A Framework for Unified Traffic Engineering in IP over WDM Networks

Jinhan Song, Saewoong Bahk
School of Electrical Engineering, Seoul National University,
Seoul 151-742, Korea
Tel: +82-2-880-8434  Fax: +82-2-880-8198
{sjh, sbahk}@netlab.snu.ac.kr

Hyogon Kim
Division of Information and Communication, Ajou University
Joong-gu, Seoul 100-095, Korea
Tel: +82-2-319-4080  Fax: +82-2-754-1182
hyogon@madang.ajou.ac.kr

Abstract
Traditionally, the traffic engineering functionality has belonged to the realm of the “lower” layers, such as WDM or ATM or FR. The primary reason is the lack of traffic engineering in IP layer, i.e., that IP routes traffic based on hop distance regardless of the actual load on the routed path. However, a new development in IP networking is an attempt to enhance IP with a traffic engineering functionality, through, most notably, MPLS. So in this article, we consider a traffic engineering framework that has both WDM layer and IP layer traffic engineering components, and capture the performance behavior of such system.

Keywords: WDM, IP over WDM, Traffic Engineering, Optimization

The need for a unified traffic engineering in IP over WDM networks

Given the ever increasing demand for network bandwidth, and the recent phenomenal advances in WDM technologies, the Internet is expected to quickly embrace the IP-based optical networking architecture. As IP over WDM networking technologies mature, a number of important control issues are surfacing. Marked among them is traffic engineering, which is the optimization of network throughput while maintaining the required QoS [1], a task traditionally the layers below IP have undertaken. The primary reason for “lower layer” traffic engineering is that IP is designed to utilize only the hop distance to the destination to determine the route. Due to this so-called “min-hop” routing, IP is practically incapable of controlling traffic distribution over a network.
But a recent development in IP networking is to enhance IP with IP-controlled traffic engineering functionality[2], through, most notably, Multi-Protocol Label Switching (MPLS)[3]. MPLS allows for the explicit setup based on actual link usage as well as hop count, and this feature can easily be leveraged for traffic engineering purpose. When the default path between two MPLS-capable routers becomes overloaded, for instance, the router pair can alleviate congestion by shedding some traffic off the default route and routing it over an alternate path for load sharing [12].

Note that the fact that IP is being equipped with its own traffic engineering functionality does not immediately obviate the need of any similar functionality on WDM layer. There is a limit in IP level traffic engineering using MPLS, since the maximum capacity between a given pair of routers is dictated by the underlying WDM topology. Graph-theoretically, the maximum capacity is equal to the “min-cut”, which is the minimum wavelength capacity crossing the network between the router pair. Unless the underlying WDM layer opens more capacity through the reconfiguration of the WDM virtual topology[4], the MPLS-based IP level traffic engineering can only allow as much traffic as the given WDM topology can accommodate. Furthermore, if MPLS is not aware of the WDM topology, it can attempt to spread out the load without knowing that it already reached the min-cut capacity. Then MPLS only spreads the congestion to neighboring links [13]. Note that this problem is not specific to MPLS. Any IP level traffic engineering scheme that utilizes explicitly utilized alternate paths to balance the load between a router pair can face the same problem.

On the other hand, the existence of WDM layer traffic engineering does not deny the need of IP level traffic engineering, either. First, the granularity of a wavelength is fixed and huge. So it could be inappropriate for WDM-level traffic engineering is to meet smaller bandwidth demand changes that could be easily met by IP-level traffic engineering such as alternate routing. Second, frequent changes in the IP connectivity as a consequence of WDM virtual topology reconfiguration may cause unwanted disruption to IP routing. So, the traffic engineering that considers these two layers has a separate operating regime for each layer in terms of time scale and bandwidth granularity. If WDM layer and IP layer both perform the traffic engineering functions, however, the lack of coordination between the two layers may lead to inefficiency and even pathologies. Thus a unified framework that couples the two traffic engineering functions is called for.

