Dynamic Routing Optimization in WDM Networks

J. Crichigno*, N. Ghani†, J. Khoury‡, W. Shu†, M. Y. Wu‡

*Engineering Department, Northern New Mexico College, Espanola - NM, USA
†Electrical & Computer Engineering Dept., University of New Mexico, Albuquerque - NM, USA
‡Dept. of Computer Science & Engineering, Shanghai JiaoTong University, Shanghai, China
jcrichigno@nnmc.edu, {nghani, jkhoury, shu}@ece.unm.edu, wu-my@cs.sjtu.edu.cn

Abstract—We present a multi-objective optimization approach for joint throughput optimization and traffic engineering, where the routing request of traffic arrives one-by-one. We provide an Integer Linear Program (ILP) that simultaneously i) maximizes the aggregate throughput, ii) minimizes the resource consumption, and iii) minimizes the maximum link utilization. We study the impact of optimizing the three different objectives simultaneously in dynamic environments, and show that better solutions than those of mono-objective approaches can be obtained. Because of the complexity of the ILP, we also propose another ILP with reduced complexity, and study its performance and the optimality gap between it and optimal solutions.

Index Terms—WDM networks, integer linear programming, multi-objective optimization.

I. INTRODUCTION

Dynamic throughput optimization and traffic engineering problem in Wavelength-Division Multiplexing (WDM) networks refers to the problem of set up paths between nodes in the network, called optical cross-connects, such that the number of connection requests successfully routed (aggregate throughput) are maximized while resource consumption and maximum link utilization (MLU) are minimized.

Routing in WDM networks has been an active research area for several years. Much of the work is focused on maximizing the aggregated throughput in terms of accepted connections, which is similar to minimizing the blocking probability [1]. The latter is defined as the ratio of the rejected connections to the total number of connection requests. Other approaches consider traffic engineering, which attempt to balance the load of the network and to improve the resource consumption. Load balancing is obtained by minimizing the utilization of the most heavily used link in the network, or MLU. This objective is fundamental to minimize the link utilization throughout the network, so that no bottleneck link exists. Although minimizing the MLU avoid hot-spots, the total resource consumption or sum of assigned wavelengths at links of paths may be higher than needed.

The problem is further motivated by the fact that there may exist the possibility of having new routing requests in the future (dynamically), which cannot be excluded. Thus, the routing scheme must be an on-line algorithm capable of handling requests in an optimal manner when the requests are not all present at once. Since traffic requests may not be known in advance, the optimal solution may need to re-establish paths for previous requests and re-optimization may be needed whenever a new request arrives or departs.

In this paper, we present a multi-objective Integer Linear Program (ILP) which jointly considers the throughput optimization and the traffic engineering problems in wavelength convertible WDM network. In this kind of networks, converter devices are available at any node, and end-to-end traffic requests must be routed such that no two connections use the same wavelength on a same link. We assume that a traffic request is composed of a source node, destination node, and traffic demand. The proposed program simultaneously maximizes the aggregated throughput, minimizes the resource consumption and MLU. The ILP can generate optimal solutions for practical sized networks in a reasonable time. For those scenarios where path disruptions are not permitted, we also propose the application of another ILP, which works on an on-demand basis. With this approach, the optimization metrics quantify the solution for the new traffic request only according to the current state of the network, avoiding full re-optimization and path disruptions.

The paper is organized as follows. Section II discusses related work. Section III formulates the joint throughput optimization and traffic engineering problem in WDM networks as an ILP. Section IV presents a suboptimal alternative ILP with reduced complexity, which avoids path disruptions and full re-optimization. Section V shows performance studies, and Section VI concludes our work.

