Capacity Planning of Survivable MPLS Networks Supporting DiffServ

Kehang Wu and Douglas S. Reeves*
Departments of Electrical and Computer Engineering and Computer Science
North Carolina State University
kwu@unity.ncsu.edu, reeves@eos.ncsu.edu

Abstract

It is essential for ISPs to offer both performance and survivability guarantees at the IP/MPLS layer. Capacity planning is needed to ensure there will be sufficient resource availability and quality of service under normal circumstances, and during network failures.

In this paper we study the issue of capacity planning for survivable MPLS networks providing DiffServ EF and BE traffic classes. Our goal is to minimize the total link cost, subject to the performance and survivability constraints of both EF and BE classes. The problem is formulated as an optimization problem, where we jointly select the routes for edge to edge EF and BE user demand pairs, and assign a discrete capacity value for each link. Capacity for the EF class is fully restorable under the single link failure model, while restorability of the BE class can be adjusted by the network operator.

We propose an efficient solution approach based on the Lagrangian Relaxation and subgradient methods. Computational results show that the solution quality is within a few percent of optimal, while the running time remains reasonable for networks with 1000 nodes, 2500 links, and 40,000 demand pairs. This represents the first work on capacity planning for survivable multiple-class-of-service IP/MPLS networks with non-linear performance constraints.

1 Introduction

The last few years witnessed the explosive growth of mission-critical data traffic carried by the Internet, and the emergence of general societal dependencies on the Internet. We are now almost as dependent on

*This work is supported by DARPA and AFOSR (under grants F30602-99-1-0540 and F49620-99-1-0264).

the availability of the communication networks as on other basic infrastructures, such as roads, water, and power. With up to 100 terabits per second of data traffic possibly carried in a single fiber with DWDM, failure of a single fiber could be catastrophic. Despite efforts to physically protect the fiber optic cables, cable cuts are surprisingly frequent according to the FCC. Survivable network design, which refers to the incorporation of survivability strategies to mitigate the impact of the given failure scenarios, has become a vital part of network design, not an afterthought.

The current network architecture is moving toward a two layers structure where IP/MPLS [15] [29] is on top of optical WDM networks. Traditionly, network resilience primarily relies upon the functionalities provided by SONET and ATM [35][31]. As mentioned in [18], most ISPs rely on the IP layer when failures occur. But IP-based failure recovery schemes suffer from long delay and/or routing instability [27].

MPLS-based recovery has become a viable solution because of the faster restoration time than IP rerouting. There are many works that investigated recovery schemes in MPLS enabled networks [14][20]. With MPLS layer protection, the failure is detected either by the MPLS layer detection mechanism (such as "exchange of Hello message"), or the signaling message propagated from the lower layer [14]. Once a failure is detected, protection switching based recovery mechanisms are used. A LSP is protected by a pre-established recovery path or path segment. The resource allocated to the recovery path, the so-called spare capacity, may be fully available to preemptible low priority traffic. The spare capacity can be shared by multiple recovery paths corresponding to different failure scenarios, to further reduce the resource redundancy [19]. With MPLS-based recovery, it is possible to differentiate the level of protection for different classes of service [5]. For example, Expedited Forwarding (EF) based service, such as virtual lease line service, can be supported by link protection with a reserve path that offers 100% restorability. Best effort (BE) traffic, on the other hand, may not need to be fully restorable, or may simply rely on IP rerouting without being assigned any spare capacity in the network.

The cost of redundancy for survivability can be very high, compared to a corresponding network

designed only to serve the working demands under normal conditions. One of the major interests in survivable network design has been providing cost-efficient resource allocation.

Most of the work on survivable network design has been concentrated on the classic spare capacity allocation problem (SCA) [13][23][22]. The SCA problem assumes that information about the operational network is given; the focus is only on the calculation of backup paths and spare capacities, not the primary path and total capacity. It has been shown in [19], however, that even though optimizing the working path and the backup path together is computationally more expensive, it offers significant advantage for the total capacity savings. Since capacity planning is usually performed infrequently (on the time scale of weeks to months), we can afford the extra running time in exchange for a lower network cost.

With the popularization of e-commerce and new value-added services over IP, it is desirable for the next generation Internet to offer Quality of Service (QoS), the ability to support multiple traffic classes with different levels of performance guarantees. MPLS and Differentiated Service [8][28] are regarded as two key components of QoS. There has been work on survivable network design touching on the issues of multi-class traffic, such as the architectural aspects of differentiated survivability [5]. However, no quantitative treatment of the subject has been done so far.

In this paper, we study the issue of capacity planning of survivable MPLS networks. The targeted MPLS networks will be providing DiffServ EF and BE traffic classes [34][11]. Since our approach requires a precise performance model for optimization, we do not include the AF traffic classes in this paper, due to the lack of a consensus on the implementation of the AF PHB [4]. We focus on the problem of resource dimensioning and traffic routing, and assume that the network topology is given. The problem is formulated as an optimization problem, where we jointly select the primary and backup LSPs for edge to edge EF and BE user demand pairs [1], and assign a discrete capacity value for each link. The goal is

[1]The term "'demand pair"' and "'demand"' are used interchangeably through out the paper. Depending on the goals of the network designer, a demand pair may correspond to a single flow, or an aggregation of flows sharing the same QoS requirements, source and destination, and path through the network.

to minimize the total link cost, subject to the performance and survivability constraints of both EF and BE classes. The non-bifurcated (i.e., single-path) routing model is used for the EF class as required by [11], so that the traffic from a single EF demand pair will follow the same LSP between the origin and the destination. Traffic in the BE class is allowed to be split across multiple LSPs. The performance constraint of EF traffic is only represented by a bandwidth requirement, as specified in [11].

Most of literature on the design of survivable network only consider the case of single link failures [14]. While there are recent works investigating the issues of double-link failure [10], the existing methods for protection switching are designed for single link failures. We assume there is only single link failure in this paper. We assume that each EF user demand is assigned a primary LSP as well as a backup LSP with enough preemptible bandwidth; therefore it will not be affected by any single link failure. The performance as well as the survivability constraints of the BE class are characterized by the upper limit on the average delay in each link, assuming no more than a single link failure. Queueing is modeled as M/G/1 strict priority queues.

