

Increasing the Link Utilization in IP over WDM Networks Using Availability as QoS

Antonio Nucci^a, Nina Taft^a, Patrick Thiran^b, Hui Zang^a and Christophe Diot^a

^a Sprint Advanced Technology Labs, Burlingame, CA 94010, USA

^b LCA-ISC-I & C, EPFL, CH-1015 Lausanne, Switzerland

ABSTRACT

In this paper we solve a mapping problem related to supporting two service classes that are differentiated based on their level of protection. The first class of service, called Fully Protected (FP), offers end-users a guarantee of survivability in the case of a single failure; all FP traffic is protected using either a 1:1 or 1+1 protection scheme at the WDM layer. The second class of service, called Best-Effort Protected (BEP), is not protected; instead restoration at the IP layer is provided. When a failure occurs, the network restores as much BEP traffic as possible, and thus BEP traffic does not receive any specific guarantees. The FP service class mimics what Internet users receive today. The BEP traffic is designed to run over the large amounts of unused bandwidth that exist in today's Internet. The motivation of this approach is to give carriers a mechanism for increasing the load carried on backbone networks without reducing the QoS received by existing customers.

In order to support two such services, the logical links at the IP layer need to be carefully mapped onto primary and backup paths at the optical layer. We incorporate into our mapping problem a number of practical requirements that reflect constraints that carriers face and policies they want to enforce. For example, we allow the FP demand to be specified via a traffic matrix at the IP layer, we include an overprovisioning factor that specifies the portion of each link that must be left unused, and we incorporate a minimal fairness requirement on how the BEP traffic is allocated among connections. Our goal is to quantify how much BEP traffic can be carried in addition to the FP traffic, without impacting the protection quality of the FP traffic even in the case of failure, and without impacting the FP load.

We provide two solutions, one is an optimal solution using an Integer Linear Program (ILP) model, and the other is an algorithm based on the Tabu Search methodology. Our heuristic algorithm allows us to solve this problem for large networks such as those spanning the continental US. We show that by having two such classes of service, the load on a network can be increased by a factor of 4 to 7 (depending upon the network). We illustrate that even with overprovisioning and fairness requirements (both of which reduce the total possible BEP load carried), we can still typically triple the total network load. We show that the location of the bottleneck can affect whether or not we see a difference in performance between 1:1 or 1+1 protection schemes. Our results illustrate the gain in terms of additional BEP load carried that can be obtained simply due to the upgrade of a single link. Our proposal provides carriers a new vehicle for generating revenue by extracting benefit from all the sources of unused bandwidth in networks.

1. INTRODUCTION

The Internet backbone contains a large amount of capacity that is not currently being used. Some of this extra capacity is reserved for protection and gets used only when failures occur. Some of this excess may not be used at all for many months because it is part of the long term growth planning of a backbone. Carriers today are very interested in carrying additional load on their networks in order to generate additional revenue; however they are concerned about not reducing the quality of service received by existing customers. To achieve these two goals simultaneously, we propose that carriers provide two classes of service, one of which would mimic today's service and a second one that would provide a lower quality of service. The idea is for the lower-grade service to be carried on the "excess" bandwidth in the backbone in such a way that has no impact on the Service

Level Agreements (SLAs) promised to the higher-grade service. The majority of the time this excess bandwidth is unused, hence the lower-grade service will have a good SLA. When this bandwidth becomes needed in a failure scenario, we would drop as many packets as needed from the lower-grade service in order to ensure there is enough bandwidth to protect the higher-grade service.

In order to achieve this, the two classes of service should be differentiated by their *level of protection* against failures. The first class, called *Fully Protected* (FP), offers users the insurance that they will not suffer service interruption in the case of a single failure. Protection is provided at the WDM layer via either a 1+1 or 1:1 protection scheme that guarantees fast recovery after a single failure. This service class represents what an Packet-over-SONET backbone provides today. The second class of service, called *Best Effort Protected* (BEP) is new. It does not provide a specific guarantee on service disruption. Instead, in the case of failure, it offers to restore as much of the affected traffic as possible. This restoration is taken care of at the IP layer, via IGP protocols such as OSPF or IS-IS which have their own mechanisms for detecting failures and computing new paths. Protection at the WDM layer is very fast (e.g., 50 ms in SONET) because precomputed backup paths are used, while restoration at the IP layer is slower since new alternate paths are computed after the failure has been detected. This idea of the two service classes, FP and BEP, was initially proposed in Thiran.¹⁵ That paper considered a small six node network and provided a proof of concept that the load in the network can be significantly increased without impacting existing traffic.

The three main reasons why the Internet contains unused capacity are because of *equipment redundancy*, *overprovisioning*, and *the link upgrade process*. Equipment redundancy typically leads to a multiplicity of links between a pair of nodes. Overprovisioning usually implies that network links are run at low utilization levels. Redundancy of equipment and overprovisioning are used to protect the backbone against failures. Single link failures due to fiber cuts are more common than one might expect; today's large international Tier-1 backbones experience 2 or 3 fiber cuts per month.¹⁴ With technologies such as WDM, a single fiber failure can bring down a large number of IP paths. Overprovisioning is the current solution adopted by carriers towards providing quality-of-service in the Internet. With overprovisioned networks, the delays and losses in the Internet are very small, as proved in Papagiannaki et al.¹² The simplicity of overprovisioning makes it cheaper in IP backbones than the alternative solutions of reservation-based and priority services, which can also offer users a low-delay, low-loss service.

Upgrading the links (i.e., converting an OC-12 to an OC-48 link) in a large backbone is a time consuming process. For example, upgrading a large inter-POP backbone link can take a few months. Thus each time a link is upgraded, a "pocket" of additional bandwidth become available. But this is not really available to users whose path traverses both the new link and older slower links since the slower links will determine the end-to-end throughput. When a network is partially upgraded, and has many pockets of bandwidth scattered over the topology, a potentially large number of users could indeed profit from this new capacity. Our proposed BEP traffic would also run over this type of additional bandwidth. We will see how the upgrade of a single link can impact the amount of BEP traffic a network can carry.

In order to support two such services in an IP over WDM network, the logical links between routers at IP

level must be carefully mapped over the physical topology. This mapping needs to be done carefully so that we can support as much BEP traffic as possible without impacting either the SLA or the protection received by the FP traffic. The general mapping problem has been studied before; we address a version of this problem that incorporates a number of elements that arise in practice. These reflect constraints that carriers face and policies they want to enforce. We thus include the following aspects in our problem (1) We allow FP demands to be heterogeneous and specified via a traffic matrix at the IP layer. (2) We consider real WDM systems where the bandwidth of a fiber is equally divided among a fixed number of channels and each channel is assigned a different wavelength. In our heterogeneous network scenarios, we have different WDM systems that differ in the number of wavelengths supported and in their total bandwidth. We also include the line card speed limits in routers at the IP layer. (3) We incorporate an *overprovisioning factor*, denoted β_{FREE} , that represents an operational requirement to leave a given fraction of each logical link unused. (4) We include a *fairness policy* that enables carriers to ensure that there is some equity in how the offered load for this new service is distributed among logical connections. We develop solutions to this problem for both the 1+1 and 1:1 protection schemes at the WDM layer for FP traffic. In this paper we do not study the restoration problem at the IP layer, instead we leave the BEP traffic unprotected at WDM layer.

We briefly comment on two of these elements, the over-provisioning factor, and the fairness policy. (These are motivated and discussed at length in Section 3.) As part of the overprovisioning practice, it is common for carriers to require that a fixed percentage of each logical link remains unused. We incorporate this practice into our model to study the impact of such a policy. The more we over-provision, the less BEP traffic we can carry. We will quantify this relationship in our test cases.

Although we want to select the mapping that allows the maximum of BEP traffic to be carried, we must be careful as to how we distribute the unused capacity among the logical connections. If we simply try to add as much BEP as possible, network-wide, by trying to maximize a global network load, it wouldn't be surprising if the distribution of the amounts of BEP offered to each logical connection were very inequitable. Experience indicates that short connections would be favored. We were thus motivated to include a fairness policy in our solutions to circumvent highly inequitable offerings of BEP. In our approach to this problem, we thus maximize the BEP traffic carried *subject to* a fairness constraint or policy. The more fairness we impose, the less BEP traffic we can carry.

We provide two solutions for the problem we address. The first uses optimization techniques to find an optimal solution based on formulating the problem as an Integer Linear Program (ILP). The model was already presented in ⁹ and is not included here for lack of space. As in most ILP models designed for real problems, our model is powerful in finding the optimal solution for small and medium sized networks, but unfortunately is not scalable for large networks due of the enormous number of variables that arise. Our second solution, presented for the first time in this paper, defines a sub-optimal algorithm based on the Tabu Search (TS) methodology ³ that can be used in practice for actual carrier backbone networks. The heuristic cannot guarantee that the solution found is the optimal one, but it is scalable and applicable for a real large network scenario.

In⁹ we evaluated the performance of our optimal ILP solution on medium-sized networks. In this paper,

after presenting our heuristic solution, we validate it against our optimal model using a large set of test cases for medium sized networks, and show that the worst case difference between the objective functions achieved by our heuristic and the optimal solution was less than 3%. Thus our heuristic solution is close to the optimal one. We then use our algorithm to evaluate the performance benefit in a typical large scale US carrier backbone network. Our ILP model used a simple minimum fairness policy, whereas our heuristic solution uses a true max-min fairness policy.