**Modeling a simple, unified traffic engineering system**

As the debate still unfolds over the IP over WDM networking model, and as views vary from forum to forum [5,6,7], drawing a complete unified traffic engineering system model at the current stage of development would be difficult and could not be so helpful. So instead, we will let an abstract analytical model serve our purpose to delineate the behavior of the unified traffic engineering system and explore the performance characteristics.
A WDM layer lightpath (that could span multiple optical cross-connects) is abstracted as a high-speed point-to-point link by the two IP nodes on either end of the lightpath. If multiple lightpaths are available and active for packet transmission between a pair of IP nodes, then the capacity is much larger but the IP connectivity does not change. So it is still modeled as a single IP link. Note that a lightpath may use a single wavelength, or multiple wavelengths if wavelength conversion is supported. But it does not affect our main results, so we will assume the former and use the terms, lightpath and wavelength interchangeably. The IP nodes use the single- or multiple-wavelength link to transport IP packets between them. It is natural to assume that the arriving packets are placed in the queue before being transmitted. Thus a WDM lightpath can be thought of as a server that serves IP packets waiting in the queue. Fig. 1 depicts such a system model.

The scheduler shown in the Fig. 1 classifies incoming IP packets and distributes them to the appropriate queues according to the traffic engineering policy of both IP and WDM layers. A few exemplary policies will be discussed later.

In the MPLS context, each queue serves forwarding equivalent classes (FECs). As packets in an FEC travel along the same label switched path (LSP), they can be mapped to one or more lightpaths. We model the IP traffic engineering as varying the input of the traffic arrival rate \( \lambda \) associated with each LSP as a means of evenly distributing the traffic. If an LSP carries traffic to the extent that (one of) its lightpath begins to be overloaded, IP traffic engineering function intervenes to resolve this situation. It creates alternate LSPs (if possible), partitions the FEC for the overloaded LSP into smaller constituents, and routes packets in some of the constituent FECs to these new alternate LSPs. Note that an FEC is not at all a load-oriented unit. When an FEC is created, we do not know how much traffic it will carry, except that it will vary as a function of time. Also note that the new LSPs should use some lightpaths connecting the IP node pair other than the overloaded one. While IP traffic engineering modulates \( \lambda \), WDM engineering can be considered varying the departure rate \( \mu \) for an IP link. If additional bandwidth is required to meet the traffic demand between a pair of IP
nodes, WDM engineering simply provides more wavelengths that could be at the cost of other IP link capacity. As the number of allocated wavelengths increases, for instance, the service rate increases and the congestion should be relieved. This system model is generic enough to abstract the class of IP and WDM traffic engineering methods that modulate $\lambda$ and/or $\mu$.

A cost metric with which we evaluate the performance of a unified traffic engineering framework can be formulated from the following intuitive observations. First, the cost in the IP layer increases as the queue builds up. This is because delay and loss probability of packets increase as the queue grows. Second, obviously, the cost in the WDM traffic engineering can be characterized as the number of allocated wavelengths since wavelengths are a critical resource. A generic cost function thus can be expressed as:

Combined Cost $C$

\[ C = (\text{cost of the IP layer}) + (\text{relative cost weighting factor}) \times (\text{cost of the WDM layer}) \]

\[ = N_{IP} + c \cdot N_{WDM} \]

where $N_{IP}$ is the sum of queue lengths in the system, $N_{WDM}$ is the total number of allocated wavelengths, and $c$ is the ratio of the unit cost of the IP layer to the unit cost of the WDM layer.

We observe that the combined cost $C$ reflects the engineering tradeoff between the IP and WDM layers. On one hand, for maximal utilization of the network, a queue should serve as much traffic as its allocated wavelength(s) can accommodate. Namely, minimizing $N_{WDM}$ inevitably implies the increase in $N_{IP}$. On the other hand, as the capacity of a queue is made smaller to reduce $N_{IP}$, more wavelengths are necessary to maintain the same level of loss and delay performance for IP packets. At an extreme, the cost of the IP layer is minimized if the number of allocated wavelengths is infinite, which means the infinite cost in the WDM layer. Likewise, if we are to minimize the cost of the WDM layer, only a minimal number of wavelengths need to be allocated. In this case, the queue builds up to infinity when the arrival rate reaches the service rate, resulting in infinite cost in the IP layer. So it is crucial for a unified traffic engineering scheme to strike the balance between these two extremes and minimize the combined cost $C$. And the value of parameter $c$ is pivotal to determine the optimal operating point of the traffic engineering in a unified framework.