II. RELATED WORK

The throughput optimization problem in WDM networks can be seen as an extension of the multi-commodity flow problem in a graph, where flows and link capacities take integer values. Flows represent connection requests, while a link capacity represents the number of wavelengths available at the corresponding link. Banerjee et al. [1] presented an ILP that minimizes the average number packet hop distance. The authors claimed that optimizing the above objective is equivalent to maximizing throughput (i.e., the number of accepted connections). However, in general these two metrics may conflict with each other. Krishnaswamy et al. [2] provided two ILPs for the routing problem in WDM networks, and relaxed LP formulations with proper rounding techniques to approximate to optimal solutions. Cavendish et al. [3] studied the routing and wavelength assignment (RWA) problem in WDM mesh networks with full wavelength conversion capabilities, and attempted to minimize multiple objectives such as the number of wavelength converters, the number of hops
and the number of wavelengths used along the path; yet the authors just optimized one metric at a time. Belgacem et al. [4] proposed a heuristic to solve large instances of the throughput optimization problem in WDM networks. The heuristic partitions the problem into several smaller ILP subproblems that are optimally solved. Jaekel et al. [5] presented two ILP formulations which minimize the MLU and the resources used to route the connection request. Christodoulopoulos et al. [6] presented an ILP to optimize the number of wavelengths to satisfy certain connection requests, and a heuristic to solve it in a polynomial time. This problem is a dual of the throughput optimization; the goal is the minimization of the number of wavelengths instead of the maximization of the throughput. Note that the latter assumes that the topology is fixed, and the number of wavelengths per link is already known. A traffic engineering scheme which minimizes congestion is proposed in [7]. To achieve this objective, traffic engineering schemes attempt to minimize the MLU. In our previous work [8], we presented an ILP for the joint throughput optimization and traffic engineering problem for static scenarios, where all the traffic requests are routed at once. Performance studies showed that, by considering the multiple objectives simultaneously, better solutions can be obtained for static cases. This paper extends the work in [8] to dynamic scenarios, where traffic requests arrive one-by-one, and presents two ILPs schemes for these scenarios.

### III. JOINT THROUGHPUT OPTIMIZATION AND TRAFFIC ENGINEERING PROBLEM

#### A. Problem Formulation

We represent a wavelength convertible WDM network as a graph $G = (V, E)$, where $V$ is the set of nodes and $E$ the set of links. Let $N$ be the set of end-to-end flows or traffic demands. Each demand is characterized by a 5-tuple $(s_n, d_n, r_n, t^n_n, c'n_n)$, which denotes the source node, the destination node, the number of connection requests, the start time, and the end time of the demand. Flows are not known in advance, but arrives dynamically on time. Let $x^n_{ij}$ be a variable representing the number of connection requests of the $n^{th}$ flow routed on link $(i, j)$, and $c_n$ be the capacity of the link $(i, j)$ in number of wavelengths. Define $f_n \leq r_n, n \in N$, as the number of connection requests allocated to flow $n$. Note that in typical throughput optimization problems, $r_n$ is not given beforehand, and the objective is to maximize $f_n$. At a time $t$, the problem is defined as shown in Fig. 1.

The objective function $F$ consists of three terms with their corresponding weights $w_1, w_2$ and $w_3$. The first term is the aggregated throughput. The second term is the resource consumption (i.e., wavelengths) utilized by the scheme, and the third term, $\alpha$, is the MLU. Minimizing the last two objectives is equivalent to maximizing the negative of them. Eq. (1) is the flow conservation constraint, and Eq. (2) is the link capacity constraint, where capacities are multiplied by the MLU. Eq. (3) restricts the number of connection requests allocated to flow $n$ to be no more than its requested number of connections $r_n$. This constraint permits us to upper bound the number of connections allocated to any flow $n$. Eqs. (4) and (5) restrict the variables to be positive integers, and Eq. (6) states that the MLU is between 0 and 1.

ILP1 can be considered as a multi-objective problem, where the objective function $F$ at a time $t$ may be expressed as an objective vector $\vec{F} = (F_1, F_2, F_3)$, where:

$$F_1 = \sum_{n | t^n_n \leq t} f_n \quad \text{(throughput)}$$

$$F_2 = -\sum_{n | t^n_n \leq t} \sum_{(i,j) \in E} x^n_{ij} \quad \text{(resource consumption)}$$

$$F_3 = -\alpha \quad \text{(MLU)}$$

The relative importance of each objective can be varied according to the weight vector $\vec{w} = (w_1, w_2, w_3)$. We will consider the routing of all connection requests as the main goal. Thus, we will set $w_1 > w_2$ and $w_1 > w_3$.