The goals of this paper are: i) to quantify how much BEP traffic can be carried on the network without impacting the FP service; ii) to determine how to allocate the BEP traffic load among all the logical connections (i.e., compute the BEP traffic matrix) such that the total BEP traffic carried by the network is maximized and the partition of the BEP traffic is as fair as possible; iii) study the impact of the overprovisioning factor β_{FREE} on the problem; iv) understand the *bottlenecks* that arise when considering implementation factors such as line cards in routers and WDM systems in OXCs; and v) evaluate the average and the worst BEP losses when a failure happens in the WDM layer. Quantifying the losses BEP may experience allows us to assess to what extent the service is degraded during failure episodes.

We study three network scenarios: two versions of a medium-sized network with heterogeneous links similar to the Italian backbone, and an homogeneous large-sized network similar to the Sprint backbone. The second medium-sized network is the same as the first with the exception of one link that we upgrade in the second version. This allows us to assess the impact of individual link upgrades. We solve this problem for our medium-sized networks using our ILP model, and for our large network using our Tabu Search heuristic. The results show that the gain of using our FP/BEP approach is that the total network load can be increased by a factor that varies between 4 and 7, depending upon the network. Even when the over-provisioning factor is at 50%, we still see a tripling of the total load that can be carried in a network if the FP/BEP approach is adopted. This shows that the potential of using a BEP traffic class to generate additional revenue is huge, and that this potential can be achieved without any impact to those desiring a high-grade protection service. We discuss the issue of whether bottlenecks lie in the WDM layer or the IP layer. We find that when the main bottlenecks lie in the WDM layer, we see marked differences in the performance of the 1:1 and 1+1 protection schemes. However, when all the bottlenecks are at the IP layer, there is no difference in performance between the two protection schemes. We demonstrate that the BEP traffic that can be carried decreases linearly as the over-provisioning factor grows.

Recently the problem of service management has gained a lot of attention in the optical community ^(4, 6, 7, 13). Proposals for different service classes in optical networks are introduced in Gerstel and Ramaswami.⁴ Ramamurthy and Mukherjee¹³ study the traditional 1+1 and 1:1 protection strategies at the WDM layer for a single class of traffic. They formulate the corresponding ILP optimization problem applicable to small networks. Mohan and Somani⁶ propose a class of service that offers a minimal level of protection to every connection. They claim that if the demands are highly dynamic, it is possible to select routes whose (shared) back-up paths have a specified maximal non-zero probability of being unavailable if a failure occurs. The difficulty here is to provide a tight upper bound on this probability, and to select lightpaths (i.e. logical connections) that have this

specified degree of protection in a fluctuating environment. Sridharan and Somani⁷ formulate the ILP problem when three different service classes co-exist. They try to minimize the capacity requested by all working and backup paths, weighted by the traffic class to which it belongs (since each class brings in a different amount of revenue). Ramamurthy and Mukherjee¹³ prove that the general problem is NP-complete for a single class of traffic. Hence the recent proposal for three classes of traffic at the WDM layer may be too complex to apply to real networks.

The remainder of this paper is organized as follows. The FP and BEP classes of service are fully defined in Section 2. In Section 3 we explain which components of the overall problem belong to which layer (physical or logical), describe our traffic matrices and give a formal problem statement. The approaches used to solve the problem are introduced in Section 4, while a detailed description of the proposed algorithm based on Tabu Search methodology is presented in 5. Numerical results for our three network scenarios are presented and discussed in Section 6. Section 7 concludes the paper.

2. DEFINITION AND PROVISIONING OF CLASSES OF SERVICE

In this section, we fully specify the two classes of service introduced earlier. The **Fully Protected (FP)** service guarantees its customers that their traffic is protected against any single point of failure in the backbone. FP traffic is protected via pre-computed, dedicated backup paths at the WDM layer, using either a 1+1 or 1:1 protection strategy. Failures are transparent to the IP layer for this class of traffic. In a 1+1 protection scheme, the FP traffic is transmitted simultaneously on two disjoint paths. The receiver selects the signal at the destination that has the better signal quality. If that path is cut, the receiver automatically switches to the other path to receive input. In a 1:1 protection scheme, the FP traffic is transmitted only on one path (called the *working* or *primary* path). If this path fails, the sender and receiver both switch to the other path (called the *backup* path). Our idea is to take advantage of 1:1 protection because the reserved but unused capacity on the backup path can be given to unprotected traffic whose packets would be dropped in the case of a failure.

The **Best Effort Protected (BEP)** service does not offer a guarantee of traffic survivability in case of failure. For BEP traffic we offer restoration and not protection; in other words, BEP traffic is entirely unprotected at the WDM layer, and instead we rely on the IP layer to carry out restoration. When a failure occurs, BEP packets may be dropped at the router before the point of congestion, until IP has been able to restore this traffic by rerouting it on an alternate IP path. Failures are detected at the IP layer via IGP routing protocols such as OSPF or IS-IS. Thus the restoration offered to BEP is a slow one, in contrast to that offered to FP. The BEP traffic of each logical connection can be routed on either the primary or the backup path, but not on both (i.e., it cannot be split over two paths). If there is no spare capacity available at all at the IP layer at the moment a fiber cut occurs, then the BEP traffic could experience a total disruption until the IGP protocol as the IP layer converges. Having spare capacity at the IP layer (“spare” here means above and beyond the capacity given to FP and BEP traffic) would allow the IP layer to reroute BEP traffic at the time of failure. This means that the BEP service would be degraded but not interrupted until IP layer recovery completes. Thus the SLA performance, in terms the packet drop rate, received by the BEP traffic depends upon the amount of overprovisioning that exists after both FP and BEP traffic have been accommodated.

We point out that in an environment in which each logical connection is protected via either a 1:1 or 1+1 scheme at the WDM layer, and in which failures happen one at a time, the logical topology will always be connected. Thus the logical topology will be always able to apply a restoration strategy at the IP layer, and does not suffer from the *failure propagation* problem described in Crochat et al.^{5,10,11}

In order to implement two such classes of service, packets need to be marked according to their class of service, and IP routers must implement class-based scheduling. Many commercial routers today already have priority-based scheduling available. In normal operation, differentiation is not needed between the two types of packets. However, upon notification of a failure, FP packets continue to be served as before, while BEP packets may be dropped until BEP traffic has been restored at the IP layer.

3. PROBLEM STATEMENT

The main problem we address is to find a mapping of IP-layer logical links to physical fibers such that (1) the FP traffic is protected, (2) the amount of BEP that can be added is maximized subject to a constraint imposing a minimally fair allocation of BEP load among all the logical connections, and (3) an overprovisioning requirement (described below) is satisfied. Our intent is to add BEP traffic into the system such that there is no impact at all on the protection quality received by the FP traffic, even in the case of a single failure. In order to protect the FP traffic, we find for each logical link, two link-disjoint fiber paths at the physical layer - a *working* physical path, and a *backup* physical path. Our solution is robust to single and even to multiple failures as long as none of them is a *critical failure*. In this context, a *critical failure* is a multiple failure scenario that brings down a set of links such that both the working and backup path of the same logical link are interrupted.

A number of the inputs to our problem, that define requirements and constraints, come from the IP layer. To clarify which components of the problem are related to the logical (IP) layer and which are part of the physical (WDM) layer, we now discuss the elements of each layer and comment on the relationship between these elements. To be clear, we state some definitions of basic terms. We use the expression *logical link* to refer to a single link between two routers at the IP layer. We use the term *logical connection* to refer to a sequence of logical links. Each logical link corresponds to a sequence of one or more physical links interconnected via OXCs.

We focus on PoP-to-PoP (Point-of-Presence) topologies at the IP layer, rather than on router-to-router topologies that consist of hundreds of routers. A PoP is an ensemble of core and access routers that usually reside in a single building in a metropolitan area. PoPs are interconnected via inter-POP links attached to the core routers. The access routers are used to connect customers to the backbone. With this topology, the logical links we map capture the inter-POP backbone links. Since optimization is much more important for inter-POP links than intra-POP links, we choose to focus on the POP-level topology. Conceptually a PoP is equivalent to a router because we have collapsed an ensemble of a few core routers into a single router. Access routers can be ignored because they do not connect directly to other PoPs or other routers in the backbone.

The **FP traffic matrix** is a part of the logical layer. We decided to focus on maximizing the amount of BEP traffic carried while letting the FP traffic be specified by an input demand matrix. The reason for this

is because capacity planning in the Internet is typically done using an IP layer traffic matrix that specifies the average amount of bandwidth that needs to flow between any two routers or POPs in a domain. After we choose an initial matrix, we scale the entire matrix up, in order to load the maximum amount of FP onto our network. By “scaling up” we mean that we multiply all elements in the matrix by a constant factor that is as large as possible. The limit on how much the matrix can be scaled up is defined by the maximum amount we can protect. In other words, we cannot scale the matrix any further if it means that some FP traffic could not be protected.

The **IP routes** are those given by either the OSPF or IS-IS protocol that operates at the IP layer. All carriers today use one of these two standardized and widely adopted protocols. Since these protocols are unlikely to undergo any fundamental changes, we assume they are a fixed component on our network environment. These protocols usually compute shortest-path routes between routers. A path specified by OSPF (or IS-IS) is thus a sequence of *logical links*.