A simple analysis

Now we investigate the behavior and the characteristics of the unified traffic engineering through a simple analysis of the system model. First, we analyze the system with an FEC and answer the question of how much optical resource should be allocated to an FEC to minimize the combined cost. Then, we extend the result to the system with multiple FECs.
For a simple analysis of our generic system model, we draw on elementary queuing theory. It turns out that depending on the load distribution algorithm in the IP level traffic engineering, the queuing model can slightly differ. If the algorithm chooses the least loaded lightpath (or its associated queue) for arriving packets, which is called `metering scheduling [8]', it can be modeled with a single queue served by multiple servers. This is because the multiple queues with their own servers behave the same as a single queue with multiple servers. On the other hand, if the algorithm chooses the queue regardless of its loading status, the system can be modeled with `randomization scheduling [8]', where each queue is served by its own server.

To analyze our generic system model, the usual assumptions come in. The packet arrival process is Poisson, and the service time is exponentially distributed. Note that actual packet arrivals could be bursty without following a Possion process, but we leave the study with other traffic models for our future work. Also, each queue reaches the steady state so that it has an equilibrium probability distribution. Let us first consider the randomization scheduling case. Given the number of queues (i.e., lightpaths) $k$, all queues have the same utilization $\rho = \frac{r}{k} \leq 1$ where $r = \frac{\lambda}{\mu}$ is the normalized offered load to the system, and $N_{IP}$ equals $k\times$ (average number of customers in an M/M/1 queue). Using the result of the M/M/1 queueing system, we can show that $C = kN_{M/M/1} + ck \geq r(\sqrt{c} + 1)^2$ where $N_{M/M/1}$ is the average number of customers in an M/M/1 queue [9]. We can also show that the equality holds when $k = (1 + \frac{1}{\sqrt{c}})r = k^*$, and the optimal utilization of each queue becomes $\rho^* = \frac{r}{k^*} = \frac{\sqrt{c}}{1 + \sqrt{c}}$. If the cost of the WDM layer engineering is dominant, e.g. $c \rightarrow \infty$, $\rho$ converges to 1 and $k$ converges to $r$. This means that the traffic load is mostly borne by the IP layer and the queue length goes to infinity since $\rho \rightarrow 1$. But if the cost of the IP layer is dominant, then $c \rightarrow 0$ and $\rho$ converges to 0 and $k$ diverges to $\infty$. Again, it means that the traffic load is mostly borne by the WDM layer and each packet is served by a separate wavelength. This behavior is intuitively correct, and confirms that $c$ determines which layer plays the major role in the unified traffic engineering.

If metering scheduling is used instead of randomization scheduling, an arriving packet is assigned to the least utilized queue to achieve high utilization and low delay. This system can be analyzed by M/M/k queuing system [8]. In this case the combined cost becomes $C = N_{M/M/1k} + ck$ where $N_{M/M/1k}$ is the average number of customers in an M/M/k queue. For brevity, we present some numerical results since it is not possible to obtain the optimal solution analytically.
In Fig. 2, the cost $C$’s of metering and randomization schedulings are plotted as a function of the number of allocated wavelengths $k$. Both the metering and randomization schedulings show the tradeoff relation between the costs of IP and WDM layers. If we push the number of active queues too low or too high, the combined cost increases. An optimal tradeoff point exists for each scheduling method, at which a unified traffic engineering mechanism tries to maintain its operating point. Fig. 2 also shows, rather obviously, the metering scheduling requires less number of wavelengths and achieves lower cost. This is because it assigns an incoming packet to the least utilized queue.

In reality, we cannot assert that the traffic arrival pattern for the IP layer follows Poisson distribution and the service time is exponentially distributed. However, irrespective of traffic distribution, the same tradeoff relation exists between IP and WDM layers, which is implied in the definition of the cost function. This is because the cost increases as more traffic arrives or as more queues are allocated.

In a system with multiple FECs, FECs contends for wavelengths to transport its packets in a short time. The number of allocated wavelengths that an FEC can have is limited by the cost of having wavelengths. For a simpler analysis of this case, we opt for randomization scheduling, and further, assume that the WDM network is ideal. It means that routing and wavelength assignment (RWA) [10] are done without delay, and no blocking occurs if wavelengths are sufficiently many. To delineate only the core effect of the algorithm, we abstract away the other specifics of RWA. Finally, let us also
assume that we can measure the packet arrival rate for each queue. Regarding the results for packet-level randomization scheduling above as the system behavior with regard to a single FEC, we can formulate a similar optimization problem for the system with multiple FECs. Namely, when the combined cost of transporting $i$-th FEC is written as

