (i,j) \in M_t, X_{ij}^t = X_{ij}^t - c$
17 End for
18 $M = \bigcup_{t \in T_f} M_t$
19 $p2mp = p2mp \cup \{(M,c)\}$
20 End While
21 End

Figure 7. Fig. 1. Sub-flow mapping algorithm

The path $P_{\beta,t}^{t,c}$, in line 11 of Figure 7, will still exist when *t* is the destination index of the selected sub-flow fraction *c* in line 9, because of the law of the conservation of flow in (2), (3) and (4). Note that at least one X_{ij}^t is eliminated, i.e. put to zero, in each iteration (6 to 20), because line 16 in Figure 7 reduces the bandwidth of link (α,β) by exactly *c* units when the destination is t in 9. Therefore, the complexity of the procedure is $O\left(\left|X_{ij}^{t}\right|\cdot\left|T_f\right|\right)$ for a single flow.

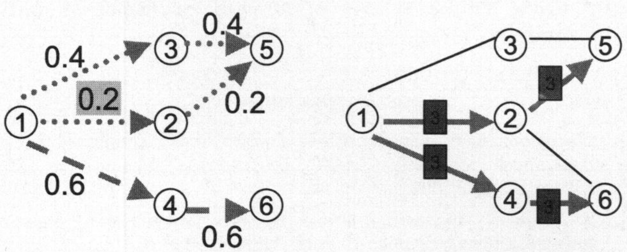
Figure 8-a. shows the sub-flow solution (X_{ij}^{t}) considered to explain the sub-flow mapping algorithm. Initially, in lines 7 to 9 of Figure 7, the set L will contain only the link (1,2) for destinations 5 and 6. Therefore, (α,β) will be (1,2) and c = 0.4 (see also Fig. 8-a.) because it is the minimum fraction for all destinations computed in (1,2). Assuming that the cycle in line 10 begins with t = 5, then path $P_{2,5}^{5,0.4}$ exists. In the same way, path $P_{2,6}^{6,0.4}$ also exists. Hence, the multicast tree built in the first iteration is $M=\{(1,2),(2,5),(2,6)\}$ (see Fig. 8-b) and the values $X_{1,2}^{5}, X_{1,2}^{6}, X_{2,5}^{5}$ and $X_{2,6}^{6}$ are reduced to 0.4.



a. Initial sub-flow fraction values b. First multicast tree: (40%)

Figure 8. First iteration of the algorithm.

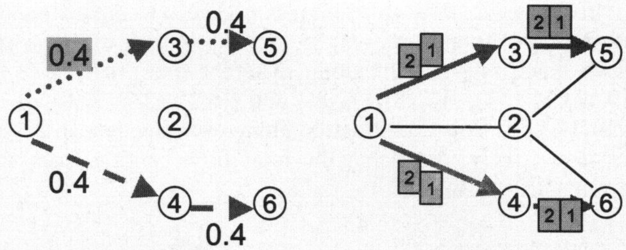
In this case, the set of remaining fraction values X_{ij}^{t} is presented in Figure 9-a. The link (α,β) selected could be either (1,2) or (2,5) because both have the smallest capacity (c= 0.2). We will select (1,2) for the example. In this case, path $P_{2,5}^{5,0.2}$ exists, but $P_{2,6}^{6,0.2}$ does not. So, $M_5=\{(1,2),(2,5)\}$ is computed with line 12 of Figure 7 and $M_6=\{(1,4),(4,6)\}$ with line 14. The resulting tree $M=\{(1,2),(1,4),(2,5),(4,6)\}$ with sub-flow fraction 0.2 is shown in Figure 9-b.



a. New sub-flow fraction values. b. Second multicast tree: (20%)

Figure 9. Second iteration of the algorithm.

Finally, the set of remaining fraction values X_{ij}^{t} is presented in 12-a. Note that all the fraction values in the last iteration are the same, in this case c = 0.4. The algorithm applied here returns the remaining tree, $M=\{(1,3),(1,4),(3,5),(4,5)\}$ (see Fig. 10-b).



a. Sub-flow fraction values before final iteration. b. Third multicast tree: (40%)

Figure 10. Final iteration of the algorithm.

V. EXPERIMENTAL RESULTS

We have analyzed the performance of the sub-flow mapping algorithm. Over the 14-node NSF (National Science Foundation) network which has 20×2 links (Fig. 1) two flows with the same source, s=N0, are transmitted. The egress nodes subsets are T1={N5, N8, N11} and T2={N8, N11, N13}, respectively. The transmission rates are 0.25 Mbps, 0.5 Mbps, 1 Mbps, 1.5 Mbps, 2 Mbps and 2.5 Mbps for each flow.

Table I shows minimal flow fraction $c_{\min}$ and the number of p2mp LSPs obtained with the mapping algorithm using the results of MHDB-S model.