The novel aspect of our problem is the fact that two traffic classes, EF and BE, with independent performance and survivability requirements, share the same capacity resource, which results in a complex non-linear performance constraint. In addition to the non-linear performance constraint, non-bifurcated routing and discrete link capacity constraints dramatically increase the degree of difficulty, and significantly limit the viable solution approaches.

There are related papers on capacity planning that do not deal with survivability issues. A survey of those works can be found in [32].

The remainder of this paper is organized as follows. In Section 2, notation and detailed assumptions and models are presented. The problem definition is given in Section 3. Section 4 shows a Lagrangean relaxation of the original problem, and describes the subgradient procedure to solve the resulting dual problem. Section 5 presents some numerical results on the use of the method. The paper is concluded in

Section 6.

2 Notation and Assumptions

The issue of capacity planning of reliable MPLS networks supporting DiffServ EF and BE traffic classes is studied in this paper. The network is considered reliable as long as the given performance constraints are satisfied under any single link failure scenarios.

A link based formulation is used throughout this paper. Our notation is defined in Table 1. Compared to our previous work [34], $B_{kj}, \delta_b^l, \hat{x}_{jkb}^{ef}, \hat{x}_{jkb}^{be}$, and r are new variables introduced to reflect the survivability constraints. Please refer to [32] for more complete discussion on the notations and assumptions.

The network topology is assumed to be given, as well as the set of candidate LSPs J_k for each O-D (Origin-Destination) pairs. As mentioned, we only study the failure scenario where there is at most a single link failure. The given networks must be two-link-connected [12], which implies that there are at least two link-disjoint paths for any O-D pair. For every candidate path j, $j \in J_k$, there is a set of set of candidate backup paths B_{kj}, which consists of LSPs that are link-disjoint from the primary LSP j, to ensure the availability of the backup path regardless of the failure scenarios. δ_b^l defines the route of backup path b. The backup paths for the EF and BE user demands on path j, will be specified by the backup path routing variables $\hat{x}_{jkb}^{ef}, \hat{x}_{jkb}^{be}$ respectively.

T_l denotes the index of available link types for link l. Only one link type can be used for each individual link. The capacity and the cost of link l, $\tilde{\psi}_l$ and $\tilde{C}_l$ respectively, are:

$$\tilde{\psi}_l = \sum_{t \in T} u_{lt} \psi_{lt}, \qquad \tilde{C}_l = \sum_{t \in T} u_{lt} C_{lt} \tag{1}$$

There is no linear relationship assumed between C_{lt} and ψ_{lt}, therefore the link cost is not necessarily a linear function of the link capacity. For an EF demand $m, m \in M_k$, we differentiate between the average arrival rate, α_{km}^{ef}, and the requested bandwidth, ρ_{km}^{ef}. ρ_{km}^{ef} is usually a value between the average arrival

Network	
L	set of links in the network
K	set of (both EF and BE) Origin-Destination (O-D) pairs
T_l	index of available link types for link $l \in L$
u_{lt}	link type decision variable; 1 if link type t is used for link $l \in L$, 0 other wise
ψ_{lt}	size of the capacity of link type $t, t \in T_l$
$\tilde{\psi}_l$	total capacity of link l, $l \in L$
C_{lt}	cost of the link type t, $t \in T_l$, $l \in L$
$\tilde{C}_l$	total cost of link l, $l \in L$

Candidate paths and routing	
J_k	set of possible candidate LSP paths for the O-D pair k, $k \in K$
δ_j^l	link-path indicator; 1 if path j uses link l, $j \in J_k, k \in K, l \in L$, 0 otherwise
B_{kj}	set of candidate backup paths, which are link disjoint from $j, j \in J_k$
δ_b^l	link-path indicator; 1 if backup path b uses link l, $b \in B_{kj}, j \in J_k, k \in K, l \in L$, 0 otherwise
x_{kmj}^{ef}	EF primary path routing variable; 1 if EF demand $m, m \in M_k$, $k \in K$ uses path $j \in J_k$, 0 otherwise
x_{kj}^{be}	BE primary path routing variable: the portion of BE demand k that uses candidate path $j, j \in J_k$. x_{kj}^{be} can be any real value between 0 and 1
$\hat{x}_{jkb}^{ef}, \hat{x}_{jkb}^{be}$	backup path routing variable, 1 if backup path $b \in B_{kj}$ is used by path $j \in J_k$

EF demand related	
M_k	set of EF demands for O-D pair k, $k \in K$
α_{km}^{ef}	average arrival rate of an EF traffic demand m, $m \in M_k$, $k \in K$
ρ_{km}^{ef}	requested bandwidth of EF traffic demand m, $m \in M_k$, $k \in K$
η_l	total requested bandwidth of EF demand on link $l \in L$
β_l^{ef}	average arrival rate of total EF traffic demand on link $l \in L$

BE demand related	
α_k^{be}	average arrival rate of a BE traffic demand, $k \in K$
β_l^{be}	average arrival rate of total BE traffic demand on link $l \in L$
d_l	average delay experienced by BE traffic on link $l \in L$
d_{lmax}	maximum value of d_l allowed for link $l \in L$
r	BE traffic restoration level
g_l	BE delay bound factor
$\tilde{y}$,$\widetilde{y^2}$	the first and second moment of packet size, (units: bits & bits^2)

Table 1: Notation

rate and the peak rate. It is noted in [11] that the packets of the EF traffic class belonging to the same flow should not be reordered. Consequently, traffic from the same EF demand can not be separated into different LSPs. For the same reason, only one backup path will be used per EF user demand. Given a failure of link F, $F \in L$, the total amount of EF user demand on link l is:

$$\eta_l = \sum_{k \in K} \sum_{m \in M_k} \rho_{km}^{ef} \sum_{j \in J_k} (x_{kmj}^{ef} \delta_j^l + \sum_{b \in B_{kj}} x_{kmj}^{ef} \delta_j^F \hat{x}_{jkb}^{ef} \delta_b^l) \tag{2}$$

where $(x_{kmj}^{ef} \delta_j^l)$ equals 1 if the EF demand m uses LSP j as its working path and the path j includes link l,

and 0 otherwise, while $(x^{ef}_{kmj}\delta^F_j\hat{x}^{ef}_{jkb}\delta^l_b)$ equals 1 only if the working path of the EF demand m is interrupted by the link failure F and the backup path b uses link l.