In next sections, we will use Pareto optimality concepts [9]. Let $\vec{w}(a), \vec{w}(b)$ be two weight vectors, and $\vec{F}(a), \vec{F}(b)$ be the objective vectors corresponding to the solutions of ILP1 with $\vec{w}(a)$ and $\vec{w}(b)$ respectively. We will say that the solution obtained by setting ILP1 with $\vec{w}(a)$ is better than that with $\vec{w}(b)$ if and only if (iff) $F_i(a) > F_i(b)$ for some $i \in \{1, 2, 3\}$, and $F_i(a) \geq F_i(b)$ for the remainder objectives. We denote this relation as $\vec{F}(a) > \vec{F}(b)$.

ILP1 is a generalized model for the joint throughput optimization and traffic engineering problem. The flexibility of the model permits us to solve not only the joint problem but also either of them by appropriately setting the weight vector $\vec{w} = (w_1, w_2, w_3)$.

#### B. Complexity of ILP1

Variables $x^n_{ij}$, Eq. (4), and $f_n$, Eq. (5), are integer variables. The objective function and constraints are linear in the variables $x^n_{ij}$, $f_n$ and $\alpha$, $(i,j) \in E$ and $n \in N$. Therefore, the model is an integer linear program and is NP-hard. The number of constraints from Eqs. (1), (2), and (3) are $O(|V| \cdot |N|)$ (because at most $|N|$ traffic requests can be simultaneously active), $|E|$, and $O(|N|)$. The number of variables from Eqs. (4) and (5) are $O(|E| \cdot |N|)$ and $O(|N|)$. Thus, the number of constraints and variables are dominated

$$\text{Max } F = w_1 \cdot \sum_{n | t^n_n \leq t} f_n - w_2 \cdot \sum_{n | t^n_n \leq t} \sum_{(i,j) \in E} x^n_{ij} - w_3 \cdot \alpha$$

$$\sum_{(i,j) \in E} f_n ; \text{if } i = s_n$$

$$\sum_{(i,j) \in E} f_n; \text{if } i = d_n$$

$$\sum_{n \in N} x^n_{ij} \leq \alpha \cdot c_{ij}; \quad (i, j) \in E$$

$$f_n \leq r_n; \quad n| t^n_n \leq t \leq t^n_n$$

$$n| t^n_n \leq t < t^n_n, (i,j) \in E$$

$$f_n \in \{0, 1, 2, ..., r_n\}; \quad n| t^n_n \leq t < t^n_n, (i,j) \in E$$

$$0 \leq \alpha \leq 1$$

![Fig. 1. Integer Linear Program 1 (ILP1).](image-url)
by the terms $O(|V| \cdot |N|)$ and $O(|E| \cdot |N|)$. If these terms are not large (i.e., the number of nodes and links in the network, $|V|$ and $|E|$, and the number of simultaneously active traffic demands are not too large), then, ILP1 can be solved in a reasonable amount of time.

IV. SUBOPTIMAL APPROACH

Note that the optimal solution of ILP1 at any time $t$ may require the re-routing and re-establishment of paths whenever a new request arrives or a previous request ends. Finding the optimal solution of ILP1 may not be a viable solution when path disruptions are not permitted, or when the number of active traffic requests, $n|\tau|_{n}^{t} \leq t < t_{n}^{t}$, is large, since the computational time increases rapidly. As an alternative approach, we also evaluate a suboptimal approach that routes traffic requests one-by-one, as they arrive. Let $k \in N$ be a traffic request currently arriving. Then, the request $k$ is routed according to ILP2:

$$\text{Max } F = w_{1} \cdot f_{k} - w_{2} \cdot \sum_{(i,j) \in E} x_{ij}^{k} - w_{3} \cdot \alpha$$

$$\sum_{j:(i,j) \in E} x_{ij}^{n} - \sum_{j'(j,i) \in E} x_{ij'}^{n} = \begin{cases} f_{k} \text{; if } i = s_{k} \\ -f_{k} \text{; if } i = d_{k} \\ 0; \text{ otherwise} \end{cases}$$

$$x_{ij}^{k} \leq \alpha \cdot c_{ij}^{r} \quad (i,j) \in E$$

$$f_{k} \leq r_{k}$$

$$x_{ij}^{k} \in \{0,1,2,...\}; \quad (i,j) \in E$$

$$f_{k} \in \{0,1,2,...,r_{n}\}; \quad (i,j) \in E$$

$$0 \leq \alpha \leq 1$$

ILP2 has the same structure as ILP1; however, it only establishes a set of paths for the traffic request $k$ considering the current state of the network. The link capacity constraint given by Eq. (11) includes the term $c_{ij}^{r}$, which is the residual capacity of link $(i,j) \in E$ at the time the traffic request $k$ arrives. The optimal solution of ILP2 may represent a suboptimal solution for the multi-commodity problem (ILP1); however, it has a lower complexity than ILP1.

