Increasing the Robustness of IP Backbones in the Absence of Optical Level Protection

F. Giroire[1], A. Nucci[2], N.Taft[2], C. Diot[2]
[1] – Ecole normale supérieure, Paris, FRANCE
[2] – Sprint Advanced Technology Labs, Burlingame, CA 94010, USA
email: frederic.giroire@ens.fr, {anucci,nina,cdiot}@sprintlabs.com

***Abstract*— There are two fundamental recent changes in IP backbone design techniques that challenge the robustness of such networks. First, SONET protection is gradually being removed because of its high cost (SONET framing is kept for failure detection purposes). Protection and restoration are provided by the IP layer that operates directly over a DWDM infrastructure. Second, constraints on Service Level Agreements force ISPs to systematically use the shortest distance path between two Points of Presence. In this context, IP backbones are extremely vulnerable to fiber cuts that can bring down a significant fraction of the IP links. We propose a heuristic to optimally map a given IP topology on a fiber infrastructure. The optimal mapping maximizes the robustness of the network while simultaneously maintaining the ISP's SLA delay requirement. In addition, our heuristic takes into consideration constraints that are faced by backbone administrators, such as a shortage of wavelengths or priority among links. The heuristic is evaluated on a real fiber network and IP topology and observations are discussed.**

I. INTRODUCTION

Most IP backbone networks are designed on top of a Dense Wavelength Division Multiplexing (DWDM) infrastructure. IP links are matched on wavelengths using static Wavelength Routing (WR). Wavelength routed IP networks contain two types of nodes, high-speed IP routers and optical cross connects (OXC). This setup allows for the definition of semi-permanent optical pipes called *lightpaths* that may extend over several physical links (i.e., fibers). At the IP layer, these lightpaths correspond to *logical links*. In DWDM, each logical link is assigned one wavelength if wavelength continuity is required, or a sequence of wavelengths if wavelength conversion equipment is present. In this environment, several logical links (each using a different wavelength) may traverse the same fiber (or the same conduit), making the IP network very vulnerable to a physical link failure such as a *fiber-cut*.

In the past, SONET was used to offer protection and fast restoration of service. However, due to the cost of optical equipment, most ISPs do not use SONET protection anymore and rely on the IP layer to protect and restore the service.[1] When a physical link fails in the optical network, IP routers detect the failure and update their routing tables with alternate paths. This approach only succeeds if the set of non-disrupted logical links still forms a connected topology. For each possible failure, we want to ensure that there will be other logical links (and other physical paths) unaffected by the failure so that alternate routes can be found by the IP routing protocol. It is thus of fundamental importance to map the logical links onto the physical topology in a way that minimizes the impact of physical topology failure on the IP network.

In this paper we focus on this *mapping problem* in the context of a *Tier-1 backbone network*. To the best of our knowledge, we offer here the first solution to this problem that incorporates real *characteristics* and *requirements* facing carriers today. One important characteristic comes from the fact that Points-of-Presence (POPs) in the backbone are usually interconnected via multiple logical links. The number of inter-POP links varies across neighboring POP pairs and is typically between two and six. Fault resilience is achieved by mapping these parallel logical links onto physical links that are as disjoint as possible. A second characteristic comes from carriers' desire to prioritize some inter-POP links as more important than others. For example, inter-POP links that tend to carry the largest amount of traffic should be assured higher fault resilience than links carrying small amounts of traffic. A third characteristic comes from the almost complete absence of wavelength converters in the DWDM layers and from the diversity of fiber quality (fibers can support between 8 and 80 wavelengths). Therefore shortage of wavelengths in such a network is not a rare event.

---

[1] SONET framing is used though to allow a fast detection of link failure.

In addition to these characteristics, the most stringent requirement comes from the need to meet *Service Level Agreements* (SLA). One of the SLA parameters is an upper bound for the maximum end-to-end delay. This value is found at each ISP and takes values between 45 ms and 60 ms for the continental US. In case of failure, carriers do not want the delay performance to go above the SLA on the newly selected logical path. Therefore, parallel paths need to be designed in such a way that the SLA is met even in case of major fiber-cut.

There is a trade-off between ensuring maximal disjointness and achieving minimal delays simultaneously. There may not necessarily exist two short delay paths that are completely disjoint. We will illustrate the trade-off between these two carrier requirements in test cases closely mimicking a real backbone network.

We develop an Integer Linear Program (ILP) model, that includes all of the above features. Because such a model cannot scale to large backbone networks, we also develop a heuristic algorithm that relies on the application of the Tabu Search meta heuristic methodology [7]. We compare the solutions found by our heuristic algorithm to the optimal solutions given by the model and illustrate the near optimal performance of our algorithm. We then apply our algorithm to the Sprint IP backbone network. We study the extent to which disjoint paths can be found, while matching various operational constraints. We believe that this is the first study to consider the requirements of fault resilience and delays simultaneously.

Our results show that for roughly 85% of all neighboring POP pairs we can find completely disjoint fiber paths for at least two of the parallel logical links. We also illustrate that without the priority mechanisms we impose, the critical heavy links would not achieve the same degree of fault resilience as with the priority mechanisms. An important application of these findings is that carriers can use our algorithm to determine the limits of the mapping and where they should add fibers in order to improve resiliency. We illustrate the limitations of the existing fiber layout by showing that with a near optimal solution, we still have about 15% of neighboring POP pairs whose entire connectivity (all the parallel links) can be disrupted by a single fiber cut. Note that these POP pairs can still remain connected by routing through other POPs, but the number of logical hops would increase.

The rest of the paper is organized as follows. In the next section, we discuss related works. In Section III, we present the problem statement taking into consideration the many facets of the problem. We formalize the problem in Section IV and provide our heuristic algorithm in Section IV-C. We briefly compare the model and the heuristic on a tractable network. In Section V, we use our heuristic algorithm to study the Sprint IP backbone network. Finally, we summarize our findings in Section VI.

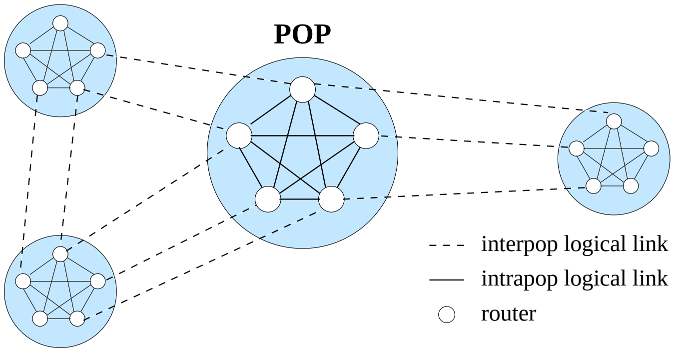

Fig. 1. POP interconnection with multiple links

II. RELATED WORK

Mapping logical links to the physical topology to assure connectivity during failures has already been studied [3], [4]. The problem is shown to be NP-complete [2]. An ILP formulation is provided in [5] and the problem is optimally solved for moderate size networks by applying a *Branch & Cut* algorithm. The major difference with our work is that these studies assume a single logical link between POP pairs and do not include delay requirements.