Similarly, average arrival rate of total EF traffic on link l can be calculated according to the following equation:

$$\beta^{ef}_l = \sum_{k\in K}\sum_{m\in M_k}\alpha^{ef}_{km}\sum_{j\in J_k}(x^{ef}_{kmj}\delta^l_j + \sum_{b\in B_{kj}} x^{ef}_{kmj}\delta^F_j\hat{x}^{ef}_{jkb}\delta^l_b) \tag{3}$$

For each O-D pair k, only one BE demand pair is defined, which can be looked as the aggregation of multiple BE demands with the same O-D. Unlike EF traffic, we allow the traffic within a single BE demand to split arbitrarily across any number of candidate LSPs, therefore the aggregation of BE traffic would potentially improve the effectiveness of traffic engineering. However, we assume only one backup path will be used for traffic on the same working path. We assume that the BE traffic may not be fully restorable in case of link failure. BE traffic restoration level r, a value between [0,1], represents the proportion of BE traffic on each link that is being protected. The ability to set the restoration level of BE traffic and accordingly differentiate the resilience level provides a powerful tool to the service provider. It allows the ISP to clearly demonstrate the advantage of premium service based on EF class, and also allows the ISP to fine tune the redundancy level and total network cost.

The total BE load (average arrival rate) on link l is:

$$\beta^{be}_l = \sum_{k\in K}\alpha^{be}_k\sum_{j\in J_k}(x^{be}_{kj}\delta^l_j + r\sum_{b\in B_{kj}} x^{be}_{kj}\delta^F_j\hat{x}^{be}_{jkb}\delta^l_b), \quad \forall F\in L \tag{4}$$

It is assumed that the requested bandwidth of a EF user demand is defined in a way such that the performance of an EF demand will be guaranteed if it is given the requested bandwidth.

$$\eta_l \leqslant \tilde{\psi}_l, \quad \forall l\in L \tag{5}$$

There are many discussions regarding the limits on the link utilization of EF user demands [9]. The revised EF PHB, RFC3246 [11], introduces an error term E_a for the treatment of the EF aggregate, which represents the allowed worst case deviation between the actual EF packet departure time and the ideal

departure time of the same packet. This revision makes it possible for the EF utilization to go up as high as 100% [37]. In this paper, we assume that there are no additional constraints on the EF utilization.

We pick the average delay as the sole performance measurement for BE traffic in this paper as suggested by [24]. We evaluate the performance of BE traffic on a per-link basis (i.e., not end-to-end). The value $\frac{\tilde{y}}{\tilde{\psi}_l}$ stands for the average transmission delay of packets. We use $\frac{\tilde{y}}{\tilde{\psi}_l}$ as the basis for the delay bound. Let $d_{lmax} = g_l \frac{\tilde{y}}{\tilde{\psi}_l}$, where g_l is a parameter defined by the network designer. The larger the value of g_l, the more bandwidth is required for link l, therefore the lower the link utilization. We assume that the performance of BE traffic is satisfactory if $d_l \leqslant d_{lmax}$. Note that the delay bound d_{lmax} remains the same under the normal condition and the failure scenarios.

Every router is modeled as a M/G/1 system with Poisson packet arrivals and an arbitrary packet length distribution. [30] concludes, through both simulation and analytic study, that even though the traffic exhibits bursty behavior at certain time scales, the relationship between the variance and the mean is approximately linear, or "Poisson-like", if they are measured over larger time scales, where the traffic can be treated as if it were smooth. Our choice of the Poisson arrival model is justified because we are more concerned about the average BE performance over a large time scale for capacity planning purposes.

From the average queueing delay formula of the priority queue [21], we obtain the performance constraint for BE traffic:

$$\begin{aligned} d_l &= \frac{\tilde{y}}{\tilde{\psi}_l} + \frac{\widetilde{y^2}}{2\tilde{y}} \frac{\beta_l^{ef} + \beta_l^{be}}{(\tilde{\psi}_l - \beta_l^{ef})(\tilde{\psi}_l - \beta_l^{ef} - \beta_l^{be})} \\ &\leqslant g_l \frac{\tilde{y}}{\tilde{\psi}_l} \end{aligned} \tag{6}$$

With some rearrangement, (6) yields

$$\tilde{\psi}_l \geqslant f(\beta_l^{ef}, \beta_l^{be}) \tag{7}$$

where

$$f(\beta_l^{ef}, \beta_l^{be}) = \beta_l^{ef} + \frac{\beta_l^{be}}{2} + \frac{\widetilde{y^2}(\beta_l^{ef} + \beta_l^{be})}{4(\tilde{y})^2(g_l - 1)} + \frac{1}{2}\sqrt{(2\beta_l^{ef} + \beta_l^{be} + \frac{\widetilde{y^2}(\beta_l^{ef} + \beta_l^{be})}{2(\tilde{y})^2(g_l - 1)})^2 - 4\beta_l^{be}(\beta_l^{be} + \beta_l^{ef})}.$$

In order to have a meaningful solution for constraint (7), $\tilde{\psi}_l > \beta_l^{ef} + \beta_l^{be}$ is required.

3 Problem Formulation

The formal problem definition is presented below.

