On the impact of relocation on network dimensions in resilient optical Grids.

Jens Buysse, Marc De Leenheer, Bart Dhoedt, Chris Develder

Department of Information Technology
Ghent University - IBBT, Ghent, Belgium
Email: jens.buysse@intec.ugent.be

Abstract—Optical networks, with their high bandwidths and low latencies, are excellent to support Grid applications, esp. those arising in so-called eScience applications. In this paper, we consider the resulting so-called optical Grids. We address the issue of resiliency against network failures and show how the Grid-specific anycast principle can be exploited in providing shared protection. Since in Grid scenarios users generally allow the Grid system to decide upon the location where jobs are executed, we exploit relocation to alternate backup sites in case of failures. We provide integer linear programming (ILP) solutions to the resulting routing and wavelength assignment problems, as well as a scalable heuristic. A case study on a European network topology shows that this relocation allows savings in the order of 16% of total wavelength capacity compared to traditional shared path protection.

I. INTRODUCTION

A. Resiliency in Optical Grids

Grid computing entails the combination of numerous dispersed computational-, storage- and network resources from multiple administrative domains in order to process highly demanding tasks. These problems can range from several research domains such as astrophysics, climate modeling and fluid dynamics. Due to the data intensive nature of these Grid jobs, the network interconnecting the different resources should be able to support high bandwidth traffic with a low latency in a reliable way. Clearly, Wavelength Division Multiplexing (WDM) has made the optical network an ideal candidate for these Grid applications, resulting in the introduction of the optical Grid concept. In this paper we will focus on the optical network using Optical Circuit Switching (OCS) technology, where bandwidth granularity is at the wavelength level since one or more wavelengths are allocated to a connection, while connectivity between source and destination is established using a two-way reservation.

An important aspect of network deployment is the ability to survive from certain network failures. The network operator can either choose to use some protective measures (Protection), or can opt to deal with the failure at the time it occurs (Restoration). In this paper we will deal with two protection schemes which protect a path from one point to another by reserving a back up path which can be used in case the primary path fails. The failure scenario we consider is where a link between two adjacent network nodes fails (single link protection) and as a consequence the network nodes cannot interchange information anymore. A traditional resiliency scheme is shared path protection where each primary path has a backup path where wavelengths can be shared between several back-up paths, as long as their corresponding primary light paths do not overlap. Its counterpart, dedicated protection, does not allow this kind of sharing. In this paper, we have extended the shared path protection algorithm by incorporating the anycast principle [1] which states that in general, multiple processing locations exist in a Grid network, so the exact location of execution (the destination of Grid jobs) is of less importance to the end user. Instead of reserving a back-up path to the original destination determined by the Grid scheduler, it could be better to relocate the job to another, possibly closer resource as illustrated in Fig. 1. This can result in an overall reduction of network capacity (albeit that there may be potential penalty in terms of Grid server resources to support the extra load for the server which receives the relocated jobs).

This paper is further structured as follows. After summarizing related work in the remainder of this section, we outline in Section II the principle of shared protection with relocation. Subsequently, in Section III we introduce ILP formulations for the considered protection schemes, both without and with relocation. We present the corresponding heuristics in Section IV and continue in Section VI with a case study validating the heuristics and investigate the influence of relocation on the network dimensions. The paper’s conclusions are presented in the final Section VII

B. Related Work

In [2] some preliminary work has been done concerning the influence of relocation on the network. In this previous work, we first dimension the computational resources by determining the best server site locations, calculating the server capacities and choosing an appropriate scheduling policy. These steps lead to a static traffic matrix which is the starting point for the ILP formulation modeling the traditional shared path protection strategy and the relocation strategy. The limited case study pointed out that it is possible to achieve a reduction of the total number of necessary wavelengths of 20%. This paper on the other hand, differs from [2] because we do not start from a static traffic demand matrix but from a demand vector expressing a number of requested connections per source. This implies that in the current work, the routing and
wavelengths assignment algorithms will include the choice of the destination site for each connection (i.e. job). In addition, we not only present integer linear programming (ILP) models, but also include a scalable heuristic. This allowed us to also evaluate larger problem instances in the case study presented further.

The relevance of focusing on protection against single link failures is demonstrated in [3]. There, the authors state that in order to provide complete protection from all dual-link failures, one may need almost thrice the spare capacity compared to a system that protects against all single-link failures. However, it has also been shown that systems designed for 100% single-link failure protection can provide reasonable protection from dual-link failures.

We will evaluate our proposed relocation strategy by formulating two Integer Linear programs (ILP). On the matter of this kind of network planning several studies have been performed. In [4] the main optical protection techniques for the WDM-layer are classified and reviewed. The authors of [5] investigate the problem of fault management in a WDM-based optical mesh network in which failures occur due to fiber cuts. Several off line-algorithms and heuristics are examined and their performance is compared through numerical examples. Our ILP formulation and heuristic for shared protection and relocation are based on the ones described in the latter paper. Yet, our differentiating contribution is to include also the case of relocation, and adapting the formulation to the Grid case (cf. the anycast principle, we also optimize the destination of the Grid jobs).