Both the FP traffic matrix and the IP routes are inputs to our problem. From these we can compute the **aggregated traffic demand on each logical link**. This demand is determined by the set of logical connections that share a given logical link. We calculate this aggregate load for each logical link by routing the FP traffic matrix over the logical topology according to the OSPF routes. Given these three things - the FP traffic matrix, the IP routes, and the logical link load (all coming from the IP layer) - we now have the inputs needed for our optimization problem at the physical layer.

As mentioned above, the optimization procedure needs to find a pair of disjoint fiber paths for each logical link. There are typically a large number of such possible pairs for each logical link. We search for combinations of pairs for all the logical links that maximize our objective function, which is to maximize network-wide load. We choose among the many candidate solutions by evaluating the corresponding maximal amount of BEP traffic that the network could carry. Since the BEP traffic is allocated as a certain amount to each logical connection, it can also be described as a **BEP traffic matrix** with the same rows and columns as the FP traffic matrix. Recall that the FP traffic matrix is scaled so that the network is saturated by as much FP traffic as can be protected. Thus the amount of BEP traffic that can be added onto the network using the remaining spare capacity is determined after we have satisfied the demands for FP traffic. Part of our optimization problem is thus to search for a mapping that will maximize the BEP traffic matrix, i.e., one that allows the largest amount of BEP to be carried. Note that this is the same as the mapping that maximizes network-wide load, since the FP traffic matrix is fixed. For a given topology of spare bandwidth, there are a multitude of ways in which supplementary BEP traffic bandwidth could be distributed among the many connections. We describe the strategies we use in our two solutions in Section 4. We will see that it is indeed possible to add a significant amount of BEP traffic.

In the Internet today, it is common practice for carriers to require that a certain percentage of all links be left free. In other words, average link utilization levels are not supposed to exceed some threshold, say 60% or 70%, for any extended period of time. Once a link starts to repeatedly exceed the specified threshold, then plans for a link upgrade are usually put in place. This type of requirement also comes from router vendors who

insist that link utilization levels shouldn't exceed about 80% or 90% otherwise routers can slow down to the point of introducing very large delays or even crash. We incorporated this **over-provisioning practice** into our problem via a factor we call β_{FREE} , which represents the fraction of each logical link that a carrier desires to leave free. Hence a $\beta_{FREE} = 20$ means that 20% of a logical link is unavailable to either FP or BEP. The value of $\beta_{FREE} = 0$ represents the case where the full capacity of the logical link can be used. This factor limits both the amount the FP matrix can be scaled up, and the amount of BEP traffic that can be added onto the network being evaluated.

This policy has an interesting side-effect. Because it usually applies to average values of load, it is acceptable to exceed these thresholds for a limited amount of time. A positive value of $\beta_{FREE} > 0$ means that there is excess capacity available that could be used in the case of failure to reroute some BEP traffic *temporarily*, until the IP layer can find another route for the BEP traffic. For example, when a failure happens one could avoid dropping all affected BEP traffic by load balancing a fraction of it on another nearby link. This alternate link would be used only during the OSPF route convergence time. As a result, it may not be necessary to drop *all* BEP traffic just after the occurrence of the failure and before OSPF has found new routes at the IP layer to restore this traffic. Thus overprovisioning allows some of the BEP traffic to experience service degradation, but not service interruption, when failures occur.

We now give the formal problem statement, incorporating all of the elements above.

GIVEN:

- i) a physical topology (which must be at least biconnected), whose nodes are optical cross connects (OXC) interconnected by optical fibers that support a limited number of wavelengths and have limited capacity.
- ii) a logical topology whose nodes are IP routers interconnected by logical links. These links have a finite limit on the total amount of traffic they can carry (including both FP and BEP). The limit comes from both their capacity and the over-provisioning factor.
- iii) an FP traffic matrix, denoted $D_{FP} = [d^{kh}(FP)] \geq 0$, that defines the FP traffic demand for each pair of routers (k, h) at the IP layer. We call these pair origin-destination (OD) pairs.
- iv) the routing paths selected at the IP layer for each OD pair of routers. This set of routes is denoted by \mathcal{R} . These are the routes determined by OSPF and specify the path through the network of logical links. (We implemented a shortest path computation to mimic OSPF's routing decisions.)
- v) an FP protection strategy at the WDM layer, either 1+1 or 1:1.

FIND

- i) the primary and backup paths for each logical link in such a way that the network is able to carry all the demand specified in the FP traffic matrix D_{FP} .

ii) the amount of BEP that can be added to each logical connection so as to maximize the total network load without impacting the protection of the FP traffic. This output is specified in the form of a BEP traffic matrix $D_{BEP} = [d^{kh}(BEP)] \geq 0$. The volume of this matrix is limited by the overprovisioning factor, β_{FREE} , defined per logical link.

4. APPROACH

We develop two solutions to this problem. This first uses optimization techniques to find an optimal solution based on formulating the problem as an Integer Linear Program (ILP). Although this approach can find optimal solutions, it is limited in its applicability since it becomes prohibitively expensive as the network size scales up. Our second solution defines a heuristic algorithm based on the Tabu Search (TS) methodology³ that can be used in practice for actual carrier backbone networks.

The objective of our ILP formulation is to maximize the total load carried by the network. Recall that this is equivalent to maximizing the amount of BEP traffic carried network-wide since the FP traffic is given as an input. The total BEP traffic is computed as the sum of the BEP traffic over all connections, or origin-destination pairs. We have two versions of our model, one for the 1+1 FP protection strategy, and one for 1:1 FP protection. We do not include the model here because it is quite lengthy and a full description can be found in .⁹

Even for moderate size networks, obtaining an optimal solution to this problem becomes quite cumbersome due to the large number of variables and constraints involved in its formulation. Indeed, a simpler version of this problem, in which one tries to optimize the network load for only one class of service, was already proven to be NP-complete.¹³ US backbone carriers can have upwards of 30 OXCs and 50 fibers in a physical topology, and upwards of 20 PoPs and 40 bidirectional logical links at the IP layer. In the image of the Sprint backbone we consider here, there were typically thousands of disjoint fiber paths for each logical link. The complexity of the optimal solution comes not only from this (and the large number of variables it implies), but also from the complexity of evaluating different allocations of BEP traffic (described below). It is thus clear that heuristic solutions are the only practical candidate solutions that carriers can consider using.

There are a multitude of ways in which BEP can be added to the spare capacity because there are many combinations of bandwidth that can be given to each connection, and each connection can route its BEP traffic on either the working or backup paths. Our ILP considers all possible strategies for adding BEP traffic and uses the maximization of the total network-wide load as the criteria for selecting the best solution. While this achieves the objective and finds a network-wide maximum, our experience shows us that this approach tends to lead to a very unbalanced distribution of the BEP load - giving large amounts of traffic to some connections and close to zero to others. In particular, single hop connections tend to receive a large amount of BEP while longer multihop connections receive very little. We believe that carriers would find this unappealing because of the unfairness. For that reason we include in our problem the concept of a *fairness policy*. In the ILP model this is represented by the Z_{min} factor that represents the minimum amount of BEP load to be assigned to each logical connection. Additional BEP load cannot be distributed among the logical connections unless each has received its minimum amount. This policy is a first step towards the policy of max-min fairness, where all

logical connections for which a logical link is the bottleneck receive the same share of the bandwidth left for BEP traffic. We will use max-min fairness,¹ which goes beyond this policy, as a heuristic to drive our Tabu Search algorithm. We chose the max-min fair policy because it is known to avoid penalizing long connections. We will see that the more fairly the BEP bandwidth is distributed, the less total BEP load the network will be able to carry. (This will be illustrated via examples for the heuristic in Section 5.1 and for the model in Section 6.4.)

Deciding how the excess bandwidth should be allocated among logical connections is a decision made at the IP layer. By choosing to follow a max-min policy, we reduce the usually vast number of possibilities. However, even using the max-min fair strategy that uniquely defines the BEP traffic allocated to each logical link, there are still two physical paths it can be mapped onto. If we have L logical links, we have 2^L possible BEP traffic matrices. The importance of fixing a BEP allocation strategy at the IP layer in the heuristic solution is that it helps to keep the complexity low, as opposed to the ILP model that considers all possible strategies.

We evaluate our solutions using two different topologies. The first mimics the Italian national backbone and represents a medium sized network. We consider two different versions of the medium network; in scenario 1 we have five OC-12 links and seven OC-48 links, whereas in scenario 2 we have four OC-12 links and eight OC-48 links because we allow one of the links to be upgraded. The results for the medium sized networks were obtained by solving the ILP model using the optimization package in². We validated our heuristic algorithm on the first of these medium networks, by comparing the heuristic solution with the optimal solution (see section 5.2). The second topology we work with mimics the Sprint US continental backbone and represents a large network. The physical WDM topology is depicted in Fig. 7 and the POP-to-POP logical topology is shown in Fig. 8. All inter-POP links are OC-48. The results for the Sprint backbone were obtained using our heuristic algorithm.