Given: $L, K, T_l, \psi_{lt}, C_{lt}, J_k, \delta_j^l, B_{kj}, \delta_b^l, M_k, \alpha_{km}^{ef}, \rho_{km}^{ef}, \alpha_k^{be}, d_{lmax}, r, g_l, \tilde{y}, \widetilde{y^2}$

Variable: $u_{lt}, x_{kmj}^{ef}, x_{kj}^{be}, \hat{x}_{jbk}^{ef}, \hat{x}_{jbk}^{be}$

Goal:

$$\min \sum_{l \in L} \tilde{C}_{lt}$$

Subject to:

$$u_{lt} = 0/1, \sum_{t \in T_l} u_{lt} = 1 \tag{8}$$

$$x_{kmj}^{ef} = 0/1, \sum_{j \in J_k} x_{kmj}^{ef} = 1 \tag{9}$$

$$\sum_{j \in J_k} x_{kj}^{be} = 1 \tag{10}$$

$$\hat{x}_{jkb}^{ef} = 0/1, \sum_{b \in B_{kj}} \hat{x}_{jkb}^{ef} = 1 \tag{11}$$

$$\hat{x}_{jkb}^{be} = 0/1, \sum_{b \in B_{kj}} \hat{x}_{jkb}^{be} = 1 \tag{12}$$

$$\tilde{\psi}_l \geqslant \eta_l \tag{13}$$

$$\tilde{\psi}_l \geqslant f(\beta_l^{ef}, \beta_l^{be}) \tag{14}$$

(8) imposes a discrete constraint on the link capacities. Constraint (9) ensures that all traffic from one EF O-D pair will follow one single path. (11)(12) denotes that only one backup path will be used for each working path. Constraint (13)(14) ensures the performance of EF and BE traffic respectively.

Because $\tilde{C}_l$ is non-decreasing with respect to β_l^{ef}, η_l, and β_l^{be}, this problem can be reformulated as:

$$\min \sum_{l \in L} \tilde{C}_l \tag{15}$$

Subject to (8)(9)(10)(11)(12) (13)(14) and:

$$\eta_l \geqslant \sum_{k\in K}\sum_{m\in M_k}\rho_{km}^{ef}\sum_{j\in J_k}(x_{kmj}^{ef}\delta_j^l + \sum_{b\in B_{kj}} x_{kmj}^{ef}\delta_j^F\hat{x}_{jkb}^{ef}\delta_b^l) \tag{16}$$

$$\beta_l^{ef} \geqslant \sum_{k\in K}\sum_{m\in M_k}\alpha_{km}^{ef}\sum_{j\in J_k}(x_{kmj}^{ef}\delta_j^l + \sum_{b\in B_{kj}} x_{kmj}^{ef}\delta_j^F\hat{x}_{jkb}^{ef}\delta_b^l) \tag{17}$$

$$\beta_l^{be} \geqslant \sum_{k\in K}\alpha_k^{be}\sum_{j\in J_k}(x_{kj}^{be}\delta_j^l + r\sum_{b\in B_{kj}} x_{kj}^{be}\delta_j^F\hat{x}_{jkb}^{be}\delta_b^l) \tag{18}$$

We refer to the problem defined by (15,8,9,10,11,12,13,14,16, 17, 18) as problem (P) in the rest of this paper. As can be seen from the above problem formulation, problem (P) is a non-linear integer programming problem, which is considered to be at least NP hard [2].

4 Solution Method

4.1 Lagrangean Relaxation

Lagrangean Relaxation is a common technique for multicommodity flow problems [2]. It has been successfully applied to the capacity planning and routing problems [16] [17] [3] [25] [26], and our previous work [33][34]. We describe its use for our problem in this section. Please refer to [32] for more detail.

Using Lagrangean Relaxation, relax (16)(17) and (18), and we have the Lagrangean as:

$$\begin{aligned} & L(x_{kmj}^{ef}, x_{kj}^{be}, \hat{x}_{jbk}^{ef}, \hat{x}_{jbk}^{be}, u_{lt}, \lambda_l^{ef}, \lambda_l^{\eta}, \lambda_l^{be}) \\ = & \sum_l(\tilde{C}_l - \lambda_l^{\eta}\eta_l - \lambda_l^{ef}\beta_l^{ef} - \lambda_l^{be}\beta_l^{be}) \\ & + \sum_{k\in K}\sum_{m\in M_k}\sum_{j\in J_k} x_{kmj}^{ef}\sum_{l\in L}(\lambda_l^{\eta}\rho_{kj}^{ef} + \lambda_l^{ef}\alpha_{km}^{ef})(\delta_j^l + \sum_{b\in B_{kj}}\delta_j^F\hat{x}_{jkb}^{ef}\delta_b^l) \\ & + \sum_{k\in K}\sum_{j\in J_k} x_{kj}^{be}\sum_{l\in L}\lambda_l^{be}\alpha_k^{ef}(\delta_j^l + r\sum_{b\in B_{kj}}\delta_j^F\hat{x}_{jkb}^{be}\delta_b^l) \end{aligned} \tag{19}$$

The Lagrangean dual problem (D) is then:

$$\max_{\lambda_l^{\eta},\lambda_l^{ef},\lambda_l^{be}\geqslant 0} h(\lambda_l^{\eta}, \lambda_l^{ef}, \lambda_l^{be}) \tag{20}$$

where:

$$h(\lambda_l^{\eta}, \lambda_l^{ef}, \lambda_l^{be}) = \min_{x_{kmj}^{ef}, x_{kj}^{be}, \hat{x}_{jbk}^{ef}, \hat{x}_{jbk}^{be}, u_{lt}} L(x_{kmj}^{ef}, x_{kj}^{be}, \hat{x}_{jbk}^{ef}, \hat{x}_{jbk}^{be}, u_{lt}, \lambda_l^{ef}, \lambda_l^{\eta}, \lambda_l^{be}) \tag{21}$$