Topology mapping with wavelength constraints has also been studied in the literature. Without wavelength converters, the problem is known as the *wavelength assignment problem*, or WAP. A lot of research has focused on this well-known problem [9], [8], [10]. This problem is similar to the *path coloring problem* in standard graphs, which is in turn equivalent to the general *vertex coloring problem* [11]. It has been proven to be NP-complete [13], [14] and numerous heuristics have been proposed for different types of topologies [8], [12], [10]. In this paper, we provide our own solution to the WAP problem. The idea is not to present an improvement of previous methods, but rather to apply it in a practical problem. We were thus able to examine the impact of a shortage of wavelengths on the amount of disjointness achieved by our mapping.

III. METHODOLOGY

A. Problem statement

A Tier-1 backbone is made up of an ensemble of Points-of-Presence (POP) interconnected by logical links, as illustrated in Fig. 1. Each POP is its own mini-network in that it typically consists of a small number of core routers and a large number of access routers. The core routers

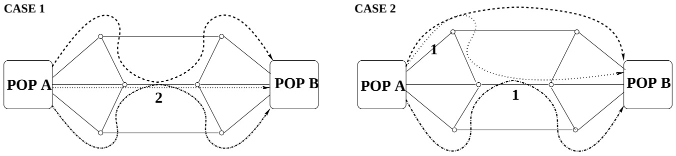

Fig. 2. Example of jointness and priorities

are fully meshed. Each of these core routers is attached to numerous access routers and customers connect into the backbone via these access routers (not represented in Fig. 1). Inter-POP links are attached directly to the core routers. The connection between two POPs is done by parallel logical links terminating at different core routers.

We define *neighboring POPs* as POPs that are directly connected by one or more logical links. Because of this structure there are typically several logical links between neighboring POPs. This number can vary between two and six for any given pair, and is thus heterogeneous across POP pairs. When we use the term *parallel links* we mean the set of logical links between a neighboring POP pair. For each logical link we need to find a physical fiber path. Backbone networks are designed this way in order to improve IP layer restoration that relies on an IP routing protocol. The improvement is obtained because load balancing is used across these parallel logical links. When one such link fails, traffic is quickly rerouted onto another parallel link. This load balancing activity requires fewer steps by the IP routing protocol for traffic to be switched to a parallel logical link than to recompute new network-wide paths. The use of parallel logical links is coupled with policies constraining the maximum acceptable load on each link. Thus in the case of failure, if one of the other parallel links is still functional and has spare capacity, it can carry the traffic of the failed link. The advantage of such a design is that the rerouted traffic does not travel on a longer path in terms of logical hop count and of delay.

B. Problem parameters

Due to the choices carriers make today in terms of equipment and POP design when building their IP backbones, the following issues and constraints need to be included in any practical version of the mapping problem. The issues to be addressed include: (1) protecting *multiple parallel logical links* between each pair of neighboring POPs, (2) meeting the SLA delay requirements and (3) handling the wavelengths limitations.

Network Protection and Disjointness. In order to ensure a high fault-resilience property, these parallel logical links need to be mapped onto the physical fiber network in such a way that a single fiber cut does not cause a failure for all of the logical links between a pair of POPs. The goal is to ensure that each pair of neighboring POPs remains connected via a single hop IP route, in the presence of physical link failures. In order to achieve this, the parallel logical links should be mapped onto physically disjoint fibers.

It is typically unachievable to map two to six parallel logical links onto disjoint fiber paths. This is because there is a limited set of conduits containing fibers in the ground and because these fibers have been layed out according to terrain constraints (mountains, rivers, etc.) and conveniences such as train tracks or pipelines rather. When we cannot find completely disjoint paths, our strategy will be to search for *maximally disjoint paths*. With this approach we try to minimize the number of logical links that are disrupted over all possible failures. Rather than define a metric for disjointness we define one for *jointness* because this is conceptually easier to understand. The objective of our ILP model and our Tabu Search heuristic algorithm will be to minimize this metric.

The jointness is defined for a pair of neighboring POPs (s,t). It expresses the level of ”sharedness” of fibers among parallel paths for a POP pair. The parallel links between s and t use a set of fiber segments (i,j). The jointness of a fiber segment for a given pair is the number of its parallel logical links sharing this fiber segment -1. Therefore, the jointness of a fiber segment used by a single link between s an t is zero. The jointness of a POP (s, t) is defined as the sum of the jointness of each segment (i, j). We call this *local jointness (LJ)* as it is defined for one pair of POPs. Note that a *local jointness* of 0 means that all the fiber-paths for a POP pair are disjoint.

Lets illustrate this definition on the example figure 2. Plain lines are fiber segments separated by OXCs. Dashed lines are parallel logical paths between POP A and POP B (3 in our example). In case 1, the jointness of POP pair (A, B) is 2 as only a single fiber segment is shared by 3 parallel links between A and B. Case 2, the jointness of POP pair (A, B) is also 2, with 2 segments being shared by 2 parallel logical links.

We define the *global jointness (GJ)* as the sum of the local jointness over all neighboring POP pairs in the backbone. We will use this metric to compare various mappings in sectionV.

Delay constraints. The Service Level Agreement (SLA) is a contract between an ISP and its customer. This contract specifies an maximum end-to-end delay between any *arbitrary POP pair* (not just neighboring POPs) that must be satisfied at any moment in time, both under normal operation and when failures occur. We introduce this constraint into the problem and assume that the delay primarily comes from propagation delay. We compute the IP route of the arbitrary POP pair according to the ISIS protocol. The delay of all the physical routes along these multi-hop IP routes must therefore not exceed the specified fixed threshold. The typical range of delay SLAs offered by continental USA carriers is 45-75ms.

The physical layout of fiber in today's network tends to yield the following situation. Two POPs that are geographically close have a high probability to have only one short fiber path and all other fiber paths are much longer (on the order 5 to 10 times longer). The more disjointness, the more we will find ourselves selecting an alternate fiber path that is very long. We want to avoid mapping logical links to physical paths that make long traversals across the USA. Such paths use a lot of resources and artificially increase the load on the network. Moreover a logical link with a long physical path is more prone to failure (than those traversing a small number of fibers). In order to avoid that, we include a second delay constraint in our problem, the *particular delay*. For each pair of neighboring POPs, we define a *default path* and require that the delay of all the parallel logical links of this pair are no longer than u% of the *default path*. In III-C we will define our different strategies to define this path.

Wavelength limitation. In DWDM networks each fiber has a fixed number of wavelengths. We need to perform wavelength assignment and verify that a sufficient number of wavelengths exist for a given mapping. In the case of no wavelength conversion, we have to verify that the same wavelength is available on all the fibers involved. Each solution we find will be composed of a set of physical paths, together with its wavelength. Therefore, limiting the number of available wavelengths significantly complicates our problem. A solution that is optimal from a disjointness standpoint might not be feasible from the wavelength allocation standpoint. In other words, assigning one wavelength to a logical link of POP pair A can reduce the possibilities of fiber path choice for POP pair B, and finally increase the jointness for other POP pairs. Therefore, our approach needs to take wavelength limitation into consideration in the computation of jointness.