$$C_i = c k_i + r_i + \frac{r_i^2}{k_i - r_i}$$

where $r_i$ and $k_i$ represent the estimated arrival rate and the number of allocated wavelengths for $i$-th FEC respectively, the optimization problem is to minimize the total cost $C = \sum_{i=1}^{N} C_i$, subject to

$$\sum_{i=1}^{N} k_i \leq L \quad \text{and} \quad k_i \geq r_i.$$ 

$L$ is the total number of wavelengths that can be allocated, and $N$ represents the number of active FECs. With a little mathematical maneuver and approximations, we can obtain that $k_i$ minimizing $C$ is

$$k_i = \min[(1 + \frac{1}{\sqrt{c}}) r_i, \frac{L}{N} r_i]. \quad (2)$$

If $L$ is very small, then the latter term in the right side of the equation will minimize the equation. In this case, wavelengths are allocated to each FEC fairly according to its load. With larger $L$, each FEC will tend to be allocated with more wavelengths. When the wavelengths are enough that the former term comes to minimize the equation, the allocation state has reached the optimal point and will not change with much larger $L$.

Note that the optimal traffic engineering scheme should try not only to allocate the resources fairly but also to reduce the number of assigned wavelengths. The unified traffic engineering achieves fairness by allocating wavelengths in proportion to the load of each FEC. On the other hand, the number of wavelengths that a node can have is limited by the cost in the WDM layer. This is a major departure from weighted fair queueing (WFQ)-like scheduling methods used in routers and switches [11]. These schemes utilize the excess bandwidth to the fullest to achieve high throughput. In our framework, however, it is possible that some wavelengths remain unassigned even with the excess wavelengths and longer alternate paths are created. In this manner the network resource can be managed more effectively. Also note that the optimal traffic engineering scheme tries to balance the roles of two layers by controlling the parameter $c$. If $c$ is set to 0, it distributes wavelengths freely and works like a WFQ scheduling method. But if $c$ is set to $\infty$, it tries to use a minimal number of wavelengths possible.

**A practical algorithm**

We can easily devise a practical algorithm to calculate the optimal number of wavelengths to be allocated to an IP node, using the results obtained previously. The input load is estimated using the moving average method. If the input load change of an FEC is bigger than a predefined threshold, the algorithm is activated. Since the required number of
wavelengths is proportional to the input load, we calculate \( \Delta k = k_{\text{estimated}} - k_{\text{current}} \) instead of tracking the changes of the input loads, and if it is larger than \( k_{th} \), the scheme decides that the reconfiguration is necessary. The current algorithm continuously checks whether the reconfiguration is necessary, while the overhead will be smaller with periodic or event-triggered activation. The practical algorithm is shown in Table 1. The \( k_{\text{estimated}} \) for an individual FEC is calculated in Step 2, and the total \( k_{\text{estimated}} \)’s calculation is done in Step 3.

**Table 1. Algorithm**

| (step 1) | Estimate the load. |
| (step 2) | Check if the reconfiguration is required using the estimated load. |
| (step 3) | If required, calculate the required number of wavelengths for each IP node. |
| (step 4) | Perform the wavelength routing and assignment. |

A simulation below shows the behavior of the unified traffic engineering system iterating the algorithm. In the simulated scenario, three FECs contend for wavelengths. And the input loads of FECs are varied in 3 intervals. Parameters used and calculated optimal \( k_i \)’s are summarized in Table 2.