4.1. Example

To illustrate the inputs and outputs of our ILP solution, we provide an example using our medium-sized network under scenario 1. Figure 1 shows our Italian backbone network consisting of 10 nodes (circle symbols) and 12 links at physical layer, and 6 nodes (rectangular boxes) and 9 links at the logical layer. Some of the physical links are OC-12 links while others are OC-48. When we speak of an OC-12 physical link we assume the link has 8 WDM channels each of which transmits at OC-12 rates (622 Mbps). Similarly, we assume an OC-48 link consists of 16 WDM channels at 2.448 Gbps each. We assume that the line cards in the routers are at OC-48 speeds.

Figure 1 also indicates all of the inputs to our problem. A sample FP traffic matrix is given in D_{FP} . The logical connections and their corresponding OSPF routing are given at the bottom of the figure. For example, the connection from POP 0 to 6 would flow over logical links (0,9) and (9,6). Two assumptions are made: i) each logical link is bidirectional and ii) all the logical connections (k, h) and (h, k) use the same multihop path defined by OSPF (i.e. the sequence of logical links). The first one implies that each logical link carries the aggregated IP traffic in both the directions. The second one implies a upper triangular traffic matrix (see the

bottom-right of Figure 1 for the FP traffic matrix), i.e. the (k, h) entry in the FP traffic matrix is the sum of the aggregated FP traffic from node k to node h , and from h to k .

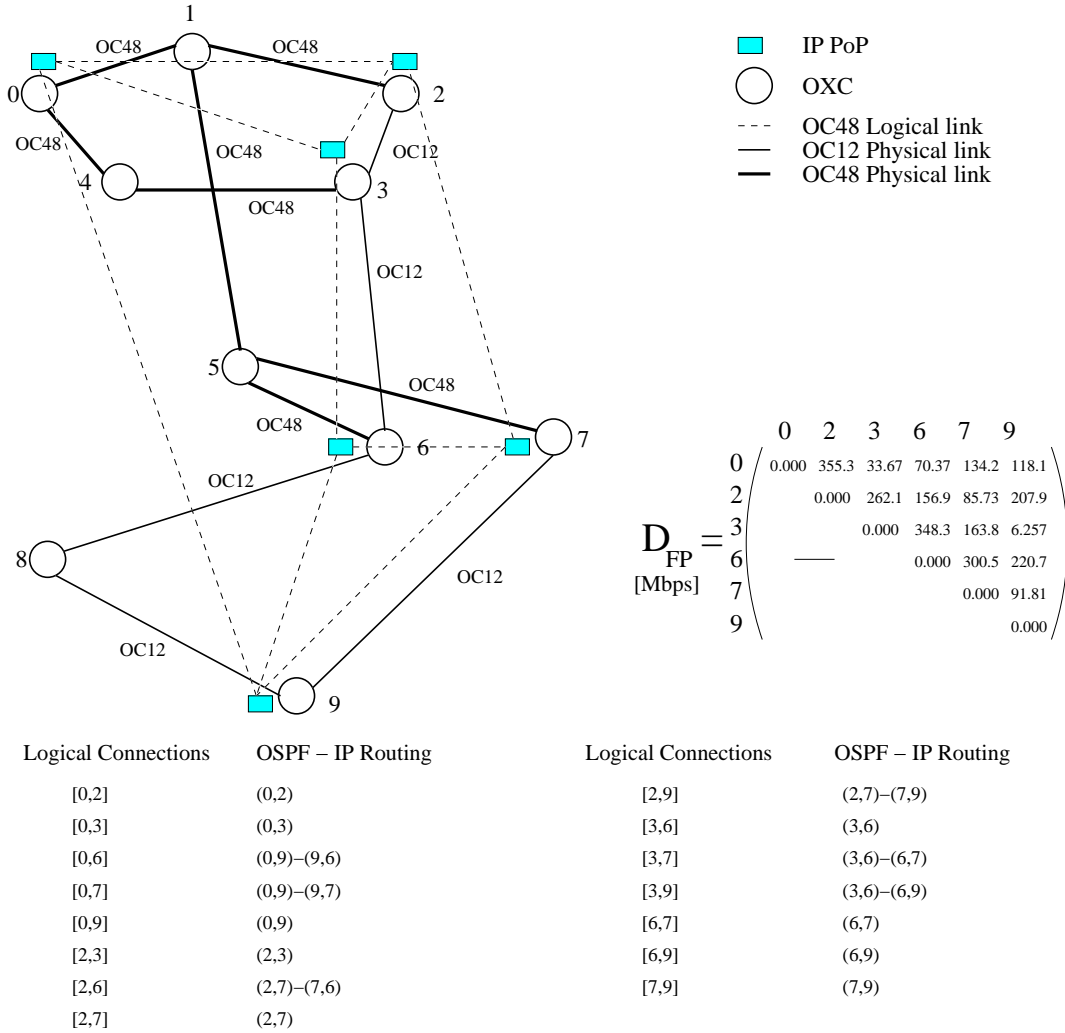


Figure 1. Italian backbone: physical and logical topologies - 10 nodes 12 links at WDM layer - 6 nodes 9 links at IP layer. 7 physical links upgraded to OC48. On the right is shown the FP traffic matrix considered D_{FP} .

The optimal solutions for this example are given in Figure 2; the solution for 1:1 protection is given on top, and for 1+1 on the bottom. Our solutions were obtained by solving the mathematical model using a Branch and Bound technique running ILOG CPLEX optimizer² over a 800 MHz Pentium III PC running Linux 6.2. The solution shows the working and backup paths selected for each logical link, the path chosen for the BEP traffic ([p]=w: working, [p]=b: backup), and the BEP traffic matrix (D_{BEP}). We also indicate the total BEP load and FP load in the figure. For both of the FP protection strategies, the gain realized, in terms of network load, by having two classes of service instead of only one is huge: 6.6 times bigger with 1:1 protection and 6 times with 1+1 protection.

This example illustrates how a bottleneck impacts the limitations on the scaling of the FP traffic matrix. In this example, we generated each element of the traffic matrix according to a uniform distribution between 1 and 50, and then scaled the matrix up as much as possible while still assuring all FP is protected. The amount of FP traffic that can be carried is limited because the router at node #9 has only OC12 interfaces. This bottleneck limits the amount of traffic that can be exchanged on logical links (9,0), (9,6) and (9,7). Since these cannot be scaled up by more than what's indicated in the traffic matrix, the rest of the FP matrix cannot be further scaled either.

Another important observation is the existence of a lot of zero entries in the BEP traffic matrix. This means there are a lot of logical connections that do not carry any BEP traffic at all. Note that these zero entries are assigned to multihop connections. As indicated in Section 3, we expected this to happen. This example illustrates our motivation to include fairness policies in our solutions to circumvent such outcomes.

1:1 FP Protection Strategy				FP load=2556 Mbps	BEP load=14313 Mbps					
Logical Links	Working Paths	Backup paths	[p]							
(0,2)	0->4->3->2	0->1->2	b		0	2	3	6	7	9
(0,3)	0->4->3	0->1->2->3	w							
(0,9)	0->1->5->7->9	0->4->3->6->8->9	b	0	0.000	2093	2414	0.000	0.000	622.0
(2,3)	2->1->0->4->3	2->3	w	2		0.000	2186	0.000	1997	0.000
(2,7)	2->1->5->7	2->3->6->8->9->7	w	3			0.000	1930	0.000	0.000
(3,6)	3->4->0->1->5->6	3->6	w	6				0.000	1827	622.0
(6,7)	6->8->9->7	6->5->7	b	7					0.000	622.0
(6,9)	6->5->7->9	6->8->9	b	9						0.000
(7,9)	7->5->6->8->9	7->9	b							

1+1 FP Protection Strategy				FP load=2556 Mbps	BEP load=13259 Mbps					
Logical Links	Working Paths	Backup paths	[p]							
(0,2)	0->4->3->2	0->1->2	b		0	2	3	6	7	9
(0,3)	0->4->3	0->1->5->6->3	w							
(0,9)	0->4->3->6->8->9	0->1->5->7->9	w	0	0.000	2093	2414	0.000	0.000	299.3
(2,3)	2->1->0->4->3	2->3	w	2		0.000	2186	0.000	1997	0.000
(2,7)	2->3->6->8->9->7	2->1->5->7	b	3			0.000	1930	0.000	0.000
(3,6)	3->4->0->1->5->6	3->6	w	6				0.000	1827	324.7
(6,7)	6->5->7	6->8->9->7	w	7					0.000	188.1
(6,9)	6->5->7->9	6->8->9	w	9						0.000
(7,9)	7->9	7->5->6->8->9	w							

Figure 2. Example of an optimal solution obtained by solving the ILP, with $\beta_{FREE}=0$ and $Z_{min} = 0$

5. HEURISTIC SOLUTION

TS is based on a partial exploration of the space of admissible solutions, starting from an initial solution usually obtained with a greedy algorithm, and ending when a stopping criterion is satisfied. The algorithm returns the best solution it found during the entire search. For each admissible solution, the algorithm defines a class

of neighboring solutions (the *neighborhood*) obtained from the current solution by applying an appropriate transformation, called a *move*. In each iteration of the TS algorithm, all solutions in the neighborhood of the current solution are evaluated, and the best one is selected as the current new solution.

In order to efficiently explore the solution space, the definition of neighborhood may change during the exploration of the solution space; this enables a *diversification* of the search in different solution regions. The TS algorithm can be seen as an evolution of the classical local optimum solution search algorithm called Steepest Descent ⁽⁸⁾. It can avoid getting trapped in local minima due to the TS mechanism that allows limited excursions toward solutions that appear worse than the current one.