C. Approach

Trying to find a solution that takes into consideration the above constraints is not an easy task as most of the time, one comes at the expense of another one. Trade-offs must be established to solve the problem. In this section we describe our approach for 2 points: how to protect the network and take into consideration the limitations of the topology (fibers, delay...); and how to handle the 2 constraints for the delay: end-to-end delay and particular delay.

Objective function. The central part of our problem is minimizing the amount of overlap in the physical topology of the logical links, i.e., minimizing the jointness. To achieve this we define two priority policies. The priority policies determine which logical paths should be favored in terms of ensuring their disjointness over other paths.

They are as follows. First, there are particular pairs of POPs in the network, called *priority POP pairs* because their inter-POP links are heavily used. For example, in the Sprint backbone these typically correspond to transcontinental east-west links. Thus when we carry out the mapping, we map these logical links by assigning fibers and wavelengths to them first. By prioritizing these POP pairs, we are effectively saying that achieving disjointness for these inter-POP links is important enough that their disjointness should be achieved even if it means more jointness for other inter-POP links.

Second, for each pair of neighboring POPs, we focus on finding *two* parallel logical links whose fibers are maximally disjoint. In the best case, the two paths will be completely disjoint. With complete disjointness, when a single failure occurs at the WDM layer, we are sure to have at least one lightpath not involved in the cut. With maximally (but not completely) disjoint paths, we minimize the likelihood of both paths being affected by a single fiber failure. We call these two paths the *priority logical links*. Note they are defined for each neighboring POP pair. Any two of the multiple logical links can be used for these priority logical links. If there are additional logical links to map, we try to find physical paths such that the entire set of physical paths (for all logical links) are maximally disjoint.

The objective function of our ILP model will be to minimize the global jointness. In the heuristic algorithm, we keep the same objective. In our approach we combine the issues of jointness and priorities together to define how we evaluate particular solutions. Because we have two priority policies: one for prioritizing some logical links over others within a neighboring POP pair, and one for prioritizing some pairs of POP among all the POP pairs, and because each of these policies has essentially two levels (high and low priority), we have an overall priority scheme with four levels. We thus work on minimizing the jointness of the mapping according to the following order.

1) minimize the Global Jointness for the two high priority logical links for the high priority POP pairs (*Priority Pairs GJ-2* in our notations).
2) minimize the Global Jointness for all the logical

links of the high priority POP pairs. (*Priority Pairs GJ-ALL*).

3) minimize the Global Jointness for the two high priority inter-POP logical links for all the pairs of neighboring POPs. (*All Pairs GJ-2*).
4) minimize the Global Jointness for all the logical links of all the pairs of neighboring POPs (*All Pairs GJ-ALL*).

To compare two solutions, we use first the first level of priority. The other levels are used successively as tie-breakers.

To illustrate the interaction of the jointness and priority mechanisms, let go back to the small example depicted in Figure 2. The local jointness of this pair of POPs for both solutions (case 1 and 2) is two. But case 2 is better since in case 1 a single fiber cut can take down all the parallel links. Our approach computes the jointness for the *priority logical links*, that is 0 for case 2 (as there exists a pair within the set of 3 logical links that are completely disjoint) and 1 in case 1. Thus the priority mechanism helps us to find the better solution when two solutions yield the same jointness.

Delay requirements. In addition to minimizing jointness, we must guarantee our two delay constraints: particular delay and maximum delay requirements. Our approach is to solve directly the particular delay constraint as our mapping effort is focused on mapping links between neighboring POPs. We will solve the end-to-end delay requirement indirectly by varying the values of u and the definition of the *default path*. We will give examples in Section V on how we convert this end-to-end delay requirement into requirements for inter-POP links.

Let now explain how we choose the *default path*. Because of the constraints of the physical fiber layout, we may not always have the luxury of selecting the shortest path as the default path length. If we used the shortest path with a small value of u, it could simply be infeasible to find paths to meet our requirements. We thus define three strategies for selecting the default path.

- *SP: Shortest Path Strategy*: the *default path* is defined as the shortest physical path between a given neighboring POP pair.
- *SSP: Second Shortest Path Strategy*: the *default path* is defined here as the second shortest physical path between a given pair of neighboring POPs.
- *SDP: Smallest Disjoint Path Strategy*: the third strategy is defined based on a different approach. For each pair of neighboring POPs we can always find two completely disjoint paths if we (temporarily) remove the constraints on path length differences and wavelengths. This is because the min cut of our network is two. Given these two disjoint paths, we select the longer of the two as our *default path*.

We will examine the impact of these strategies in our network. The consideration of different strategies allows a wider diversity of path selection that meet a large number of requirements simultaneously.

IV. FORMALIZATION OF THE PROBLEM

A. Problem definition

The problem that we addressed can be stated as follow:

GIVEN

- A physical topology of fibers, OXCs and IP routers with optical sections; the set of fibers, with their delay and the number of wavelengths supported for each fiber.
- The IP topology made of logical links interconnecting POPs

FIND

- Maximally disjoint physical paths for the logical links of all pairs of neighboring POPs, such that they satisfy the *particular delay* requirement.
- An assignment of wavelengths to logical links.

Note that the search for disjoint paths and the wavelength assignment tasks are conducted in parallel because the wavelength assignment has a direct impact of the feasibility of physical paths.

In a post-computation step, we compute the end-to-end delay for all arbitrary POP pairs, that is the maximum delay experienced over all alternate paths for each IP route. This route comes from the ISIS routing tables (based on link weights and shortest hop count) for all pairs of POPs given as input. We select between all the equivalent solutions for our problem the one with the minimum worst end-to-end delay.

B. ILP Model

We formulate the mapping problem as an Integer Linear Program (ILP) whose objective is to minimize the Global Jointness of the whole network. We carry out a computation step in which we compute all the default path lengths between each pair of neighboring POPs as defined in section III-C. These paths are used as input to the model. The maximum SLA achievable is calculated as a post-computation once all possible solutions have been identified by the model.

1) Notation: Let $\mathcal{E} = \{(i,j)\}$ denote the set of fibers and $\mathcal{S} = \{(s,t)\}$ denote the set of neighboring POP pairs. We use n^{st} for the number of inter-POPs links between the two POPs s and t. Let $\mathcal{S}_{priority} \subset S$ represent the subset of the priority pairs.

We let w_{ij} represent the number of wavelengths for fiber (i,j), and w_{max} the number of wavelengths available on the fiber with the most wavelengths. It will be used as bound for the channel index in the constraints. We introduce ${}^{(x)}a_{ij} \in \{0,1\}$ for all $(i,j) \in \mathcal{E}$ and $x \in \{1,2,...,w_{max}\}$ such that ${}^{(x)}a_{ij} = 1$ if the wavelength x belongs to fiber (i,j).