Since $\eta_l, \beta_l^{ef}, \beta_l^{be}$ and $x_{kmj}^{ef}, x_{kj}^{be}, \hat{x}_{jkb}^{ef}, \hat{x}_{jkb}^{be}$ are independent variables,

$$\begin{aligned} \min L \;=\; & \sum_l \min(\tilde{C}_l - \lambda_l^{\eta}\eta_l - \lambda_l^{ef}\beta_l^{ef} - \lambda_l^{be}\beta_l^{be}) \\ & + \sum_{k\in K}\sum_{m\in M_k} \min[\sum_{j\in J_k} x_{kmj}^{ef} \sum_{l\in L}(\lambda_l^{\eta}\rho_{kj}^{ef} + \lambda_l^{ef}\alpha_{km}^{ef})(\delta_j^l + \sum_{b\in B_{kj}} \delta_j^F \hat{x}_{jkb}^{ef}\delta_b^l)] \\ & + \sum_{k\in K} \min[\sum_{j\in J_k} x_{kj}^{be} \sum_{l\in L} \lambda_l^{be}\alpha_k^{ef}(\delta_j^l + r \sum_{b\in B_{kj}} \delta_j^F \hat{x}_{jkb}^{be}\delta_b^l)] \end{aligned} \tag{22}$$

4.2 Solving the Subproblems

Equation (22) shows that the problem (21) can be separated into the following three set of subproblems:

Subproblem (i):

$$\min(\tilde{C}_l - \lambda_l^{\eta}\eta_l - \lambda_l^{ef}\beta_l^{ef} - \lambda_l^{be}\beta_l^{be}) \tag{23}$$

$$\text{Subject to: } (8)(13)(14) \tag{24}$$

Subproblem (i) can be solved by the gradient projection method [7].

Subproblem (ii):

$$\min[\sum_{j\in J_k} x_{kmj}^{ef} \sum_{l\in L}(\lambda_l^{\eta}\rho_{kj}^{ef} + \lambda_l^{ef}\alpha_{km}^{ef})(\delta_j^l + \sum_{b\in B_{kj}} \delta_j^F \hat{x}_{jkb}^{ef}\delta_b^l)] \tag{25}$$

Since we have a small set of candidate path for each O-D, we can simply set the link cost to $(\lambda_l^{\eta}\rho_{kj}^{ef} + \lambda_l^{ef}\alpha_{km}^{ef})$, and then enumerate the choice of working paths and backup paths, picking the combination of working and backup paths with the least total cost. The resulting choice of working path and backup path is cached. In the subsequent iteration, only paths whose cost is changed will be compared to the previous solution, and hence save the running time.

Subproblem (iii):

$$\min[\sum_{j\in J_k} x_{kj}^{be} \sum_{l\in L} \lambda_l^{be}\alpha_k^{ef}(\delta_j^l + r \sum_{b\in B_{kj}} \delta_j^F \hat{x}_{jkb}^{be}\delta_b^l)] \tag{26}$$

Similar to Subproblem (ii), the solution is to enumerate the choice of working path and backup path, then choose the one with the least combined cost. The results are similarly cached to speed up the program.

4.3 Subgradient Method

The subgradient method is used to update λ_l^{η}, λ_l^{ef} and λ_l^{be}. For an in depth description of the subgradient method and its variants, please refer to [6].

Given the initial values of λ_l^{η}, λ_l^{ef} and λ_l^{be}, once we solve the problem (D), subgradients for the three set of Lagrangeans are computed as follows:

$$\omega_l^{\eta} = \eta_l - \sum_{k\in K}\sum_{m\in M_k}\rho_{km}^{ef}\sum_{j\in J_k}(x_{kmj}^{ef}\delta_j^l + \sum_{b\in B_{kj}}x_{kmj}^{ef}\delta_j^F\hat{x}_{jkb}^{ef}\delta_b^l) \tag{27}$$

$$\omega_l^{ef} = \beta_l^{ef} - \sum_{k\in K}\sum_{m\in M_k}\alpha_{km}^{ef}\sum_{j\in J_k}(x_{kmj}^{ef}\delta_j^l + \sum_{b\in B_{kj}}x_{kmj}^{ef}\delta_j^F\hat{x}_{jkb}^{ef}\delta_b^l) \tag{28}$$

$$\omega_l^{be} = \beta_l^{be} - \sum_{k\in K}\alpha_k^{be}\sum_{j\in J_k}(x_{kj}^{be}\delta_j^l + r\sum_{b\in B_{kj}}x_{kj}^{be}\delta_j^F\hat{x}_{jkb}^{be}\delta_b^l) \tag{29}$$

To improve the convergence speed, the surrogate subgradient method is used, where only one of the three subproblems is solved at a time. The subsequent Lagrangean multipliers are updated using the latest value.

$$\lambda_l^{\eta} \leftarrow \min(0, \lambda_l^{\eta} + t\omega_l^{\eta}), \forall l \in L \tag{30}$$

$$\lambda_l^{ef} \leftarrow \min(0, \lambda_l^{ef} + t\omega_l^{ef}), \forall l \in L \tag{31}$$

$$\lambda_l^{be} \leftarrow \min(0, \lambda_l^{be} + t\omega_l^{be}), \forall l \in L \tag{32}$$

where the step size, t, is defined by:

$$t = \phi\frac{h^* - h}{||\omega_l^{eta}||^2 + ||\omega_l^{ef}||^2 + ||\omega_l^{be}||^2} \tag{33}$$

where h^* is the value of the best feasible solution found so far, and ϕ is a scalar between 0 and 2. ϕ is set to 2 initially in our study and is halved if the solution does not improve in 10 iterations.

At each iteration, the solution of $x^{ef}_{kmj^*}, x^{be}_{kj^*}, \hat{x}^{ef}_{jkb}$, and $\hat{x}^{ef}_{jkb}$ for the primal problem (P) can be obtained from the solution of subproblem (ii) and(iii). The link capacity $\tilde{\psi}_l$ is computed according to (13). Consequently, the primary objective function can be derived. As the iteration proceeds, we store the best solution found so far for the primal problem (P). In this way, we are always able to obtain a feasible solution. The maximum number of iterations is set to 400 in the implementation [33][2].

The solution of the dual problem provides a lower bound for the primal problem. Therefore, the solution quality can be assessed by the duality gap, which is the difference between the solutions of problem (P) and problem (D). Note that because the duality gap is always no smaller than the actual difference between the obtained feasible solution and the optimal solution, it is a conservative estimate of the solution quality.