The TS method introduces the use of a *Tabu list* to prevent the algorithm from cycling among already visited solutions. The Tabu list stores the latest accepted moves; as long as a move is stored in the Tabu list, it cannot be used to generate a new one. The choice of the Tabu list size is a key parameter of the optimization procedure: too small a size could cause the cyclic repetition of the same solutions, while too large a size can severely limit the number of applicable moves, thus preventing a good exploration of the solution space. The TS heuristic ends when a stopping criteria is reached. A common stopping criteria is simply to stop after some fixed number of iterations has been carried out.

5.1. Our Algorithm

We now state our algorithm by specifying how we implement each of the elements of a TS heuristic. We have added a preprocessing step that speeds up the rest of the search procedures.

1. *Preprocessing Step.* Generate the set of all the admissible pairs of disjoint physical paths that could be used for each logical link. This determines the *admissible* solutions, each of which contains a particular mapping for each logical link. Admissibility here only refers to the fiber paths being disjoint.
2. *Initial Solution.* For each logical link, randomly select one pair of disjoint physical paths. Choose randomly within the pair which physical path is assigned as the working path and which is assigned as the backup path. The aggregated BEP traffic flowing on each logical link can be sent either on the working path or on the back-up physical path. The path leading to the largest value of the objective function is chosen.
3. *Create Neighborhood.* Select a logical link at random. Keep the working path fixed and change the physical backup path. The set of all the admissible backup paths for the selected logical link defines the neighborhood of the current solution. (If the solution makes no improvement for 50 iterations, apply a different move based on the diversification criteria - described below - to generate a larger more diverse neighborhood.)
4. *Evaluation of Solutions in Neighborhood.* We need to evaluate each solution in the neighborhood and pick the best one. Only the solutions generated by selecting a *logical link* and *the couple of physical paths* not present in the Tabu List are analyzed during this step.

- (a) Check the capacity of each solution to ensure the protection requirements for the FP matrix are satisfied. If enough resources are not available on the two physical paths to protect the FP traffic (according to either 1+1 or 1:1 depending upon which is used), then the solution is discarded as infeasible.
 - (b) Determine an allocation of BEP traffic onto the spare bandwidth that maximizes our objective function. Consider putting BEP traffic on either the working or backup paths.
 - (c) Elect the best solution found in the neighborhood as new current solution.
5. *Update.* Update the Tabu list by adding the latest move used to generate the new current solution and removing the oldest. Update the best-solution-seen-so-far if the new current solution analyzed shows a larger value of the objective function.
6. *Repeat.* If number of iterations is less than threshold, go to step 3, else stop.

We now comment on some of these steps in more detail. The move we apply to create the neighborhood has two nice properties. The first one is the guarantee that all solutions in this neighborhood are admissible¹. The second property is that this kind of move makes it easy to implement a *diversification* step. For example, we can select a different number of logical links at each iteration, which will move us rapidly to another region of the solution space. We decided to apply the *diversification* only when a certain number of successive iterations fail to yield improvement. In our simulations this number is set to 50; when this number is reached we build a new solution by selecting a random number of logical links between 3 and 5. Note that after the diversification move has been done once, we return to the regular move based on perturbing a single logical link.

We check the feasibility of a solution by routing all the logical connections into the logical topology using the standard OSPF IP routing algorithm. Then each logical link (s, t) is routed over the physical topology using the physical paths selected by the TS metaheuristic. If sufficient resources are not available to protect the entire aggregated FP traffic on each logical link, then the solution is discarded. The next solution is then analyzed. Once we find an feasible solution, we move to step 4b.

As described in the previous section, the evaluation of all possible ways to add BEP traffic onto the spare capacity is a hard task. After candidate BEP bandwidth allocations are assigned to connections, we route these connections and calculate the aggregate BEP load per logical link. We must then assess whether this traffic should flow over the working or backup physical paths. Since this evaluation needs to be carried out for each solution in the neighborhood, using a computationally extensive algorithm becomes prohibitive. Rather than visit all possible ways of adding BEP traffic onto the network (as our ILP solution would do), we define a strategy for adding BEP traffic based upon carrier practices. Our strategy is to add BEP traffic onto the links according to a max-min fair allocation.¹ Our motivation for this approach comes from our experience using our optimal ILP solution on a variety of small networks. If the objective of the optimal solution is to find a

¹Note that we distinguish between *admissibility* that refers to two fiber paths being disjoint, and *feasibility* that refers to a set of paths having enough capacity to satisfy the protection needs

BEP traffic matrix that maximizes the total BEP load carried, then frequently the BEP load is very unevenly distributed - giving large amounts of traffic to some connections and close to zero to others. In particular, single hop connections tend to receive large amounts of BEP while multihop connections receive very little. We believe that carriers would find this unappealing because of the unfairness. Hence we adopt a max-min fair strategy, and define a greedy algorithm whose computational complexity is very small.

We implement max-min fairness as follows. The algorithm starts by routing all the logical connections into the logical topology using the OSPF IP routing algorithm, and by assigning zero BEP traffic to each connection. Then the amount of BEP traffic allocated to each connection is increased in equal increments until a logical link gets saturated. At this point, the BEP bandwidth allocated to all logical connections sharing this bottleneck is frozen (at an equal level for all of them). All the other connections, which do not share this bottleneck, can still receive additional BEP traffic, without impacting the bandwidth allocated to the frozen connections. We then proceed to increase in equal increments the bandwidth to all remaining unfrozen connections, until a new logical link becomes a bottleneck (i.e., saturated). The bandwidth assigned to connections traversing the new bottleneck are now frozen. The algorithm repeats until all the logical connections are frozen. At this point the bandwidth of each logical connection is determined by its own bottleneck.

To illustrate the heuristic, a small example is shown in Fig. 3, with three logical links (continuous lines) between four IP routers. Four logical connections (dashed lines) are routed over the logical topology. Each logical link is marked with a number that represents its capacity; it is equal to the minimum value between its own line card speed limit and the capacity of the WDM physical path taken by the logical link. Using our algorithm, the first logical link that will be saturated is (R_b, R_c) . Since the bandwidth available to BEP traffic on this link is 9 units, and the link is crossed by three logical connections (C_{ad} , C_{ac} and C_{bc}), each of them will receive 3 units of BEP traffic. Next our algorithm will assign 9 units of BEP to the remaining logical connection C_{cd} . Note that the *max-min fairness* algorithm assigns the minimum amount of 3 BEP units to each logical connection and the total BEP load carried by the network is equal to 18 units. Consider what would happen if fairness were not imposed and the goal was merely to maximize the total BEP traffic carried. We could do so by assigning 9 units to C_{bc} and 12 to C_{cd} , for a total of 21 units (3 more than the fair solution). This simple example illustrates the comment we have made before that imposing a fairness requirement on BEP allocations reduces the total BEP carried network-wide.

We fix the size of the *Tabu list* to be seven. This number was chosen based upon our experience running simulations for different kinds of network topologies and FP traffic matrices. The searching procedure is stopped when a given number of iterations is reached. The number of iterations should be chosen relative to the size of the network and to achieve a good trade-off between computational time needed and the quality (distance from the optimal solution) of the solutions reached. We set this parameter to 1500 for the medium-sized network (Italian backbone) and 5000 for the large-sized network (Sprint backbone).

5.2. Validation

Whenever one defines a heuristic algorithm it is important to compare its performance to an optimal solution on a network small enough for the optimal solution to handle. We used our medium sized network and created

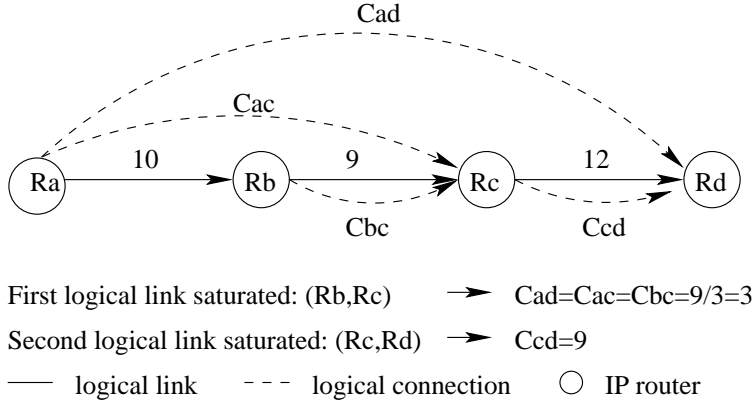


Figure 3. Example of FHLA algorithm.

a number of instances of this topology in which a different subset of the OC-12 links were upgraded to OC-48 links. For each instance of the topology we considered 30 FP traffic matrices. For all of these scenarios, both of the FP protection strategies were considered. We compared the solutions found by Tabu Search and the ILP model for this large set of test cases. Overall we found that the solutions were very close, and in some cases they were exact. Among all these test cases, the worst case difference between the objective functions achieved by our heuristic and the optimal solution was less than 3%.

6. NUMERICAL RESULTS

We now evaluate the performance of our two service proposal on our medium and large sized networks. Our goal is to quantify how much additional load can be carried on a network supporting two such services, to assess any performance degradation BEP traffic experiences in the case of failures, and to understand the interplay of other factors such a bottlenecks at each layer, upgrades and the overprovisioning factor. We use our ILP model to analyze the medium sized network and our heuristic to analyze the large commercial style network.

