Sub-flow assignment model of multicast flows using multiple P2MP LSPs

Fernando Solano, Ramón Fabregat
Institut d'Informàtica i Aplicacions, Universitat de Girona
Av. Lluis Santaló s/n. EPS P4, (17071) Girona, España.
{fsolanod, ramon}@eia.udg.es

and

Yezid Donoso
Departamento de Ingeniería de Sistemas y Computación, Universidad del Norte
Km. 5 vía Pto. Colombia, Barranquilla, Colombia.
ydonoso@uninorte.edu.co

Abstract

In previous works, a multi-objective traffic engineering scheme (MHDB-S model) using different distribution trees to multicast several flows were proposed. Because the flow assignment cannot be mapped directly into MPLS architecture, in this paper, we propose a liner system equation to create multiple point-2-multipoint LSPs based on the optimum sub-flow values obtained with our MHDB-S model.

Keywords: Multiobjective Optimization, Multicast, MPLS, Sub-flow assignment

1.Introduction

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 [22]. 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], [7] and [8], 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 [13]: 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 [17], [18], [19] and [20] 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 [31], depending on the solution obtained, between the ingress node and the set of egress nodes.

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. To solve this problem, a sub-flow assignment solution based on a linear equation system 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, we describe some related studies. In section 3, we explain the multi-objective scheme for static multicast routing (MHDB-S model) [17] that multicast several flows and solve the traffic split ratio for multicast trees. The sub-flow assignment problem is analyzed in section 4. In section 5, we propose a linear equation system for creating multiple P2MP LSPs based on the optimum sub-flow values obtained with the MHDB-S model. The proposed assignment solution is evaluated in section 6. Finally, in section 7, we give our conclusions and suggestions for further study.

2. Related work

2.1. Multipath routing: splitting flows

Several papers ([26], [12], [28], [30] and [14]) address the splitting multipath problem of unicast traffic, motivated by its importance in complete traffic engineering solutions. Traffic splitting is executed for every packet in the packet-forwarding path. A simple method to partition the input traffic is on a per-packet basis, for example in a round-robin fashion. However, this method suffers from the possibility of excessive packet reordering and is not recommended in practice.

Reference [29] tries to balance the load among multiple LSPs according to the loading for each path. In

MPLS networks [14-CLEI] multiple paths can be used to forward packets belonging to the same "forwarding equivalent class" (FEC) by explicit routing.

In [26], Rost and Balakrishnan propose a multi-path transmission between sources and destinations. The current configurations in computer networks provide an opportunity for dispersing traffic over multiple paths to decrease congestion. In [26] dispersion involves (1) splitting, and (2) forwarding the resulting portions of aggregate traffic along alternate paths. The authors concentrate on (1): methods that allow a network node to subdivide aggregate traffic, and they offer a number of traffic splitting policies which divide traffic aggregates according to the desired fractions of the aggregate rate. Their methods are based on semi-consistent hashing of packets to hash regions as well as prefix-based classification.

In [11], the performance of several hashing schemes for distributing traffic over multiple links while preserving the order of packets within a flow is studied. Although hashing-based load balancing schemes have been proposed in the past, [11] is the first comprehensive study of their performance using real traffic traces.

2.2. Support of multicasting in MPLS networks

In MPLS, unicast and multicast packets have already been assigned to different type code in the link-layer header. Therefore, MPLS routers know whether a packet is from a unicast or a multicast flow. In the case of unicast forwarding the event of an incoming flow leads to the forwarding of exactly one flow. The packet duplication mechanism that is implemented in IP routers to support the IP multicast can be used to duplicate MPLS packets. MPLS routers at the bifurcation of a multicast routing tree duplicate packets and send copies of the same packet on different outgoing links. Although MPLS natively supports multicasting in its design, the MPLS community has focused its efforts mainly on the label switching of unicast IP traffic, leaving the sections on multicasting in the main MPLS documents ([25] and [3]) virtually empty, to be addressed in future studies. Based on this, there are some proposals for supporting multicasting in MPLS networks.

A framework for MPLS multicast traffic engineering proposed by Ooms et al. [24] gives an overview of the applications of MPLS techniques to IP multicast. Another proposal explains how to distribute labels for unidirectional multicast trees [5] and for bi-directional trees' label distribution [15].

To provide MPLS Traffic Engineering [5] to a P2MP-application, existing MPLS point-to-point (P2P) mechanisms have to be enhanced to support the P2MP TE LSP setup. Reference [32] presents a set of requirements for P2MP extension to MPLS TE.

On march 1 2004, two different solution drafts ([33] and [1] for TE P2MP LSPs were presented at a MPLS working group meeting in Seul, but the chairs and the meeting strongly encourage the authors to both need to get together and converge on a single solution. The computation of P2MP TE paths is implementation dependent and is beyond the scope of those solutions. Some off-line or on-line algorithms can compute path information, e.g. the MHDB-S model presented in the next section.

Reference [33] describes a solution for P2MP TE, which extends [6] and [9] in order to establish, maintain, and teardown a P2MP TE LSP. In this case, a P2MP TE LSP is established by setting up multiple standard P2P TE LSPs from an ingress node to one of the leaf nodes of the P2MP TE LSP. The calculation for a P2MP requires three major pieces of information. The first is the route from the ingress LSR of a P2MP path to each of the egress LSRs, the second is the traffic engineering related parameters, and the third is the branch capability information.

Reference [1] describes how RSVP-TE can be used for P2MP TE. It relies on the semantics of RSVP that RSVP-TE inherits for building a P2MP TE tree. P2P TE LSPs are set up between ingress LSR and egress

LSRs. These P2P TE LSPs are appropriately merged by the network using RSVP semantics to result in a P2MP TE LSP.

Various traffic engineering solutions using programming techniques to balance loads by multiple routes have been designed and analyzed in different studies (see [17] and [18] for a detailed explanation of these proposals). It should be pointed out that several proposals could be applied to MPLS networks. In [17], we show that the multi-objective model produces a better result than various mono-objective models. In [19], we present an enhanced model (MHDB-D) for multicasting dynamic groups, and in [18] and [20] we present two heuristics algorithms to solve the previous models.

2.3. The lack of labels 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 labels that are needed to handle a particular set of flows, and may also reduce the amount of label distribution control traffic needed [25]. 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 [25], the label based forwarding mechanism of MPLS can also be used to route along multi-point to point (MP2P) LSPs. In [27] and [10], **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 [4], 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 [23]. 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 [25]. In [21], 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.

Aggregated multicast is a scheme to reduce multicast state [16]. The key idea is that, instead of constructing a tree for each flow, there can be multiple multicast flows share a single aggregated tree to reduce multicast state and, hence, tree maintenance overhead and the number of used labels. Data packets from different flows are multiplexed in the same distribution tree, called aggregated tree. Each data packet of each group is encapsulated and travels through the aggregated tree.

3. Optimization model

The following model is a summary of that presented in [17], [18], [19] and [20]. In [17], we show that the multi-objective model produces a better result than various mono-objective models. In [19], we present an enhanced model (MHDB-D) for multicasting dynamic groups, and in [18] and [20] 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

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

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

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

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

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

where

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

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

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