A. Complexity of ILP2

Similar to ILP1, variables $x_{ij}^{n}$, Eq. (13), and $f_{n}$, Eq. (14), are integer variables. The objective function and constraints are linear in the variables $x_{ij}^{n}$, $f_{n}$ and $\alpha$, $(i,j) \in E$ and $n \in N$, which makes ILP2 also an integer linear program and NP-hard problem. However, since only one traffic request is routed at a time, the number of constraints and variables are $O(|V|)$ and $O(|E|)$ instead of $O(|V| \cdot |N|)$ and $O(|E| \cdot |N|)$. Thus, the running time of solving an instance of ILP2 is inferior than that of solving an instance of ILP1.

V. PERFORMANCE STUDIES

Consider the network network topology shown in Fig. 3. We generated 200 hundreds traffic demands. The duration of each traffic demand is exponentially distributed (10 seconds), and the inter-arrival time is randomly distributed between 0 and 50 seconds. We solved ILP1 and ILP2 with four different weight vectors: $\vec{w}(a) = (1000, 0, 0)$, $\vec{w}(b) = (950, 50, 0)$, $\vec{w}(c) = (950, 0, 50)$, and $\vec{w}(d) = (900, 0, 0)$. The first objective of the problem is the throughput maximization. Thus, by setting $w_{1} >> w_{2}$ and $w_{1} >> w_{3}$, we give more importance to $F_{1}$ than $F_{2}$ and $F_{3}$.

A. Impact of $\vec{w}$ in ILP1

Consider Fig. 4(a). The curves for ILP1 were obtained with $\vec{w}(a)$ ($F_{1}(b)$, $F_{1}(c)$, and $F_{1}(d)$ are equal to $F_{1}(a)$, because $w_{1} >> w_{2}$ and $w_{1} >> w_{3}$, for all weight vectors). For ILP2, the curves were obtained with $\vec{w}(d)$, which produced highest throughput (to be explained in Section V-B). The parameter $r$ is the per-flow connection requests; i.e., for all flow $n \in N$, $r_{n} = r$. Note that the gap between ILP1 and ILP2 increases as $r$ increases. For $r = 1$, the curves are overlapped. The impact of the $\vec{w}$ can be noted in Figs. 4(b)-(d), which shows the resource consumption in wavelengths ($F_{2}$), the MLU ($F_{3}$), and the average link utilization (ALU), when $r = 1$. The ALU is defined as

$$\frac{1}{|E|} \sum_{(i,j) \in E} \sum_{n \in N} x_{ij}^{n} \cdot c_{ij}$$

and measures the link utilization in the network, on average. Fig. 4(b) shows the resource consumption. If $F_{2}$ is ignored as an objective function ($\vec{w}(a) = (1000, 0, 0)$ and $\vec{w}(c) = (950, 0, 50)$), the resource consumption increases with respect to those results obtained by slightly weighting $F_{2}$ with a positive weight of 50. Similarly, Figs. 4(c) and 4(d) show that the MLU and ALU are considerably reduced when a positive weight $w_{3} > 0$ is used. Note that even $\vec{w}(c) = (950, 50, 0)$ reduces the MLU and ALU; i.e., the minimization of the resource consumption indirectly reduces MLU and ALU.