The notation pertaining to delays is as follows. Let $l_{ij} \geq 0$ be the length of the physical link (i,j) for all $(i,j) \in \mathcal{E}$. The values would typically be in the millisecond range. Let d^{st} for all $(s,t) \in \mathcal{S}$ be the delay between the POPs s and t using the *default path*. The maximum delay difference among all parallel links between each pair of neighboring POPs is specified via the parameter u.

2) Decision Variables: To compute the routing we define $\pi_{ij}^{st}(m)$ for all $(i,j) \in \mathcal{E}$, $(s,t) \in \mathcal{S}$, $m \in \{1,2,...,n^{st}\}$. We have $\pi_{ij}^{st}(m) = 1$ if the m^{th} physical link of the POP pair (s,t) traverses the fiber (i,j).

We now define the decision variables used to handle wavelengths. We use ${}^{(x)}\lambda^{st}(m)$, defined for all $(s,t) \in \mathcal{S}$, $m \in (1..n^{st})$, and $x \in n_{max}^{st}$, where ${}^{(x)}\lambda^{st}(m) = 1$ if the m^{th} logical link of (s,t) uses the wavelength x. We also define ${}^{(x)}\lambda_{ij}^{st}(m) \in \{0,1\}$ where ${}^{(x)}\lambda_{ij}^{st}(m) = 1$ if the m^{th} logical link of (s,t) traverses either the fiber (i,j) or (j,i) using wavelength x.

The decision variables for handling the SLA are as follows. Let $\Lambda^{st}(m)$ be the total length of m^{th} logical link of (s,t) for all $(s,t) \in \mathcal{S}$ and $m \in \{1,2,...,n^{st}\}$. The length of logical link is defined by: $\Lambda^{st}(m) = \sum_{(i,j)\in\mathcal{E}}(\pi_{ij}^{st}(m) * l_{ij})$. Let Λ_{max}^{st} be a length longer than the longest logical link of (s,t).

The jointness is computed in the model with two variables q and q' where q' denotes the jointness for the two *priority logical links* and q represents the jointness for all logical links. Having two variables for jointness allows us to compare separately how we do in terms of jointness just for the top two priority logical links (of all neighboring pairs) versus how we do in terms of jointness for all logical links of neighboring pairs. We define $q_{ij}^{st} \geq \sum_{m=1}^{n^{st}}(p_{ij}^{st}(m) + p_{ji}^{st}(m)) - 1$ for all $(i,j) \in \mathcal{E}$ and $(s,t) \in \mathcal{S}$. It is the number of paths of (s,t) - 1 that use the fiber (i,j). We define $q'^{st}_{ij} \geq \sum_{m=1}^{2}(p_{ij}^{st}(m) + p_{ji}^{st}(m)) - 1$ for all $(i,j) \in \mathcal{E}$ and $(s,t) \in \mathcal{S}$. If the 2 paths use the fiber (i,j), q'^{st}_{ij} is equal to one, otherwise it is null.

3) Constraints:

- The flow continuity constraints for the physical paths of the inter-POPs links of the pair of POPs (s,t) are:

$$\begin{gathered}\sum_{j\in V:(i,j)\in\mathcal{E}} \pi_{ij}^{st}(m) - \sum_{j\in V:(j,i)\in\mathcal{E}} \pi_{ji}^{st}(m) = \\ \begin{cases} 1 & \text{if} \quad i = s \\ -1 & \text{if} \quad i = t \\ 0 & \text{otherwise} \end{cases} \\ \forall i \in V, \forall((s,t),m) \in \mathcal{S} \times \{1..n^{st}\}\end{gathered} \tag{1}$$

Equation (1) defines the physical path associated with each logical link.

- Wavelength assignment.$\forall((s,t),m) \in \mathcal{S} \times \{1..n^{st}\}$,

$$\sum_{1\leq x\leq w_{max}} {}^{(x)}\lambda^{st}(m) = 1 \tag{2}$$

Equation (2) does the wavelength assignments for all the paths. One and only one wavelength is used on all the fibers of the path.

- The following equation ensures that the physical paths use only fibers where wavelengths are available. $\forall(i,j) \in \mathcal{E}, \forall 1 \leq x \leq w_{ij}, \forall((s,t),m) \in \mathcal{S} \times \{1..n^{st}\}$,

$$\pi_{ij}^{st}(m) \leq (1 - {}^{(x)}\lambda^{st}(m)) * B + {}^{(x)}a_{ij} \tag{3}$$

If the m^{th} path of the pair (s,t) uses the wavelength x, since $(1 - {}^{(x)}\lambda^{st}(m)) = 0$, the constraint becomes $\pi_{ij}^{st}(m) \leq {}^{(x)}a_{ij}$. And we see that $\pi_{ij}^{st}(m)$ has to be null if the fiber (i,j) does not support this wavelength.

- Equation (4) ensures that one wavelength can only be used once per fiber.

$$\begin{gathered}\sum_{(s,t)\in\mathcal{S}} (\sum_{m=1}^{n^{st}} ({}^{(x)}\lambda_{ij}^{st}(m) + {}^{(x)}\lambda_{ji}^{st}(m))) \leq 1 \\ \forall(i,j) \in \mathcal{E} : i < j, \forall 1 \leq x \leq w_{ij}\end{gathered} \tag{4}$$

For each fiber (i,j) and each wavelength x, only one ${}^{(x)}\lambda_{ij}^{st}(m)$ or ${}^{(x)}\lambda_{ij}^{st}(m)$ can be used, for all the logical links of all the paths.

- Constraints on ${}^{(x)}\lambda_{ij}^{st}(m)$. $\forall((i,j),x) \in \mathcal{E} * [1..w_{ij}] : i < j, \forall((s,t),m) \in \mathcal{S} \times \{1..n^{st}\}$,

$${}^{(x)}\lambda_{ij}^{st}(m) \geq \lambda_{ij}^{st}(m) + \pi_{ij}^{st}(m) + \pi_{ji}^{st}(m) - 1 \tag{5}$$

$${}^{(x)}\lambda_{ij}^{st}(m) \leq {}^{(x)}\lambda^{st}(m) \tag{6}$$

$${}^{(x)}\lambda_{ij}^{st}(m) \leq \pi_{ij}^{st}(m) + \pi_{ji}^{st}(m) \tag{7}$$

Equations (5), (6) and (7) ensure that ${}^{(x)}\lambda_{ij}^{st}(m) = 1$ if both ${}^{(x)}\lambda^{st}(m) = 1$ and $\pi_{ij}^{st}(m) = 1$, and 0 otherwise.

- We incorporate our constraint on the relative path lengths as follows.$\forall((s,t),m) \in \mathcal{S} \times \{1..n^{st}\}$,

$$\Lambda^{st}_{max} - \Lambda^{st}(m) \geq 0 \qquad (8)$$

$$\Lambda^{st}_{max} \leq d^{st} * (1+u) \qquad (9)$$

Equation (8) forces Λ^{st}_{max} to be longer than all the physical paths of the pair of POPs (s,t). The minimization process will search for solutions less than this largest value. Equation (9) requires this largest value to be within $u\%$ of the delay of the default path length for (s,t).