TABLE I. MINIMAL FLOW FRACTION ($c_{\min}$) AND NUMBER OF P2MP LPS OBTAINED IN FLOW 1 AND FLOW 2

Flow rates (Mbps)	$c_{\min}$	Number of P2MP connections	
		Flow1	***Flow2***
0.25 – 0.5 – 1.0	33%	2	2
1.5	10%	8	5
2.0 – 2.5	10%	9	6

In particular, the four obtained trees for flow rates {0.25, 0,5, 1} are shown in Figures 11 and 12. In figure 11, flow 1 from ingress node N0 to egress nodes {N5, N8, N11} and rate={0.25, 0,5, 1} is split into two sub-flows. Each one is sent along different trees: {(0,2), (2,7), (7,8), (2,4), (4,5), (4,10), (10,11)} and {(0,1), (1,6), (6,5), (6,9), (9,8), (9,11)}. Sub-flow fraction along each tree is *2/3* and *1/3* respectively. Note that the total flow coming from each egress node is 1.

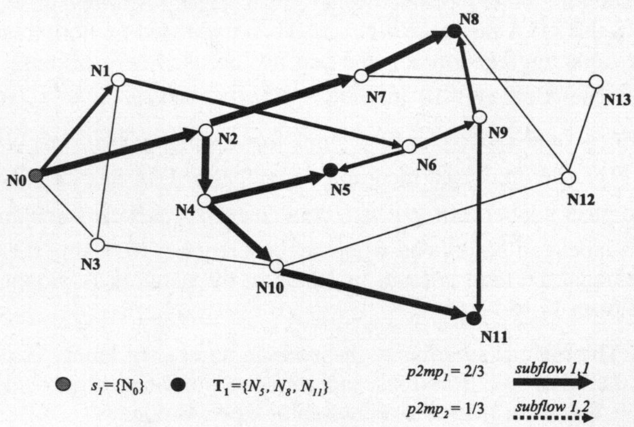


Figure 11. Sub-flows of flow 1

In figure 12, flow from ingress node N0 to egress nodes {N8, N11, N13} and rate={0.25, 0,5, 1} is split into two sub-flows. Each one is sent along different trees: {(0,3), (3,10), (10,12), (10,11), (12,13), (12,8)} and {(0,2), (2,7), (7,8), (7,13), (8,9), (9,11)}. Sub-flow fraction along each tree is also *2/3* and *1/3* respectively. Note that the total flow coming from each egress node is also 1.

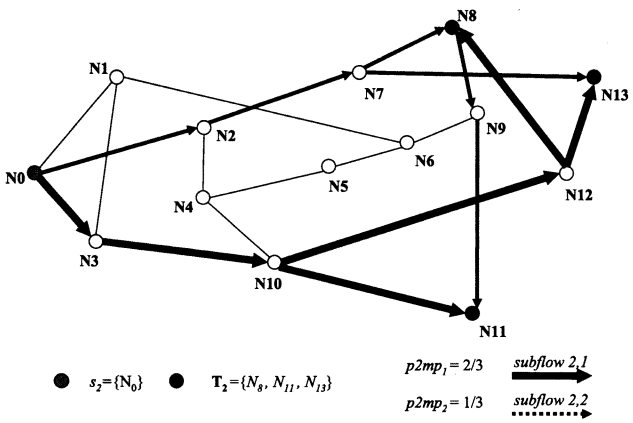


Figure 12. Sub-flows of flow 1

Due to the sub-flow mapping algorithm proposed using a considerable amount of trees, we are forced to face the lack of label problem explained before. To solve this, a label aggregation and stacking algorithm is used. If two p2mp LSPs follow the same tree from an intermediate node to any egress node subset, the aggregation algorithm allocates the same labels in these branches. In the stacking algorithm, non-branched p2mp segments are joined (tunneled) by using a single label and the MPLS label stack is used in order to save each segment's forwarding information. Table 3 shows the average number of labels used per node in the network with both label space reduction algorithms (second row) and with neither of them (first row).

To illustrate these label space reduction algorithms, suppose that $p2mp_1$ and $p2mp_2$ are two trees (see Fig. 13) of flow 2 which optimize the MHDB-S model objectives. The branch starting at N10 and ending at {N8, N13} through N12 can use a single label since both $p2mp_1$ and $p2mp_2$ share exactly the same paths between branch starting N10 and ending {N8, N13}. Although N0-N3-N10 is a path used by both trees, previous aggregation scheme can not be used here because it will cause either N10 forwards $p2mp_2$ packets to N11 (i.e. packets duplication), or N10 stops forwarding $p2mp_1$ packets to N11 (i.e. multicast incomplete replication). In this case, N0 can push a label into $p2mp_1$ and $p2mp_2$ packets stack and this label can be popped when packets reach N10. Using these methods, the total amount of labels in the network is dropped off from 17 to 12.

The table II summarize the average number of labels found in the MHDB-S solutions applying the presented algorithms. Note that a considerable reduction of labels is done.