B. Impact of $\vec{w}$ in ILP2

Consider now Figs. 5(a)-(d), which show the results obtained with ILP2. Here, we set $r = 5$ to better illustrate the effect of difference weight vectors. In contrasts with ILP1 where the four different weight vectors produced the same value of $F_{1}$, the throughput obtained with ILP2 varies with the weight vector. According to Fig. 5(a), the higher throughput is obtained with $\vec{w}(b) = (950, 50, 0)$ and $\vec{w}(c) = (950, 50, 0)$. These weight vectors produced even higher throughput than that obtained with $\vec{w}(a) = (1000, 0, 0)$, which optimizes $F_{1}$ only. This example demonstrates that to optimize throughput, other objectives such as resource consumption must be
simultaneously considered. Fig. 5(b) shows the resource consumption. Note again the good performance of ILP2, obtained with \(\vec{w}^{(b)}\) and \(\vec{w}^{(c)}\): with these weight vectors, not only the throughput is maximized but also the resource consumption is minimized. Fig. 5(c) illustrates the MLU. Since \(r = 5\), the scenario represents a high-loaded case, and the MLU is 1 most of the time, independently of \(\vec{w}\). As shown in Fig. 5(d), the ALU is also minimized with \(\vec{w}^{(b)}\) and \(\vec{w}^{(c)}\). Thus, by considering the 3 objectives simultaneously, better solutions are obtained. Note that these solutions may not be even tradeoff solutions, but Pareto solutions.

C. Throughput Surcharge

Since the weight vectors used in this paper attempts to primarily maximize throughput, we now compare the throughput surcharge between the best solutions, in term of throughput, of ILP2 (obtained with \(\vec{w}^{(d)}\)) with those of ILP1 (for ILP1, any weight vector produced the same throughput). We define the throughput surcharge \(\delta\) as the gap between the optimal throughput obtained with ILP1, \(F_1(\text{ILP1})\), and the throughput obtained with ILP2, \(F_1(\text{ILP2})\):

\[
\delta = \frac{F_1(\text{ILP1}) - F_1(\text{ILP2})}{F_1(\text{ILP1})}. \tag{16}
\]

Fig. 6 shows the results for different values of \(r\). When \(r = 1\) and \(r = 2\), ILP2 produces the same throughput as ILP1, and the throughput surcharge is 0. As \(r\) increases, the difference of throughput between ILP2 and the optimal increases up to 60% for \(r = 5\). Note, however, that until \(t = 50\) s, \(\delta\) is mostly below 40%. For \(t \geq 50\), there are no more connection request arrivals; remember that ILP1 considers, at any time \(t\), all the active flows \(n \in N|t^s_n \leq t < t^n_e\), while ILP2 does not reroute traffic requests previously arrived (no flow disruption). This is one of the reason of the surcharge gap. Another reason of the bigger gap when \(t > 50\), for \(r \geq 4\), is the fact that ILP1 reroutes traffic requests that were not able to be routed when the network was congested. E.g., let assume that a traffic request \(n\) arrived at \(t^s_n = 40\) s, with an ending time \(t^n_e = 70\) s. At \(t = 40\) s, assume that the network is congested, and \(n\) cannot be routed. Assume also that at \(t = 60\) s, the congestion is reduced and resources can be allocated to request \(n\). ILP1 then routes traffic request \(n\), while ILP2 does not.

VI. Conclusion

We have presented a multi-objective ILP for the joint throughput optimization and traffic engineering problem in dynamic WDM networks (ILP1), where traffic requests arrive one-by-one. The scheme considers the possibility of having new routing requests dynamically, and can be considered as an on-line scheme capable of handling requests in an optimal manner when the requests are not all presented at once. The proposed program simultaneously maximizes the aggregated throughput, minimizes the resource consumption and MLU. The ILP can generate optimal solutions for practical sized
networks in a reasonable time. We have also presented a suboptimal approach based on integer linear programming (ILP2), which does not require the re-establishment of paths for previous traffic requests already routed. The complexity of ILP2 is much lower than the ILP1, and is a good alternative scheme according to simulation results. Performance studies show that, by considering the multiple objectives simultaneously, better solutions can be obtained. For ILP2, we have noticed that, to maximize connection requests, the objective to be optimized should not only be throughput but also resource consumption and MLU. Future work includes the development of a systematic procedure to determine appropriate values of the weight vectors.

**Acknowledgement.** This work was funding in part by the US National Science Foundation under Award CNS-0806637. The authors are very grateful for this support.

**REFERENCES**


Fig. 6. Throughput surcharge obtained according to Eq. (16).