4) Avoiding loops: The flow continuity constraints (1) are insufficient to guarantee that our physical paths avoid loops. To solve this problem, we add some constraints as proposed in [15]. The principle is to make sure that a path uses only fibers that are a subset of the physical topology called a covering tree.

5) Objective function: The objective function is to minimize:

$$\begin{array}{c} B^3 * \sum\limits_{(i,j)\in\mathcal{E}} \sum\limits_{(s,t)\in\mathcal{S}_{priority}} q'^{st}_{ij} + \\ B^2 * \sum_{(i,j)\in\mathcal{E}} \sum_{(s,t)\in\mathcal{S}_{priority}} q^{st}_{ij} + \\ B * \sum\limits_{(i,j)\in\mathcal{E}} \sum\limits_{(s,t)\in\mathcal{S}} q'^{st}_{ij} + \\ \sum_{(i,j)\in\mathcal{E}} \sum_{(s,t)\in\mathcal{S}} q^{st}_{ij} \end{array} \qquad (10)$$

The four components of the objective function correspond to the four priority levels defined in section III-C. The first term is to minimize the jointness for the two priority logical links for the priority POP pairs. The second, third and fourth terms correspond to items 2,3 and 4 in the list in section III-C. For each component, we are trying to minimize the corresponding jointness. The parameter B represents a large number much larger than the sum of all the jointness parameters. By multiplying the first term by the largest factor B^3 we ensure that this type of jointness is minimized first. Whenever there is a tie, i.e., two solutions produce the same jointness for term one, then term two is used to break ties. The rest of the objective function is structured the same way. At each step, the model will find completely disjoint paths if possible and otherwise it will find maximally disjoint paths.

C. Tabu Search Heuristic

1) The general idea: The heuristic we propose for the solution of the problem relies on the application of the Tabu Search (TS) methodology [7]. TS is based on a partial exploration of the space of admissible solutions, finalized to the discovery of a good solution. The exploration starts from an initial solution that is generally obtained with a greedy algorithm. When a stop criterion is satisfied, the algorithm returns the best visited solution.

For each admissible solution, a class of neighbor solutions is defined. A neighbor solution is defined as a solution that can be obtained from the current solution by applying an appropriate transformation, and is also called a *move*. The set of all admissible moves uniquely defines the *neighborhood* of each solution.

At each iteration of the TS algorithm, all solutions in the neighborhood of the current one are evaluated, and the best is selected as the new current solution. Note that, in order to efficiently explore the solution space, the definition of neighborhood may change during the solution space exploration; in this way it is possible to achieve an *intensification* or a *diversification* of the search in different solution regions.

A special rule, the *Tabu list*, is introduced in order to prevent the algorithm to deterministically cycle among already visited solutions. The Tabu list stores the last accepted moves; while a move is stored in the Tabu list, it cannot be used to generate a new move. The choice of the Tabu list size is very important in the optimization procedure: too small could cause the cyclic repetition of the same solutions, while too large would severely limit the number of applicable moves, thus preventing a good exploration of the solution space.

2) Fundamental Aspects of our TS: Before describing more precisely our heuristic, we need to point out an important issue. During the search of an optimal solution, we have to go *outside the space of the admissible solutions*. By non admissible solutions, we mean solutions that require more wavelengths on some fibers than present in the WDM topology. All our solutions, even not admissible will always satisfy the SLA requirements. As a matter of fact, with some topologies with few wavelengths on fibers, find a single solution is hard as we have to solve the WAP. So we have to be able to start the heuristic outside of the space of admissible solution. We will operate a *strategic oscillation* (see [7]) between the space of admissible solutions and the space of non admissible solutions. When we are inside the space of admissible solutions we try to improve the current solution, when we are outside we try to come back inside by applying special kind of moves described later.

We now describe the fundamental elements of our heuristic in the 7 following points.

Precomputational step. Before running the TS, we evaluate and store some information that will help each round of the Tabu Search to be faster.

- For each pair of neighbor POPs, we compute the length of the *default length path* according to the three strategies described in III-C.
- For each pair of neighbor POPs, we build the set of the physical paths satisfying the particular maximum length constraint, i.e. that satisfy the *u parameter* requirement. This set is then sorted according to their length, from the shortest to the longest physical path.
- We build the IP route of all the arbitrary POP pairs according to the ISIS routing protocol.

Initial solution The choice of the initial solution is very important since it can significantly reduce the convergence time by investigating a smaller solution space. We choose the shortest physical path between neighboring POPs to be our initial solution.

Moves and Neighborhood generation. Since during the exploration we visit admissible and non admissible solutions, we define two different moves related to what kind of solution space we are visiting. The basic idea is that when the search is focused on the space of admissible solutions, the move adopted will select solution without considering the wavelength constraint, while when the search takes place out the space of admissible solutions the move will try to minimize the number of logical links that share fibers on which a shortage of wavelengths is present. We describe more precisely the two different kinds of moves as follows:

- *Admissible Space* Given an admissible solution for the problem studied, the new one will be generated following the three steps: i) random selection of a neighboring POP pair, ii) random selection of one of its parallel links not present in the Tabu List and then iii) change the physical path of this link by picking out a new path satisfying the maximum length constraint. All the other physical paths associated to all the other logical links will not be changed.
- *Not Admissible Space* When the current solution is not able to meet the wavelength constraint, the move tries to force the solution to become admissible by looking at the fibers on which the shortage of wavelengths was experienced. The new solution will be built following these three steps: i) random selection of a fiber that experiences a shortage of wavelengths, ii) random selection of a logical link that used this fiber and then iii) change the physical path of this link. All the other physical paths associated to all the other logical links will not be changed.

Then, given a current solution, its neighborhood is built by applying the *move* defined above for a random subset of all the physical paths. The cardinality of this subset defines the size of the neighborhood.

Wavelength Assignment Problem (WAP). While solving the mapping problem, we need to solve the WAP. As we explained previously, the WAP is an NP-complete problem. Since this problem must be solved for each solution visited during the exploration, a simple heuristic that is able to reach a good trade-off between running time and quality of the solution is needed. We remind to the reader that we focus on the mapping problem, and we try to provide only a soft solution for the WAP problem. The principle of the algorithm is to assign a wavelength when we select a new physical path. We look for the smallest channel index available on all the fibers of this path.

Tabu List used. In our implementation we use a *Static Tabu list*. We store the most recent move that we made, composed by its neighboring POP pair and its logical link. Each element is not allowed to be re-selected while kept into the list.

Diversification. To avoid exploring bad areas of the solution space for long time that are far from an optimal solution, we change the definition of the *move* if we do not see any improvements of the solution for a certain number of iterations (usually after 100). In this case, and only for this iteration, we introduce a perturbation by i) selecting randomly a neighboring POP pair and ii) changing all the physical paths of its parallel links. The selection of the paths is a random process. After having applied this perturbation, the traditional move defined previously is applied. We also apply the diversification for the wavelength assignment. After a given number of iterations, we reassign the wavelengths of the paths of a given POP pair. The principle is to try to find channels with lower indexes.