6.1. Methodology

We use six performance metrics to assess our approach. The first three of our performance measures are computed in a network without failures. The last three metrics are calculated over a number of scenarios in which single failures occur. For example, since there are 12 links in the Italian backbone, there are 12 failure scenarios. In each scenario we assume a different single link fails. Our six metrics are as follows.

1. We measure the BEP load that can be carried when there are no failures. The FP load is defined by the input traffic matrix since we carry all of it. The total network load is thus the sum of the BEP and FP loads.
2. The average and maximum utilization of the logical links when no failures occur.
3. The average and maximum utilization of the physical links when no failures occur.

4. The average amount of BEP traffic lost when a failure happens. We compute the amount of BEP lost in each failure scenario and present the mean, averaged over all the failure scenarios.
5. The average and maximum utilization of the logical links when a failure happens.
6. The average and maximum utilization of the physical links when a failure happens.

In the last two metrics, the average and maximum are taken over all failure scenarios. In all the utilization graphs, the levels indicated include both FP and BEP traffic.

Our ILP model itself does not explicitly take into account the various failures scenarios. In other words, the model does not compute routes that are optimal over all failure scenarios, but rather that are optimal only in the case of no failures (i.e., the full topology). To evaluate the performance of the ILP solution under a failure scenario we do the following. All of the FP and BEP traffic is routed according to the routes computed by the ILP model in the full topology. We then consider what happens if a single link fails. All logical connections not affected by the physical fiber failure retain their original physical routing paths. If the failure affects the working or backup path of a logical link, then the FP and BEP flows are moved over to the functional path. Recall that a physical link failure can affect multiple logical connections. We compute how much BEP traffic can be retained in each of the failure scenarios. The numerical results we present are averaged over all possible failure scenarios.

We present our six performance metrics as a function of the overprovisioning factor β_{FREE} . Recall that the β_{FREE} factor is a requirement on the *logical* link. Both 1:1 and 1+1 protection strategies were analyzed. For each test case we considered 20 different FP traffic matrices, and thus each point in the graphs below represents a value averaged over the 20 traffic matrices. Each traffic matrix is generated randomly according to a uniform distribution, where each entry is selected uniformly between 1 and 50 Mbps. Each matrix is then scaled up as much as possible.

6.2. The Bottleneck Issue

Before presenting the results, we first describe what we mean when we say the bottleneck is either at the WDM layer or the IP layer. We will see further below how the location of the bottleneck impacts the results. In order to send packets over OC-48 links, both the router and the optical cross-connect need to have the appropriate interface card. The upper limit of a *logical link* will be 2.5 Gbps (622 Mbps) if both the source and destination *routers* have OC-48 (OC-12) interface cards, respectively. The upper speed limit of a *physical connection* will be 2.5 Gbps if all the *OXC's* in both the primary and backup paths have OC-48 interface cards. If an OXC in one of those paths uses OC-12 cards, then that path will be the bottleneck. Let $IP_c(l)$ denote the capacity limit of logical link l at the IP layer. Let $WDM_c(l)$ denote the capacity limit of the primary and backup paths at the optical layer for logical link l . More precisely, $WDM_c(l)$ is the maximum of the capacities of the primary and backup path if FP traffic is protected on a 1+1 basis, and it is the sum of the capacities of the primary and backup path if FP traffic is protected on a 1:1 basis. If $WDM_c(l) < IP_c(l)$ then the bottleneck for l is at the WDM layer, otherwise it is at the IP layer. In all of our sample networks, we assume that the routers have OC-48 interface cards. The interface cards of the OXC's are indicated in each of the figures.

To illustrate this bottleneck issue, we return to the example in Figure 2. If we look at the working and backup paths enumerated here for the nine logical links, we see that 6 of the logical links have one physical layer path that is OC-48 and the other that is OC-12. For these links the total capacity at the *logical* layer is OC-48, while the total capacity available to that logical link at the *physical* layer is the sum of OC-48 plus OC-12 (for the 1:1 case which allows us to use backup bandwidth). Hence we consider the bottleneck to be at the IP layer because the logical connection will not be able to fill all of the capacity available to it at the physical layer. (Recall that we allocate one wavelength to a logical connection.) However for 3 of these links, both their working and backup paths have OC-12 rates, and hence their bottleneck is at the WDM layer.

6.3. Medium-Sized Heterogeneous Networks, Scenario 1

Our first set of results is given in Figure 4 for the Italian network with 5 OC-12 links and 7 OC-48 links. This set of results was obtained using our ILP model with the Z_{min} fairness factor equal to 0. When $\beta_{FREE} = 0$, the amount of BEP that can be carried is 4.5 times the FP load with a 1+1 protection strategy, and 5 times the FP load with a 1:1 protection strategy. Thus the total load increases by a factor of 5.5 under 1+1, and by 6 under 1:1. This demonstrates that the potential to increase the traffic carried on today's networks is huge. Even with $\beta_{FREE} = 0.5$, the amount of BEP is 7.5 Gbps for 1:1, which still allows for a tripling (3 FP and 7.5 BEP) of the regular load (FP only). It is intuitive that the additional BEP load carriable decreases linearly as β_{FREE} increases, since β_{FREE} is defined at the logical layer. We see that more BEP can be carried under a 1:1 protection strategy than under 1+1 protection. This was expected because in 1+1 protection, the reserved bandwidth is actually used for double transmission, while in 1:1 the reserved bandwidth is unused and hence available for BEP.

We computed (not shown in the graphs) that with FP alone, the average link utilization in the logical topology is 15%. (This matches typical utilization levels observed in many commercial backbones.) We see in the top-middle figure that with BEP, the average logical link utilization increases to 80% (with 1:1) and to 75% (with 1+1) for $\beta_{FREE} = 0$. Even when links are overprovisioned to leave 50% free ($\beta_{FREE} = 0.5$) we can increase the average logical link utilization from 15% to 50% (for 1:1) and to 44% (for 1+1). The fact that the utilization of the maximally loaded link decreases from 100% to 50% as β_{FREE} increases from 0 to 50% demonstrates the correctness of our implementation, because our target is for the maximally loaded link to be equal to $(1 - \beta_{FREE})C_l$ where C_l is the capacity of the maximally loaded link l . The curve for the two maximums is in fact the line $(1 - \beta_{FREE})C_l$ converted to percentages.

In Figure 4 on the top-right, we see that the maximum physical link utilization is at 55%. The average physical link utilization lies between 30% for $\beta_{FREE} = 0$ and 20% for $\beta_{FREE} = 0.5$; this is intuitive since increasing the β_{FREE} factor leads to a decrease of the *logical* link utilization and yields less traffic in the physical network. The fact that the slope of the decrease in average utilization at the optical layer is smaller than at the logical layer, is because the physical layer is not highly used since each logical connection is mapped to a single channel and we have 16 channels per fiber. This comes from the fact that 6 of the 9 logical links have their bottleneck at the IP layer. We also see that the 1:1 protection strategy requires a little bit less physical bandwidth than 1+1.

On the bottom-left figure, we see that when failures happen, we lose on average about 20% of the BEP traffic for both the protection strategies. The maximum BEP lost is approximately the same for 1:1 (50%) and 1+1 (48%). As β_{FREE} increases, there is a slight increase in the amount lost by 1:1, relative to 1+1. This makes sense since 1:1 carries a larger amount of BEP traffic than 1+1, it also loses more.

Examining the utilization levels under failure scenarios, we see that the logical topology is on average still well loaded even under failures. This is because 80% (on average) of the BEP traffic is in fact retained. The average logical link utilization drops by a corresponding amount of 20% while the average physical link utilization drops by around 10% (regardless of the protection strategy).

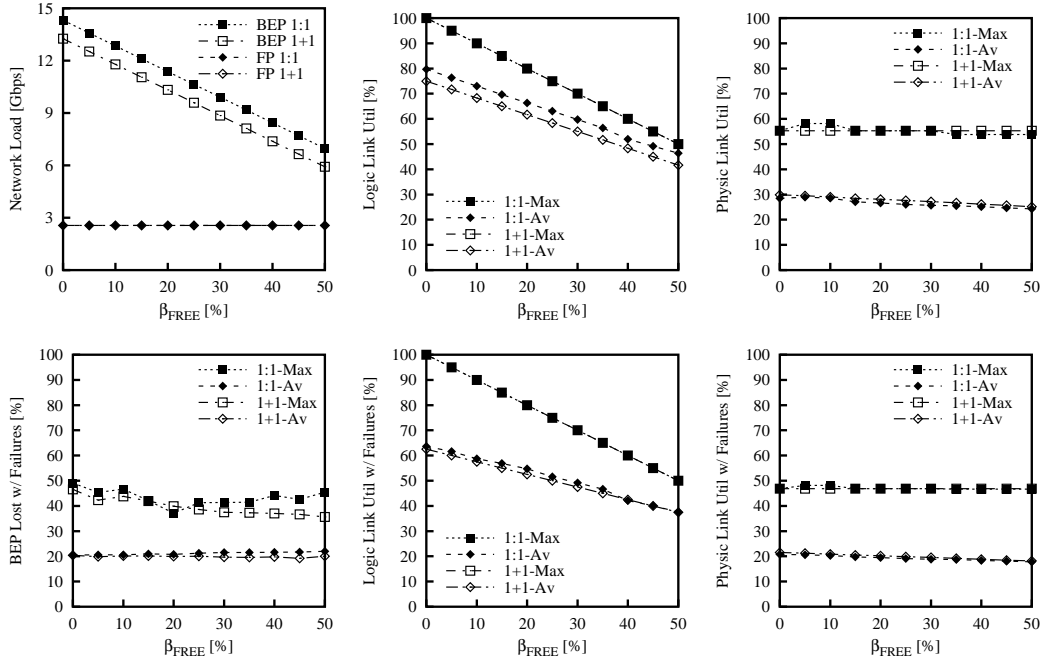


Figure 4. Italian backbone. ILP solution. 7 OC-48 links. $Z_{min}=0$ Mbps

6.4. Medium Sized Heterogeneous Networks, Scenario 2

We now present our six metrics for this second version of our medium-sized network in Figure 6. This topology differs from the previous one only in terms of link (7,9) which has now been upgraded from an OC-12 to an OC-48 link. One of the purposes of our FP/BEP approach is to take advantage of additional pockets of bandwidth that arise when one link is upgraded at a time. We study this topology to assess the gain our approach brings from a single upgrade.