TABLE II. AVERAGE LABELS PER NODE

Flow rates (Mbps)	Average labels per node		Reduction ratio
	Without label space reduction	***With label space reduction***	
0.25 – 0.5 – 1.0	26 /14 = 1.86	20 / 14 = 1.43	23.07 %
1.5	81 / 14 = 5.79	46 / 14 = 3.29	43.21 %
2.0 – 2.5	99 / 14 = 7.07	52 / 14 = 3.71	47.47 %

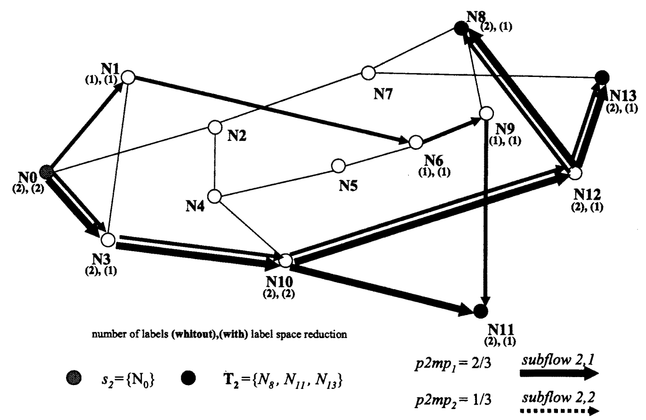


Figure 13. Flow from ingress node N0 to egress nodes {N8, N11, N13} is split into two sub-flows $p2mp_1$ and $p2mp_2$.

VI. CONCLUSIONS

In this paper we consider a multi-objective traffic engineering scheme using different distribution trees to multicast several flows and we propose a sub-flow mapping solution based on a a heuristic algorithm to create multiple p2mp LSPs. Despite the merging of LSPs reducing the number of LSPs (hence the label space), label aggregation and stacking algorithms were also considered because they reduce the number of labels needed even more. In the example shown, this reduction is between 23.07 and 47.47 %.

Although the sub-flow mapping methods proposed make mapping LSPs for an MPLS implementation easy, in the future we plan to define a model which includes an extra index to identify sub-flows. The usefulness of sub-flow mapping in p2mp LSPs with a variety of network scenarios should be demonstrated.

REFERENCES

[1] Changhoon Kim, Yanghee Choi, Yongho Seok, Youngseok Lee. "A constrained Multipath Traffic Engineering Scheme for MPLS Networks". ICC 2002.

[2] R. Ahuja, T. Magnanti, J. Orlin. "Network flows: Theory, algorithms and applications". Prentice-Hall, NJ, USA, 1993.

[3] M. Bazaraa, J. Jarvis, H. Sherali. "Linear Programming and Network Flows". John Wiley & Sons, 2nd ed, USA, 1990.

[4] M. Bazaraa, H. Sherali, C. M. Shetty. "Nonlinear Programming, Theory and Algorithms". John Wiley & Sons, 2nd ed, USA, 1993.

[5] J.C. Chen, S.H. Chan. "Multipath routing for video unicast over bandwidth-limited networks". GLOBECOM 2001

[6] Y. Donoso, R. Fabregat, L. Fàbrega. "Multi-Objective scheme over multi-tree routing in multicast MPLS networks". ACM/IFIP LANC'03.

[7] Y. Donoso, R. Fabregat and J. L. Marzo. "Multi-objective optimization algorithm for multicast routing with traffic engineering". IEEE ICN'04.

[8] Y. Donoso, R. Fabregat, J. L. Marzo. "Multi-Objective optimization scheme for dynamic multicast groups". IEEE ISCC'04.

[9] Y. Donoso, R. Fabregat, J. L. Marzo. "Multi-Objective Optimization Model and Heuristic Algorithm for Dynamic Multicast Routing". IEEE & VDE Networks 2004.

[10] Y. Wang, Z. Wang, L. Zhang. "Internet Traffic Engineering without Full Mesh Overlaying". INFOCOM 2001.

[11] F. Solano, R. Fabregat, Y. Donoso. "Sub-flow assignment model of multicast flows using multiple p2mp LSPs". 30th Conferencia Latinoamericana de Informática (CLEI2004). 27 September 2004.

[12] E. Rosen, A. Viswanathan, R. Callon. "Multiprotocol Label Switching Architecture". RFC 3031. January 2001.

[13] H. Saito, Y. Miyao, M. Yoshida. "Traffic engineering using multiple multipoint-to-point LSPs". INFOCOM 2000.

[14] S. Bhatnagar, S. Ganguly, B. Nath. "Label space reduction in multipoint-to-point LSPs for traffic engineering". ECUMN2002.

[15] D. Applegate, M. Thorup. "Load optimal MPLS routing with N+M labels". INFOCOM'03.

[16] Y.K. Oh, D.K. Kim, H.J. Yoen, M.S. Do, J. Lee. "Scalable MPLS multicast using label aggregation in Internet broadcasting systems". ICT'03.

[17] A. Gupta, A. Kumar, R. Rastogi. "Exploring the trade-off between label size and stack depth in MPLS Routing. INFOCOM'03.