Stop criterion. The search procedure is stopped after a fixed number of iterations is reached. The number of iterations is defined based on the size of the network studied and a good trade-off between computational time needed and quality of the solutions reached. We set this parameter to 3500 for the network studied.

D. Validation

We compare the performance of our heuristic to that of the ILP model. As a matter of fact we have to prove that the heuristic works well in case of medium networks, so that we can guess that it will work well for big networks. Medium sized topology is defined as a limit topology that we can solve with the ILP model. We consider a physical topology with 15 nodes and 23 fibers that is a simplified version of the Sprint backbone. We generate *random logical topologies* made of 6 to 8 POPs and 6 to 10 neighboring POP pairs with 2 to 4 parallel links. For each of these topology choices, we ran several simulations with different number of wavelengths on the fibers. On average after

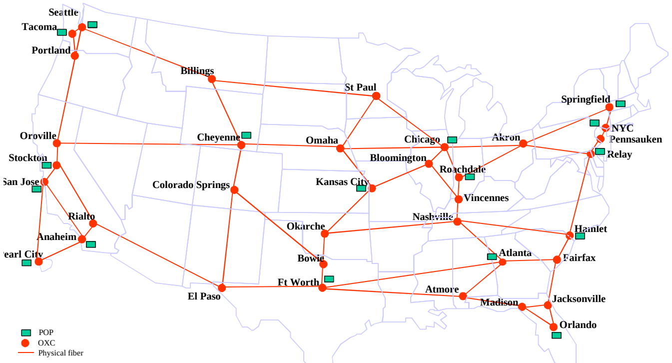

Fig. 3. Sprint physical topology

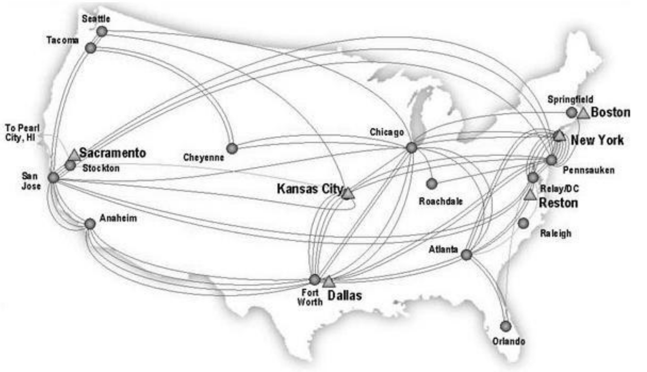

Fig. 4. Sprint logical topology

1000 iterations of the heuristic the results for the GJ differ by less than 3%.

V. RESULTS

A. *Topologies and metrics*

We use our TS heuristic algorithm to map the logical topology onto the physical topology of the Sprint US continental IP backbone.

Simplified views of the physical topology and of the IP topology are shown respectively in Fig. 3 and Fig. 4.[2] The WDM layer is composed by 51 OXC and 77 DWDM fibers that have between 16 to 48 channels. In the subsections V-C and V-D to point out the impact of wavelength shortage we reduced the number of wavelengths to 16 along a transcontinental path from San Jose to New York. The delay associated with each fiber is proportional to its length. The logical topology consists of 101 logical links, and 17 POPs that form 35 neighboring POP pairs. Each neighboring POP pair has a minimum of two parallel logical links, a maximum of six and an average of three. We have six priority neighboring POP pairs. The links between these pairs are typically transcontinental links. There is a total of 136 arbitrary POP pairs, whose IP route is between 1 and 4 hops long.

We use four metrics to evaluate our approach. The first two are the *Global Jointness (GJ)* and the *Local Jointness (LJ)* that we have defined in section III-B. The last two metrics are the *end-to-end delay* and the *POP-to-POP delay*. The End-to-End Delay is defined as the maximum delay experienced for an arbitrary POP pair. The POP-to-POP delay is defined for each pair of neighboring POPs.

B. *Jointness and SLA requirements*

In Fig. 5 we plot the four performance metrics identified above.

Our four performance metrics are plotted in Fig. 5. The terms SP, SSP and SDP in the legend refers to the strategies defined in section III-C. The suffix *-2* means that we give results for the two high priority parallel logical links, while the suffix *-ALL* is used for the results for all the logical links. The top-left plot illustrates the tradeoff between the disjointness of paths and the particular delay requirement u. We see that the global jointness decreases as the relative delay requirement increases. This is expected because the larger the value of u, the larger the set of acceptable paths, which makes it easier to find disjoint paths. If we restrict u to small values, then the set of possible paths is small and hence it is less likely to find solutions that avoid overlapping paths. The figure indicates that for values of u below 20%, the jointness is high. We also see that the curves flatten out for $u > 0.4$.

Fig. 5 also illustrates the impact of the strategies for selecting the default path length. The shortest path strategy produces the largest amount of jointness, while the smallest disjoint strategy yields the minimal jointness. This result is consistent for all the logical links as a whole and for the priority links. The gap between these policies is quite significant: with $u = 40\%$, the SP strategy yields a GI-ALL=96 while the SDP strategy can achieve GI-ALL=52. This illustrates an important property of the physical topology: it is often the case that only one short path exist between neighboring POPs while all the other paths are substantially longer. As a result the SP strategy may fail to find disjoints paths of acceptable delay.

The top-right graph of Fig. 5 plots the worst case end-to-end delay over all the paths of the arbitrary POP pairs as a function of u for each of the three strategies. As expected the overall delays increase as the particular delay constraint u increases. The gap between SP and SSP is not significant, however under the SDP strategy, the end-to-end delays are substantially larger.

By considering the top two plots in this figure together we can examine the relationship between global jointness and end-to-end delays. As we relax the particular delay

---

[2] For obvious reasons, we don't show exact data, but we use them in our computation of the mapping problem.

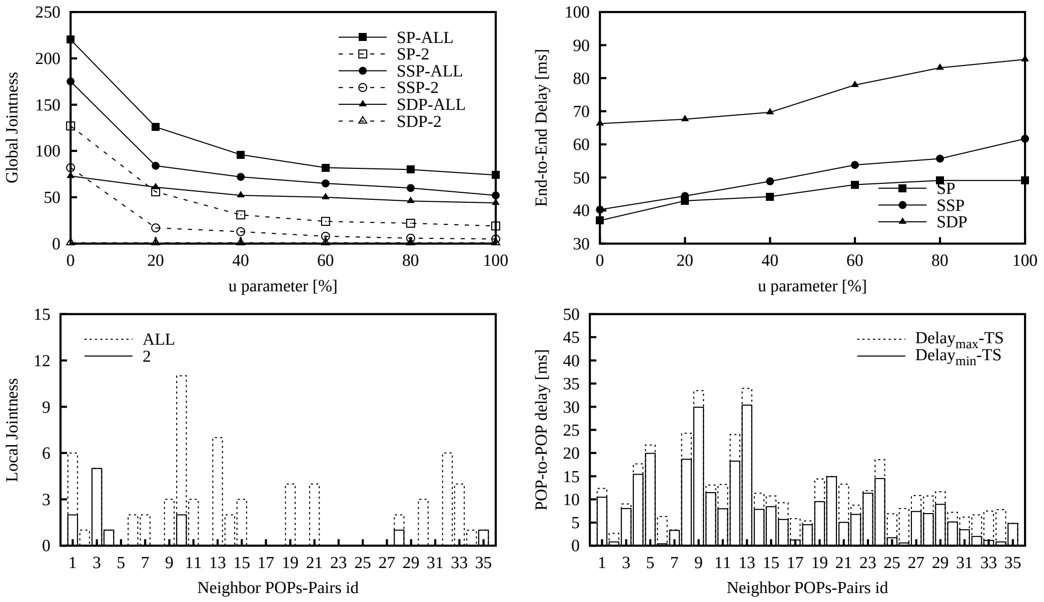

Fig. 5. The four performance metrics for Sprint Backbone: Global Jointness, End-to-End delay, Local Jointness and POP-to-POP delay.