First of all, we examine the impact of the the Z_{min} factor. Figure 5 shows the amount of BEP load that can be carried for this network. First we see that the BEP load decreases by a seemingly fixed amount each time Z_{min} increases by 100 Mbps. This illustrates the tradeoff that in order to give *each* logical connection some BEP traffic, we must reduce the total amount of BEP traffic carried.

Second we observe that the curves for $Z_{min} = 300$ and $Z_{min} = 400$ stop at some point as we increase β_{FREE} . This illustrates an important impact of the overprovisioning factor, namely that if we require a certain amount of overprovisioning, it can limit the minimum amount of BEP traffic we can offer. For example, if a carrier requires β_{FREE} to be 40%, then it is not possible to guarantee each logical connection a minimum of 400 Mbps of BEP traffic. Hence there is a tradeoff between overprovisioning and fairness. Intuitively this is reasonable since both require additional bandwidth, a finite resource that needs to be partitioned between these two objectives.

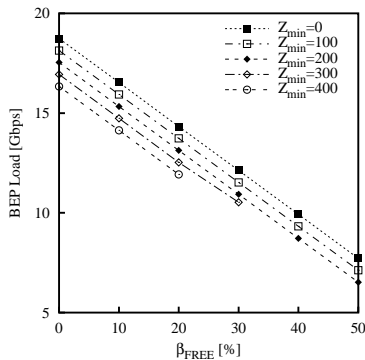


Figure 5. Italian backbone. ILP solution. 8 OC48 links.

We now present our six metrics for this second version of our medium-sized network in Figure 6. For this test case we set $Z_{min} = 200$ Mbps. The impact of the upgrade of link (7,9) is the following. All 9 logical connections now have one of their physical paths with an OC-48 rate and the other with an OC-12 rate. The bottleneck has thus been moved to the IP layer for all logical links. The most striking observation from all the graphs in Figure 6 is that the performance difference between 1:1 and 1+1 has disappeared. This reason for this is because the bottleneck is now at the IP layer. Recall that with two physical paths of OC-48 and OC-12, the total capacity available at the physical layer is 2.5 Gbps plus 622 Mbps. However since the routers have only OC-48 interface cards, the maximum amount of aggregated traffic a logical link can put into the network is 2.5 Gbps. When the bottleneck is at the IP layer, one cannot take advantage of idle backup bandwidth, such as in 1:1 protection.

In this network the total BEP load when $\beta_{FREE} = 0$ is at 17.5 Gbps; this is 6.25 times as much as the FP traffic, yielding an increase factor of roughly 7 for the total load. In scenario 1 for this network, the maximum increase in the total load was a factor of 6. The main point here is to illustrate that upgrading a single link can allow for a great deal of extra BEP to be carried in the network. This is advantageous because most carriers upgrade their backbone networks slowly; each link upgrade can take a few months. Thus carriers often find themselves in a position in which their backbone links are heterogeneous. Our FP/BEP proposal allows carriers to extract benefit from this heterogeneity.

We see other ramifications of the fact that the bottleneck has been moved to the IP layer. First, the average and maximum logical link utilizations are the same. Second, note that the maximum physical link utilization

in Figure 4 with seven OC-48 links and 14 Gbps of BEP load was at 55%, while in Figure 6 with eight OC-48 links and 18 Gbps of BEP load, the maximum physical link utilization is only at 50%. This is because we have upgraded link (7,9) and the bottleneck is now at the IP layer.

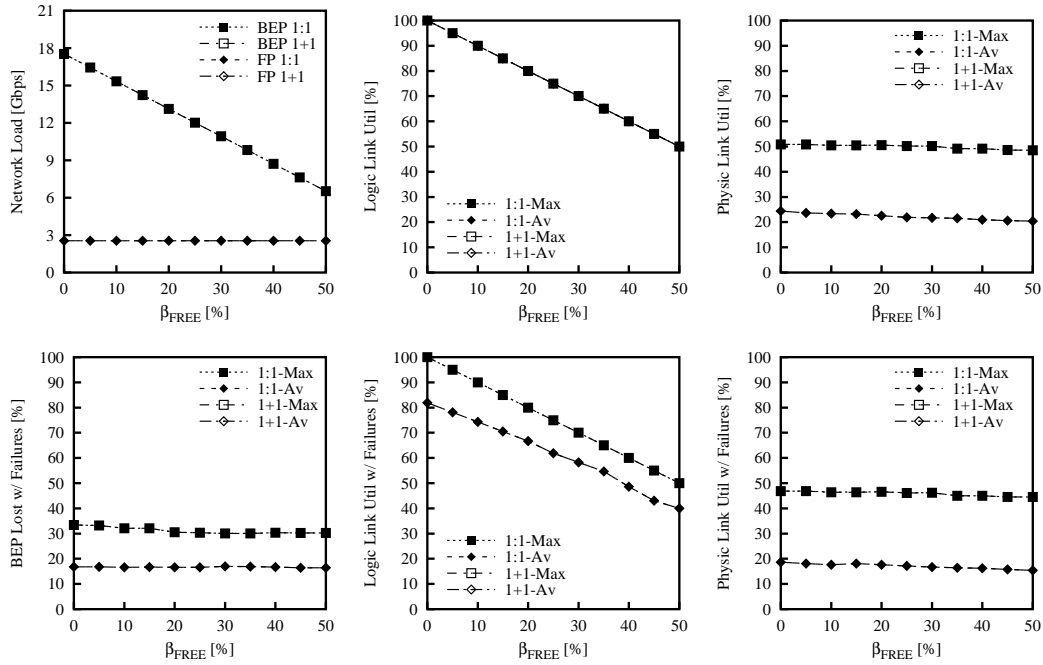


Figure 6. Italian backbone. ILP solution. 8 OC-48 links. $Z_{min}=200$ Mbps

6.5. Large Sized Homogeneous Network, Scenario 3: Sprint backbone

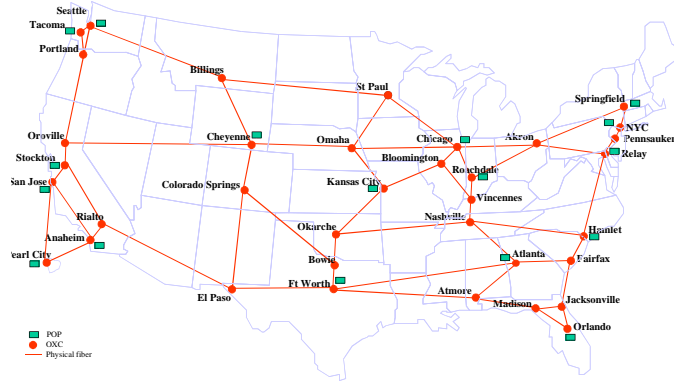


Figure 7. WDM Sprint Topology. All OC-48 links.

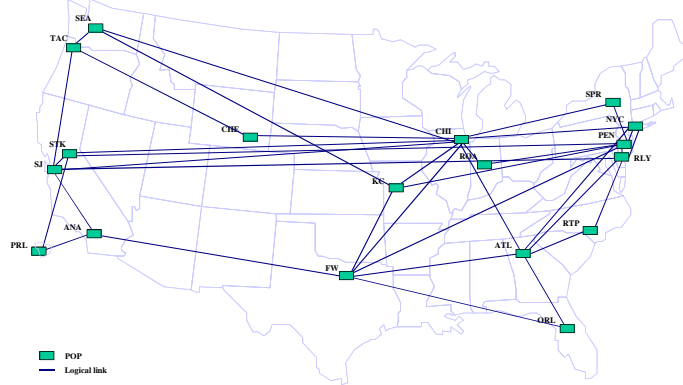


Figure 8. PoP to PoP Sprint Topology. All OC-48 line cards.