5 Computational Results

In this section, we present numerical results based on experimentation. The objective of our experiment is to evaluate the solution quality and running time of the algorithm. The program is implemented in C and the computation is performed on a Pentium IV 2.4GHz PC with 512M memory, running Redhat Linux 7.2.

The network topologies are generated using the Georgia Tech Internetwork Topology Models (GT-ITM) [38]. The locations of origins and destinations are randomly selected. For each O-D pair, 10 candidate working paths are calculated using Yen's K-shortest path algorithm [36]. 10 candidate backup paths are also generated for each candidate working path.

If not specified, EF demand pairs are randomly generated with the average rate uniformly distributed from 0 Mbps to 10 Mbps. The requested bandwidth of a EF demand pair is a value between the average rate and the peak rate. It is randomly specified to be 150%-300% of the average rate in our experiments. The average rate of BE demand is uniformly distributed from 10Mbps to 50Mbps. The number of EF demands for the same O-D pair is uniformly distributed from 1 to 10. The number of candidate link types

for link l is uniformly distributed from 5 to 10. The capacities of link types are set to be multiples of 45Mbps, while the costs of the link types for link l are randomly generated in such a way that the cost goes higher and the unit cost per Mbps goes down as the link capacity increases. We use average packet size $\tilde{y} = 4396$ bits and second moment of packet size $\tilde{y^2} = 22790170$ bits^2 for all the test cases. They are calculated based on a traffic trace (AIX-1014985286-1) from the NLANR Passive Measurement and Analysis project [1]. The restoration level of BE traffic is determined by the network designer, depending on the budget and the desired BE class resilience. It is set to 0.5 in all of the experiments.

The BE delay bound factor, g_l, is set to 2 for all links in our experiments. In practice, g_l should be carefully chosen to reflect the desired BE performance and the expected link utilization. Figure (1) shows the figure of $f(\beta_{ef}^l)$ with respect to β_{ef}^l for various values of g_l. As can be seen from the figure, the link utilization varies significantly as g_l changes. The average link utilization, the value of $\frac{\beta_{ef}^l + \beta_{be}^l}{f(\beta_{ef}^l)}$, is about 60% when g_l equals 2.

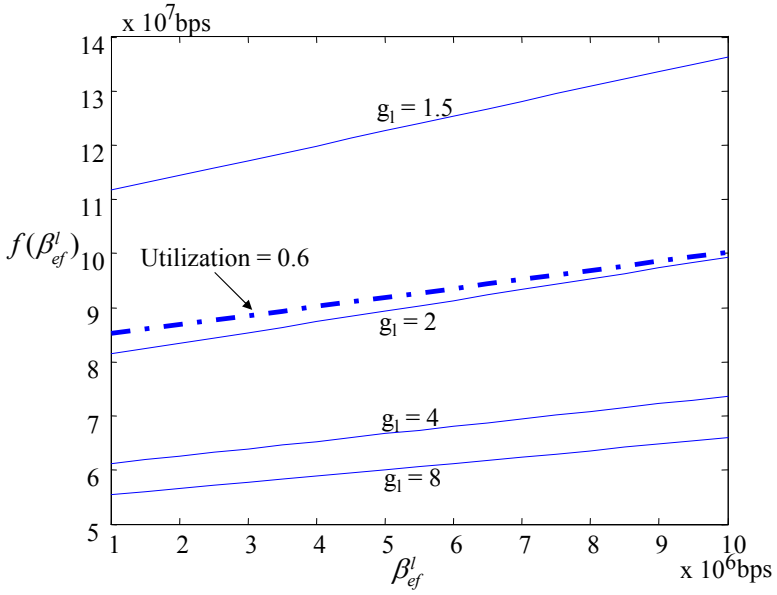

Figure 1: $f(\beta_{ef}^l)$ vs. β_{ef}^l, $(\beta_{be}^l = 5 \times 10^7$ bps)

Two additional greedy type heuristics were implemented for the purpose of comparison:

1. The first heuristic method starts with one randomly chosen O-D pair k, finds out both the best working path j and backup path b for both EF and BE demand among all candidate working and backup paths that would result in the least total cost, and then repeats the process for all $k \in K$.

2. The second heuristic is an iterative method inspired by the heuristic used in [23]. The primary difference here is that [23] deals only with the spare capacity planning problem, while we have to find out the working path first. In the first iteration, for each O-D pair k, the algorithm finds out only the best working paths j for both EF and BE demands among candidate working paths that would result in the least total cost. After all the working paths have been determined, the algorithm looks for the best backup paths b for both EF and BE demands among candidate backup paths. In the subsequent iteration, the algorithm goes through each O-D pair to see whether any change of working path or backup path will lower the total cost. The iteration will continue as long as there is at least one adjustment of path in the current iteration.

We will call these two methods (H) and (S), respectively.

The algorithms were tested on 8 different sizes of networks, ranging from 10 nodes to 1000 nodes. Some details of the network topologies are listed in Table 2. Note that the O-D demand number shown in Table 2 includes both EF and BE demands. To obtain confidence intervals, we generated 30 different topologies for each network size, with the same number of nodes, links, and O-D pairs.

$$\text{Duality Gap} \;=\; \left|\frac{s_p - s_d}{s_d}\right| \tag{34}$$

The solution quality is represented by the duality gap, which is the percentage difference between the solution of the primal problem and the dual problem. s_p and s_d are the solutions of primal problem and dual problem respectively. A value close to zero means the solution is very close to optimal.