constraint (i.e., increasing u), we decrease the jointness at the expense of end-to-end delay. We also see that the strategy for choosing default path lengths which performs best in terms of disjointness (SDP), also performs the worst in terms of end-to-end delay, and vice versa.

Using these two plots together, carriers can select the trade-off they want between fault resilience and delays. For example, if carriers select 70 ms as their maximum allowable end-to-end delay, they could use a $u = 0.4$ and the SDP strategy that would do best in terms of jointness (GI-ALL=60). However, if carriers insist that their maximum SLA be under 45ms, they will have to use $u = 0.2$ with the SSP strategy and a jointness of GI-ALL=80 (33% worse). On the other hand, with an SLA of 70 ms, we can have completely disjoint paths for all the priority logical links (GI-2=0).

For the remainder of our figures, we choose what we believe to be an operating point that achieves a reasonable trade-off. We use the SSP policy and u value of 50%. This corresponds to an end-to-end SLA of 50 ms and the global jointness for the priority links (GJ-2) is very close to zero. The two bottom plots in Figure 5 are based on these parameters.

The lower left plot shows the local jointness achieved for all pairs of neighboring POPs. We see that only 6 pairs cannot find two completely disjoint physical paths($LJ - 2 > 0$). For these pairs, and as a consequence of the delay constraint, we cannot guarantee that all the parallel logical links will not be disrupted simultaneously. For seven POP pairs we were able to find completely disjoint paths for all their parallel logical links (LJ-ALL =0).

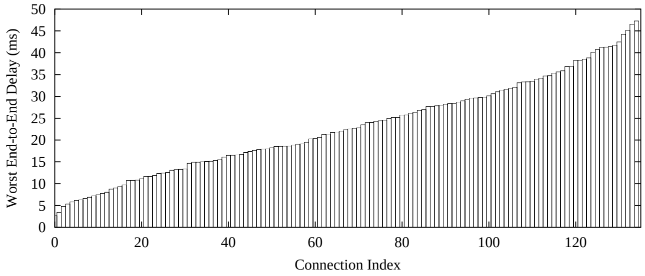

Fig. 6. Worst end-to-end delay for all arbitrary POP pairs

The lower right plot shows the *POP-to-POP delay* of the biggest and of the smallest physical path chosen by the heuristic. We see that for many pairs the difference between the minimum and maximum delay is small. There are also a large number of POP pairs whose minimum delay is very small (*e.g.*, the pairs with id 2, 3, 26). The second shortest path strategy (SSP) allows the network to achieve a reasonable diversity of fiber paths without a severe impact on the delays. Fig. 6 shows the impact of these POP-to-POP delays on the end-to-end delays of all the arbitrary POP pairs. We see that the limit of 50 ms is respected.

C. *The Impact of Priorities*

We now examine the impact of including our priority policies. Our model and heuristic essentially offers the network resources (wavelengths) to the priority POP pairs first. Recall that the priority pairs are more important to the operator because they carry a large amount of traffic and are traversed by a large number of IP routes. We had six priority POP pairs in our network, and two priority links per POP pair. Table I shows the global jointness metric for our network with and without priorities. The jointness is given for each of the four priority levels in our objective function.

	Priority Pairs		All Pairs	
	GJ-2	GJ-ALL	GJ-2	GJ-ALL
With priorities	0	20	33	103
Without	2	25	30	108

TABLE I

PROTECTING THE PRIORITY PAIRS

We see that with priorities all six of our priority POP pairs are guaranteed to have at least two completely disjoint paths for their logical links (GJ-2=0). If we remove the priorities on logical links, then two of these six pairs will no longer achieve complete disjointness; i.e., one third of the priority pairs can no longer be completely protected. Moreover the global jointness for these pairs is increased by 25% (20 to 25). Note that protecting the priority paths has a price since the global jointness over the whole network (all POP pairs) is worse with priorities (33) as opposed to without priorities (30).

D. Improving the Network Design

One of the critical steps for network designers to plan the future development of their network is to determine where to add either new fibers or new wavelengths. Our algorithm can be used as a tool by operators to determine where to add fibers and/or wavelengths in order to improve fault resilience.

In order to examine more closely the fault resilience achieved by our algorithm for the Sprint network, we look at the impact of specific fiber failures. In fig. 7 the grayness level indicates the fraction of logical links of a given pair using a given fiber. A black square means that 100% of the logical links of the POP pair use the fiber, while a white square means the POP pair does not use that fiber at all. For example, we see that all the logical links for pair # 10 use fiber #20. This figure demonstrates than 10 out of the 77 fibers will cause a pair of neighboring POPs to lose all of their direct connectivity. This corresponds to 5 POP pairs; hence 5/35 POP pairs can lose all of their inter-POP logical links due to a single fiber cut. The ids for these POP pairs are 1, 3, 4, 10 and 29. The rest of the network is well protected since in 66% of the cases of fibercuts, none of the other POP pairs would lose more than 50% of their parallel logical links.

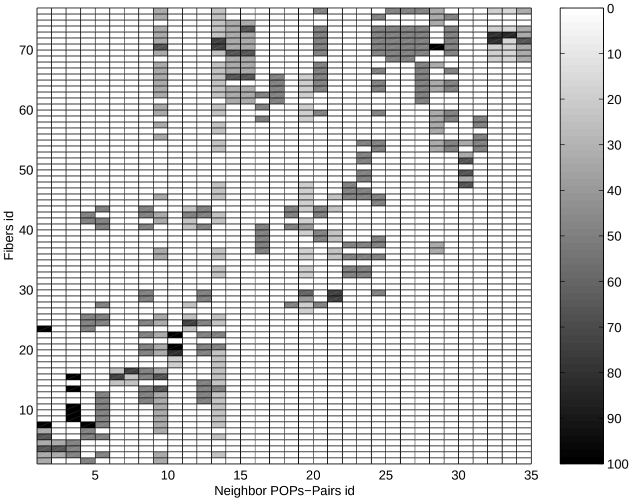

Fig. 7. Fiber network upgrade.