We now present the results of supporting the FP and BEP service classes on a large commercial network. Figures 7 and 8 display the two simplified versions of the WDM and IP layers actually used in the Sprint Backbone. The WDM layer consists of 36 OXC and 55 WDM fibers, while 18 PoPs and 36 logical links are present at IP layer. We assume that each logical link can carry no more than 2.448 Gbps of traffic (i.e., has router interfaces at OC48) and that each physical link consists of a WDM system with 40 channels OC48. In this network the amount of FP carried is around 9 Gbps and the amount of BEP carried varies from 52 Mbps (when $\beta_{FREE} = 0$) to 24 Gbps (when $\beta_{FREE} = 0.5$). Again we see that the total load carried under an FP/BEP system is about a factor of 5 bigger than the load carried on a network with only FP traffic. Even if carriers want to overprovision their network by 50%, they can still triple the load with the FP/BEP approach. Since FP traffic represents the current grade of service offered by carriers, our analysis illustrates the potentially large additional revenues, even in commercial networks, that our proposal enables. We did not include the failure cases here because there won't be any losses in this version of the Sprint topology. This is because the router cannot inject more than OC-48 rates of traffic, while both the working and backup paths are OC-48 (i.e., there is twice as much capacity at the physical layer as at the logical layer).

We see that the maximum logical link utilization is at the proper target of $(1 - \beta_{FREE})C_l$. This indicates that our heuristic is working properly. The average and the maximum logical link utilization are the same, as in the case of the second version of the Italian backbone. Again, the reason is because the bottleneck layer is at the IP layer, which thus fills its logical links to their maximum.

7. CONCLUSIONS

In this paper we defined two classes of service differentiated by their type of protection that can be used by carriers to increase the load carried on IP over WDM backbones. This problem is appealing and timely because carriers today operate their networks at low utilization levels and are interested in increasing their revenue without impacting the existing traffic. The FP class offers users the insurance that they will not suffer service interruption in the case of a single failure. Each logical link is protected at the WDM layer via either a 1+1 or

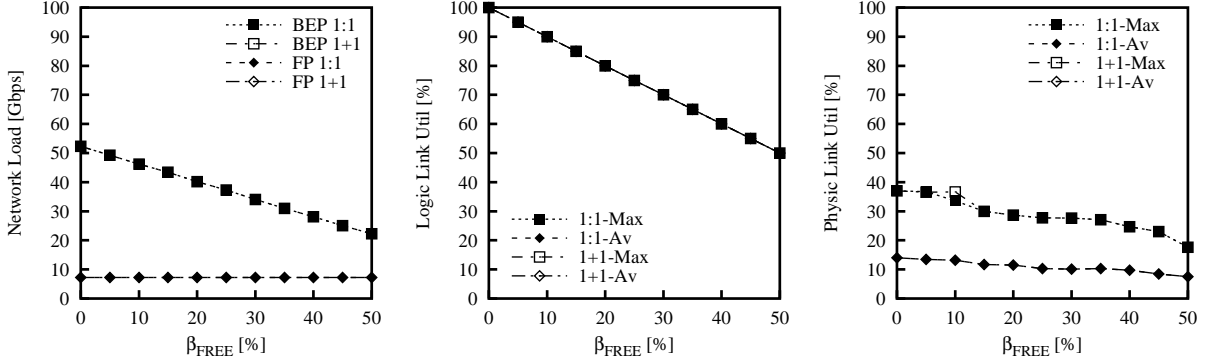


Figure 9. Sprint Backbone. Tabu Search solution.

1:1 protection scheme that guarantees fast recovery after a single failure. The lower-grade BEP class of service is new. When there are no failures in the network, BEP uses the “excess” bandwidth to generate additional revenue. In the case of failure, it does not provide a specific guarantee on service disruption, but it offers to restore as much of the affected traffic as possible. This restoration is taken care of at the IP layer.

In order to support two such services in an IP over WDM network, the logical links between routers at IP level must be carefully mapped over the physical topology. The mapping problem we studied includes a number of elements that arise in practice. We allow FP demands to be heterogeneous and specified via a traffic matrix at the IP layer. We consider real WDM systems at the WDM layer and heterogeneity in router and OXC interface cards at the IP layer. We incorporate an *over-provisioning* factor β_{FREE} , that represents an operational requirement to leave a given fraction of each logical link unused after both FP and BEP are allocated. Our service definition takes advantage of this because overprovisioning enables BEP to experience a degraded, but not disrupted, service during failure episodes. We introduce a *fairness policy* that enables carriers to ensure that there is some equity in how the offered load for this new service is distributed among logical connections; the more fairness we impose, the less BEP traffic we can carry.

We provide two solutions for the problem. The first finds an optimal solution using an ILP model and works well for small and medium sized networks. The second is a heuristic algorithm based on the Tabu Search methodology that achieves near optimal performance (in terms of an objective function) and yet is fast and scalable. We use this algorithm to study a large commercial US backbone.

We studied the gain of such an approach and found that it allows the total network load to be increased by a factor that varies between 4 and 7, depending upon the specific network scenario. We found that even if carriers want to over-provision their networks by 50%, and at the same time require a fair portion of the BEP load among the logical connections, we can still increase the total network load by a factor of 3.

We point out that the BEP traffic is never added at the expense of carrying less FP traffic. Recall that we scaled up the FP traffic demand to the maximum carriable that could still be protected. BEP traffic is added once we can no longer carry any more FP traffic. Thus by design, under our two service approach, the BEP

load has no impact on the maximum carryable FP load. Our simulations illustrated that this design goal is always achieved.

In our medium-sized network with heterogeneous links and interface cards, we observed that on average 20% of the BEP traffic gets dropped during a typical single link failure scenario. That means that even though we provide no protection guarantees at the WDM layer for BEP traffic, 80% of it remained unaffected by such a failure. In our large network with homogeneous interface cards at both the OXC's and the routers, there was no loss of BEP traffic. The reason for this is that all of the bottlenecks were at the IP layer. The IP layer cannot take advantage of the excess capacity at the optical layer in this scenario because it cannot transmit more than its own interface card limit. Also, when the capacity of each of the primary and backup paths is the same (i.e., homogeneous) as the capacity at the logical connection, then all of the traffic at the IP layer (FP plus BEP) can always be protected. We thus conclude that our proposal for FP/BEP service classes has the most gain or benefit in networks that are heterogeneous in their links and interfaces.

We demonstrated that the location of the bottleneck plays an important role. If in a network, there are some logical links whose bottleneck resides at the WDM layer, then we see a difference between the 1:1 and 1+1 protection strategies. In particular, the 1:1 strategy can carry more BEP traffic than 1+1. However if all of the logical links have their bottleneck at the IP layer, then there is no difference between the two protection strategies.

Finally we illustrated here that upgrading a single link can have a very large impact on the amount of BEP traffic carried. Since the process of upgrading all the links in a large backbone can take many months, our FP/BEP proposal allows a carrier to make use of dispersed pockets of additional bandwidth to increase their revenues.

In our ongoing efforts, we are incorporating the failure scenarios into our problem so that the routes computed are optimal over all possible failure scenarios. We are also studying scenarios in which all the protection or restoration is done at the IP layer because some carriers are considering eliminating SONET from their backbones.

REFERENCES

1. Bertsekas and Gallager. *Algorithm - Data Networks*, volume 2nd edition. Prentice- Hall, 1992.
2. Ilog cplex software optimization suite. <http://www.ilog.com/products/cplex/>.
3. E. Taillard F. Glover and D. De Werra. A user's guide to tabu search. *Annals of Operations Research*, 41:3-28, 1993.
4. O. Gerstel and R. Ramaswami. Optical layer survivability: A services perspective. *IEEE Communication Magazine*, 38(3):104-113, 2000.
5. E. Modiano and A. Narula-Tam. Survivable routing of logical topologies in wdm networks. *in Proc. INFOCOM 2001*, 1:348-357, Anchorage, April 2001.
6. G. Mohan and A. K. Somani. Routing dependable connections with specified failure restoration guarantees in wdm networks. *Proc. Infocom 2000, Jerusalem*, April 2000.
7. M.Sridharan and A. K. Somani. Revenue maximization in survivable wdm networks. *Opticomm, Dallas*, October 2000.

8. G. L. Nemhauser, A. H. G. Rinnoy Kan, and M. J. Todd. *Optimization - Handbooks in Operations Research and Management Science*, volume 1. North- Holland, 1989.
9. A. Nucci, N. Taft, P. Thiran, H. Zang, and C. Diot. Increasing the link utilization in ip over wdm networks. *Proc. Spie OPTICOMM, Boston*, August 2002.
10. J O.Crochat and J.Y.Le Boudec. Design protection for wdm optical networks. *IEEE Journal on Selected Areas in Communication*, 16, n.7:1158–1165, September 1998.
11. J O.Crochat, J.Y.Le Boudec, and O. Gerstel. Protection interoperability for wdm optical networks. *IEEE Transaction on Networking*, 8, n.3:384–395, June 2000.
12. K. Papagiannaki, S. Moom, C. Fraleigh, P. Thiran, F. Tobagi, and C. Diot. Analysis of measured single-hop delay from an operational backbone network. *IEEE Infocom 2002, New York*, June 2002.
13. R. Ramamurthy and B. Mukherjee. Survivable wdm mesh network. *Proc. Infocom 1999, New York*, March 1999.
14. Sprint operations center. *Private Communication*.
15. P. Thiran, N. Taft, C. Diot, H. Zang, and R. MacDonald. A protection-based approach to qos in packet-over-fiber networks. *International Workshop on Digital Communications*, September 2001. Taormina, Italy.