Table 2 shows the Duality Gaps of the three solution methods, expressed in terms of the 95% confidence intervals. It is easy to see from the table that the Lagrangian Relaxation together with the subgradient

Node Number	Link Number	O-D Number	Duality (%)		
			LR	H	S
10	25	30	0.12-3.24	3.82-21.89	0-3.40
20	50	90	0.01-2.96	0.2-15.21	0-2.3
50	125	350	0.17-1.78	1.7-13.70	0-1.87
100	250	1000	0.16-1.90	0-8.36	0-1.4
200	500	3000	0-1.64	0-10.21	0-1.55
500	1250	12000	0-2.59	0-11.47	
700	1750	20000	0.2.82	0-9.35	
1000	2500	40000	0-1.97	0-15.09	

Table 2: Network topology information and experimental results

method produces reasonable results as the duality gap is bounded by no more than 3.3%. Note the primal problem itself is approximated when reducing the size of candidate path set for all possible path set. But according to our experimental results, more than 97% of the time, the final solution is chosen among the 5 shortest candidate paths. Therefore, 10 candidate paths are considered adequate. Having more than 10 candidate paths will have minimal impact on the solution quality, while significantly increasing the running time. Given the large number of networks being tested, we are confident that the solution should have good quality for other sizes of networks. Heuristic (S) achieves comparable solution quality in networks sized up to 200 nodes, while method (H) yields solutions with a duality gap of more than 20%, which is not very desirable.

Figure 2 shows the running time with respect to the network size. In all 240 test cases of Lagrangian Relaxation based method, the algorithm converges without difficulty. As can be seen from the figure, heuristic (S) takes much longer time than the other two methods.

From the experiments shown above, Lagrangian Relaxation based approach demonstrates a good trade-off between the solution quality and the running time. The size of the largest network evaluated in this paper is representative of a large network, and is much larger than the test cases used in most work on network design. It is fair to predict that the running time of the algorithm will stay reasonable for practical sized networks.

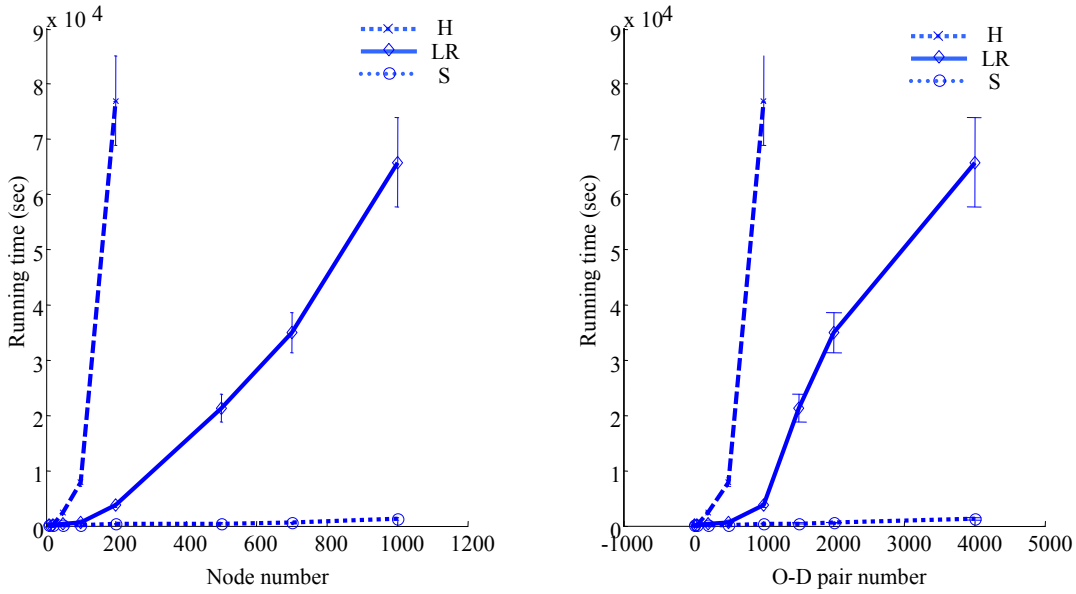

Figure 2: Running Time vs. Network Size

The flow aggregation based lagrangian Relaxation method proposed in this paper can be applied to more general nonlinear integer multicommodity flow problems. We are seeking the opptunitues to extend the method to other types of network design problems, such as design of optical networks and overlay networks.

6 Conclusions and Future Directions

In this paper, we addressed the problem of link dimensioning and routing for survivable MPLS networks supporting DiffServ EF and BE traffic. We formulate the problem as an optimization problem, where the total link cost is minimized, subject to the performance constraints of both EF and BE classes under normal circumstance and single link failures. The variable here is the routing of working and backup LSP of both EF and BE user demands, and the discrete link capacities.

This paper presents a preliminary investigation of the capacity planning issue for survivable multiclass MPLS networks. The novelty of the problem presented in this paper is that it involves two traffic classes,

EF and BE, which have totally different forms of performance requirements and results in non-linear performance constraint. Unlike SCA problems, this paper deal with the the working and backup paths at the same time, which is considered to be much more complex. The combination of targeting multiple traffic classes with non-linear performance constraints, optimizing working path and backup path simultaneously, and the use of non-linear cost function produces a problem more difficult than most related works.

We presented a Lagrangian Relaxation-based method to effectively decompose the original problem. A subgradient method is used to find the optimal Lagrangian multiplier. We investigated experimentally the solution quality and running time of this approach. The results from our experiments indicate that our method produces solutions that are within a few percent of the optimal solution, while the running time stays reasonable for practical sized networks.

The problem formulation and solution approaches may be applied to other traffic classes, such as the Assured Forwarding (AF) class. The technique of relaxing flow balance equations enables the Lagrangian Relaxation techniques to be applied to difficult integer non-linear programming problems, which is not possible otherwise. We are investigating the adaptation of this technique to the general non-linear multi-commodity flow problems.

References

[1] NLANR passive measurement and analysis project, http://pma.nlanr.net, 2001.

[2] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows, Theory, Algorithms and Applications.* Prentice Hall, New Jersey, 1993.

[3] A. Amiri. A System for the Design of Packet-Switched Communication Networks with Economic Tradeoffs. *Computer Communications*, 21(18):1670–1680, Dec. 1998.