Using this visualization a carrier can quickly identify the location of fiber bottlenecks and use this information to plan future backbone upgrades. For example, we see[3] that there are two critical areas in the US: the North-West of the country where the alternative paths are very long and the North-East because of the high concentration of POPs and links in this area. Thus adding fibers in the northwest would increase the disjointness without paying a large price in terms of delay.

We carried out an upgrade example by adding three carefully placed fibers in our network, by using our algorithm to do the mapping, and then by recomputing the same type of performance data. We added fibers between the towns of Portland-Oroville, Oroville-San Jose, and Everett-Billings. With these fiber additions we were able to reduce the number of cases of critical fiber cuts (those disrupting all parallel paths for a given POP pair) from ten to four. The number of neighboring POPs with complete disruption of their parallel logical links was reduced from five to three. Hence the addition of a few new fibers in well chosen locations can seriously improve the protection of the logical topology while still meeting SLA requirements. This experiment also indicates the power of our algorithm to serve as an evaluation and capacity planning tool for backbone network designers.

It is usually more cost effective to improve one's network by upgrading existing fibers via adding wavelengths than by deploying new fibers in the ground. To know where it is interesting to upgrade fibers, we ran our algorithm without rejecting the solutions with shortage of wavelengths. The output of the algorithm gives us the

---

[3] The indexes of the fibers and neighboring POP pairs follows roughly a geographical order from West to East.

fibers that should be upgraded in order to eliminate the shortage of wavelengths. In the Sprint backbone, the results show that we can decrease by 66% the general jointness for the two priority links (GJ-2 decreases from 33 to 12) and by more than 25% the one for all the paths (GJ-ALL decreases from 103 to 71). Only an upgrade of 6 fibers (with an average of 5 wavelengths each) is necessary to achieve this result. This clearly illustrates that a small shortage of wavelengths has a huge influence on the robustness of the network.

VI. CONCLUSION

We propose a method to increase the robustness of IP backbones in the absence of optical level protection. The goal of the method is to minimize the jointness of parallel IP paths between POPs while minimizing the maximum delay of these paths. Our method takes into account a number of real constraints faced by carriers today, such as a limited number of wavelengths and priorities among logical links. This is the first effort to solve the mapping problem, in the context of large IP backbone networks, that incorporates the requirements of such networks.

Our method has been implemented as an ILP model and as a heuristic based on Tabu Search. We applied the method to a real IP backbone network. Our heuristic can provide an efficient mapping that matches most of the constraints. We evaluate the jointness for the Sprint backbone and identify a good operating point that achieves a balanced trade-off between disjointness, delays and priorities. At this operating point, only 5 out of 35 neighboring POP pairs (15%) could potentially lose their entire inter-POP connectivity via a single fiber cut. In terms of fibers, there are 10 out of 77 fibers whose disruption would bring down the entire set of parallel inter-POP logical links. In summary, we have shown that we can have a well-protected network with a maximum end-to-end delay of 50 ms despite the limitations of the existing fiber layout.

We also explain how our heuristic can be used to identify the areas where fibers or wavelengths should be added to the network in order to increase the robustness to fiber failures.

Our solution can be used by backbone networks designers. Carriers are motivated to study this problem because it has been observed that a single fiber cut does indeed bring down many links simultaneously and that such events can happen as frequently as once a month [16]. Our algorithm can be used to identify the trade-off between fault resilience requirements on path disjointness and SLA requirements on delay.

At this point, we plan to validate our tool with various topologies. We also need to compare our optimal mapping to the state of the art in various backbone networks. We also want to package our heuristic in such a way that it can be used by network administrators to help them increase the robustness of their network and minimize the SLA.

REFERENCES

[1] A. Fumagalli, L. Valcarenghi “IP Restoration vs. WDM Protection: Is there an Optimal Choice?”, *IEEE Network*, pp.34-41, November/December 2000.

[2] J.Armitage, O.Crochat, J.Y.Le Boudec “Design of a Survivable WDM Photonic Network”, *in Proc. INFOCOM ’97* pp.244-252, April 1997

[3] J O.Crochat, J.Y.Le Boudec “Design Protection for WDM Optical Networks”, *IEEE Journal on Selected Areas in Communication*, Vol.16, n.7, pp.1158-1165, September 1998.

[4] J O.Crochat, J.Y.Le Boudec, O.Gerstel “Protection Interoperability for WDM Optical Networks”, *IEEE Transaction on Networking*, Vol.8,n.3, pp.384-395, June 2000.

[5] E. Modiano, A. Narula-Tam “Survivable routing of logical topologies in WDM networks”, *in Proc. INFOCOM ’01* Vol.1, pp. 348-357, Anchorage, Alaska, April 2001.

[6] `http://www.ilog.com/products/cplex/`

[7] F. Glover, E. Taillard and D. De Werra “A User’s Guide to Tabu Search”, *Annals of Operations Research*, Vol. 41, pp.3-28, 1993.

[8] V. Kumar “Approximating arc circular colouring and bandwidth allocation in all-optical ring networks”, *First International Workshop on Approximation ALgorihms for Combinatorial Optimization Problems (APPROX’98)*,1998.

[9] B. Beauquier, J-C. Bermond, L. Gargano, P. Hell, S. Perennes and U. Vaccar “Graph problems arising from wavelength-routing in all-optical networks”, *2nd IEEE Workshop on Optics and Computer Science, part of IPP’97*, Genova, Switzerland, April 1997. IEEE Press.

[10] V. Auletta, I. Caragiannis and P. Persiano “Randomized path coloring on binary trees”, *3rd International Workshop on Approximation ALgorihms for Combinatorial Optimization Problems (APPROX’00)*, Vol.1913, pp.60-71, Springer-Verlag, 2000.

[11] I. Chlamtac, A. Ganz and G. Karmi “Lightpath communications: An approach to high bandwidth optical WAN’s”, *IEEE Transaction on Communications*, Vol.40, n.7, pp.1171-1182, July 1992.

[12] T. Erlebach, K. Jansen, C. Kakalamanis, M. Mihail and P. Persiano “Optimal wavelength routing on directed fiber trees”, *Theoretical Computer Science*, Vol.221, n.1-2, pp.119-137, 1999.

[13] D.S. Hochbaum editor “Approximation Algorithms for NP-Hard Problems”, *PWS Publishing Company*,1997.

[14] Eric J.Harder, Sang-Kyu Lee and Hyeong-Ah Choi “On Wavelength Assignment in WDM Optical Networks”, *Proceedings of the 4th International Conference on Massively Parallel Processing Using Optical Interconnections (MPPOI’97)*,1997.

[15] A. Nucci, N. Taft, P. Thiran,H. Zang and C. Diot “Increasing the Link Utilization in IP over WDM Networks” *Opticom 2002*

[16] G. Iannaccone, C. Chuah, R. Mortier, S. Bhattacharyya and C. Diot “Analysis of Link Failures in a Large IP Backbone” *To appear in Proc. 2nd ACM Sigcomm Internet Measurement Workshop (IMW),November,2002*