**Table 2. Simulation Parameters**

<table>
<thead>
<tr>
<th>Common Parameters</th>
<th>( c = 100 )</th>
<th>( L = 120 )</th>
<th>( k_{th} = 1.25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 &lt; t &lt; 300</td>
<td>( r_1 = 10 )</td>
<td>( r_2 = 30 )</td>
<td>( r_3 = 100 )</td>
</tr>
<tr>
<td></td>
<td>( k_1 = 8 )</td>
<td>( k_2 = 26 )</td>
<td>( k_3 = 86 )</td>
</tr>
<tr>
<td>Period 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300 &lt; t &lt; 600</td>
<td>( r_1 = 30 )</td>
<td>( r_2 = 30 )</td>
<td>( r_3 = 30 )</td>
</tr>
<tr>
<td></td>
<td>( k_1 = 33 )</td>
<td>( k_2 = 33 )</td>
<td>( k_3 = 33 )</td>
</tr>
<tr>
<td>Period 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>600 &lt; t &lt; 900</td>
<td>( r_1 = 50 )</td>
<td>( r_2 = 30 )</td>
<td>( r_3 = 5 )</td>
</tr>
<tr>
<td></td>
<td>( k_1 = 55 )</td>
<td>( k_2 = 33 )</td>
<td>( k_3 = 6 )</td>
</tr>
</tbody>
</table>
Fig. 3 plots the simulation results. The first period simulates an overloaded situation where optimal $k_i$’s cannot be allocated. In this case, wavelengths are allocated in proportion to the input loads. In the second period, all FECs have the same input rates. All FECs share the wavelengths fairly. In the third period, the system becomes underloaded and FEC 1 fluctuates with a higher variation. This is because in the underloaded region, the contention for wavelengths is relieved and idle wavelengths can easily be allocated. We can see that the algorithm performs well when wavelengths are fairly and economically allocated in face of varying system load.

Table 3. Cost comparison

<table>
<thead>
<tr>
<th></th>
<th>$C_{optimal}$</th>
<th>$C_{simulation}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>period 1</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>period 2</td>
<td>10890</td>
<td>14769</td>
</tr>
<tr>
<td>period 3</td>
<td>10285</td>
<td>12186</td>
</tr>
</tbody>
</table>

The optimal cost $C_{optimal}$ and the cost obtained through simulation $C_{simulation}$ are compared in Table 5. It shows that
$C_{\text{simulation}}$ is always larger than $C_{\text{optimal}}$. This is because allocation states change frequently due to the input load estimation error in the simulation, so the queues do not reach their steady states. This violates the steady state assumption in the optimal cost analysis, and incurs some additional costs. Also there exists a tradeoff relation between the reconfiguration interval and the adaptation speed. If we lengthen the reconfiguration interval, and thereby reconfigure the network less frequently, the system can reach the steady state and the cost approaches the optimal value. But obviously, this method can only work well under moderate input load fluctuations. On the other hand, if we are to have the fast adaptation due to wild load fluctuations, we must shorten the reconfiguration interval at the cost of the increased cost. Since the current algorithm continuously checks the input load and updates the configuration of networks, it has a minimal reconfiguration interval, which results in fast adaptation but increased cost.

**Summary**

This article explored how the traffic engineering in IP over WDM networks should be designed or re-engineered. Harnessing the phenomenal advances in WDM technologies, IP over WDM networks will soon become a reality. A new development in IP networking that calls for the re-engineering in the emerging IP over WDM architecture is the addition of traffic-engineering functionality on IP layer through MPLS. We observed that the IP level traffic engineering (achieved through MPLS) and traditional WDM virtual topology reconfiguration have disparate operating regime in terms of time scale and bandwidth granularity, so that they should be mutually complementary rather than obviating the need for each other. As for the mode of complementary operation of the unified two-layered traffic engineering, we found that they are in tradeoff relation. A generic analytic formulation led to a simple solution for a unique optimal operating point, where the total cost of operation is minimized. The algorithm implementation based on the analytical solution performed well in face of changing load condition, which was demonstrated in simulations.

**Jinhan Song** received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Korea, in 1996 and 1998, respectively. He is currently working toward the Ph.D. degree in Seoul National University. His research interest includes architectures and protocols for WDM networks, and traffic engineering for IP over WDM networks.

**Saewoong Bahk** received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Korea, in 1984 and 1986, respectively, and the Ph.D. degree from the University of Pennsylvania, Philadelphia, in 1991. From 1991 to 1994, he was with the Department of Network Operations Systems at AT&T Bell Laboratories as a Member of Technical Staff, where he worked on the projects for the development of network operation systems for AT&T networks. He then joined the School of Electrical Engineering at Seoul National University and now serves as an associate professor. His areas of interests include performance analysis of communication networks, high-speed network protocol
design, routing, and resource allocation in wireless/wired/WDM networks.

**Hyogon Kim** is an assistant professor at Ajou University, Suwon, Korea. Formerly, he was with Bellcore (now Telcordia Technologies), where he was a research scientist in the area of Internet QoS, optical networking architecture, Internet protocols, and mobile networking. He holds Ph.D. in computer and information science from the University of Pennsylvania.

**References**