[4] Shigehiro Ano, Toru Hasegawa, and Nicolas Decre. Experimental TCP performance evaluation on diffserv AF PHBs over ATM SBR service. *Telecommunication Systems*, 19(3-4):425–441, Apr. 2002.

[5] Achim Authenrieth and Andreas Kirstadter. Components of MPLS recovery supporting differentiated resilience requirements. In *IFIP Workshop on IP and ATM Traffic Management WATM'2001*, Paris, France, September 2001.

[6] D.P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, MA, 1995.

[7] D.P. Bertsekas and R.G. Gallager. *Data Networks.* Prentice Hall, Englewood Cliffs, NJ, 1987.

[8] S. Blake and et al. An Architecture for Differentiated Service, IETF RFC 2475. Dec. 1998.

[9] A. Charny and J.-Y. Le Boudec. Delay Bounds in Network with Aggregate Scheduling. In *First International Workshop on Quality of Future Internet Service*, pages 1–13, Berlin, Germany, Sep. 2000.

[10] Hongsik Choi, Suresh Subramaniam, and Hyeong-Ah Choi. On double-link failure recovery in wdm optical networks. In *Proceeding of IEEE Infocom '02*, June 2002.

[11] B. David, A. Charny, and et Al. An Expedited Forwarding PHB, IETF RFC 3246. Mar. 2002.

[12] Reinhard Diestel. *Graph Theory*. Graduate Textbooks in Mathematics 173. Springer-Verlag, 2 edition, 2000.

[13] C. Dovrolis and P. Ramanathan. Resource aggregation for fault tolerance in integrated service networks. *ACM Computer Communication Review*, 28(2):39–53, 1998.

[14] C. Huang et Al. Building reliable MPLS networks using a path protection mechanism. *IEEE Communications Magazine*, pages 156–162, Mar. 2002.

[15] D. Awduche et al. Requirement for traffic engineering over MPLS, IETF RFC 2702. September 1999.

[16] B. Gavish and I. Neuman. Capacity and Flow Assignment in Large Computer Networks. In *IEEE Infocom '86*, pages 275–284. IEEE, Apr. 1986.

[17] B. Gavish and I. Neuman. A System for Routing and Capacity Assignment in Computer Networks. *IEEE Transactions on Communications*, 37(4):360–366, Apr. 1989.

[18] F. Giroire, A. Nucci, N. Taft, and C. Diot. Increasing the robustness of IP backbones in the absence of optical level protection. In *Infocom '03*. IEEE, May 2003.

[19] R. Iraschko, M. MacGregor, and W. Grover. Optimal capacity placement for path resotration in STM or ATM mesh survivable networks. *IEEE/ACM transcations on Networking*, 6(3):325–336, June 1998.

[20] S. Kim and K. G. Shin. Improving dependability of real-time communication with preplanned backup routes and spare resource pool. In *Proceeding of IWQoS 2003*, June 2003.

[21] L. Kleinrock. *Queueing Systems, Volume II: Computer Application*. Wiley Interscience, 1976.

[22] M. Kodialam and T.V. Lakshman. Dynamic routing of bandwidth guaranteed tunnels with resotration. In *Proceeding of IEEE Infocom '00*, Mar. 2000.

[23] Y. Liu and D. Tipper. Approximation optimal spare capacity allocation by successive survivable routing. In *Proceeding of IEEE Infocom '01*, Anchorage, AL, April 2001.

[24] Jim Martin and Arne Nilsson. On service level agreements for IP networks. In *IEEE Infocom 02'*, New York, 2002. IEEE.

[25] D. Medhi. Multi-Hour, Multi-Traffic Class Network Design for Virtual Path-based Dynamically Reconfigurable Wide-Area ATM Networks. *IEEE/ACM Transactions on Networking*, 3(6):809–818, Dec. 1995.

[26] D. Medhi and D. Tipper. Some Approaches to Solving a Multihour Broadband Network Capacity Design Problem with Single-path Routing. *Telecommunication Systems*, 13(2-4):269–291, 2000.

[27] S. Nelakuditi, S. Lee, Y. Yu, and Zhi-Li Zhang. Failure insensitive routing for ensuring service availability. In *Proceeding of IWQoS 2003*, June 2003.

[28] K. Nichols, V. Jacobson, and L. Zhang. A Two-bit Differentiated Services Architecture for the Internet, IETF RFC 2638. Jul. 1999.

[29] E. Rosen, A. Viswanathan, and R. Callon. Multiprotocol label switching architecture, IETF RFC 3031. January 2001.

[30] Xusheng Tian, Jie Wu, and Chuangyi Ji. A unified framework for understanding network traffic using independent wavelet models. In *IEEE Infocom '02*, volume 1, pages 446–454, New York, 2002. IEEE.

[31] Paul Veitech and Dave Johnson. ATM network resilience. pages 26–23, September/October 1997.

[32] Kehang Wu. *Flow Aggregation based Lagrangian Relaxation with Applications to Capacity Planning of IP Networks with Multiple Classes of Service*. PhD thesis, North Carolina State University, Raleigh, 2004.

[33] Kehang Wu and Douglas S. Reeves. Capacity planning of DiffServ networks with Best-Effort and Expedited Forwarding traffic. In *IEEE 2003 International Conference on Communications*, Anchorage, Alaska, May 2003.

[34] Kehang Wu and Douglas S. Reeves. Link Dimensioning and LSP Optimization for MPLS Networks Supporting DiffServ EF and BE traffic classes, 2003.

[35] Tsong-Ho Wu. *Fiber Network Service Survivability.* Artech House, 1992.

[36] J.Y. Yen. Finding the K Shortest Loopless Paths in a Network. *Management Science*, 17(11):712–716, Jul. 1971.

[37] Jiang Yuming. Delay bounds for a network of guaranteed rate servers with FIFO aggregation. *Computer Networks*, 40(6):683–694, December 2002.

[38] E.W. Zegura, K. Calvert, and M. Jeff Donahoo. A Quantitative Comparison of Graph-based Models for Internet Topology. *IEEE/ACM Transactions on Networks*, 5(6):770–783, Dec. 1997.