II. SHARED PATH PROTECTION WITH RELOCATION

In a standard fibre-optic path protection scheme where wavelengths can be shared, a light path is protected by a link disjoint light path going from the source to the destination of that light path. These back-up paths can be shared, in case the corresponding primary paths are unlikely to fail at the same time. Now, we can extend this technique by incorporating the anycast principle. In anycast, there is a one-to-many association between network endpoints: when a user creates a job, there is a set of several resources which are able to execute it and only one of them is chosen, generally by the Grid scheduler. Now, instead of protecting the path by allocating bandwidth from the source to the indicated destination, we can create a back-up path to another resource which could be closer to the destination in terms of back-up wavelengths.

We aim to evaluate the aforementioned relocation scheme against the traditional shared protection, from a network dimensioning perspective. The first thing we have to do is choose the K best server locations \( \Delta \) (hence \( |\Delta| = K \)), which is a K-medoid problem for which we refer to [2].

In contrast with most network dimensioning techniques, we do not start from a static demand matrix where for every connection source and destination are given. We rather start from a demand vector where each source specifies how many connections must be established between it and some resource in the network. This means that the scheduling of the primary resource is also included in our programs. The solution of the ILP will be a global optimum for the given demand vector.

Furthermore, we assume that every OXC in the network is capable of full wavelength conversion which we will refer to as the Virtual Wavelength Path (VWP) network. Thus, there will not be a so-called wavelength continuity constraint to be met.

III. ILP

A. Notation

Our network is modeled via the following variables.

- We have the graph \( G(V, E) \) with \( V \) the set of vertices and \( E \) the set of links.
- \( \phi \in \beta \) is a connection variable with source \( s \) and destination \( d \).
- \( |V| = N \) represents the number of network nodes, \( |E| = L \) is the number of links.
- \( \Delta \) is the set with the Grid resources (and hence a subset of \( V \)).
- \( P_{i,j}^\phi \) binary variable which is 1 if link \((i, j)\) is used for the primary path for connection \( \phi \).
- \( R_{i,j}^\phi \) binary variable which is 1 if link \((i, j)\) is used as part of a protection path for connection \( \phi \).
- \( m_j^\phi \) is 1 if node \( j \) is a resource which is used for connection \( \phi \).
- \( \pi_{i,j} \) the number of wavelengths on link \((i, j)\) used for a backup path.

B. Shared Path Protection

We start with the objective function which minimizes the number of primary- and backup wavelengths.

\[
\min \left( \sum_{i,j} \pi_{i,j} + \sum_{i,j} \sum_{\phi} P_{i,j}^\phi \right) \quad (1)
\]

Equation 2 incorporates the demand constraints and the flow conservations for the primary paths.

\[
\sum_{i:(i,j)\in E} P_{i,j}^\phi - \sum_{k:(j,k)\in E} P_{j,k}^\phi = \begin{cases} -1 : & j = s \\ m_j^\phi : & \text{else} \end{cases} \quad \forall \phi \in \beta, \forall j \in V \quad (2)
\]
For every connection there must be exactly one resource which is chosen as the destination and only Grid resource nodes ($\delta \in \Delta$) can be selected.

$$\sum_{\delta \in \Delta} m^\phi_\delta = 1, \forall \phi \in \beta$$  \hspace{1cm} (3)

$$m^\phi_\delta = 0, \forall \delta \notin \Delta,$$  \hspace{1cm} (4)

Equation 5 represents the flow conservations and the demand constraints for the back up paths. Because we are modeling the traditional protection scheme we have to use the same $m^\phi_j$ variable for the cases $j \neq s$ as in case of the primary paths. Indeed, in classical shared protection, the protection path ends at the same node as the primary path.

$$\sum_{i:(i,j) \in E} R^\phi_{(i,j)} - \sum_{p:(j,p) \in E} R^\phi_{(j,p)} = \begin{cases} -1 : j = s \\ m^\phi_j : \text{else} \end{cases}$$  \hspace{1cm} (5)

We continue with the constraint stating that a primary path and a back-up path protecting that primary path cannot overlap.

$$R^\phi_{(i,j)} + P^\phi_{(i,j)} \leq 1$$  \hspace{1cm} (6)

$$\forall \phi \in \beta, \forall (i,j) \in E$$

Now we have to introduce the variables which are used to count the number of wavelengths for secondary paths. Therefore, we define auxiliary binary variables $\Theta^\phi_{(i,j),(k,l)}$ which are 1 if a wavelength is used on link $(i,j)$ for protecting connection $\phi$ which uses a primary path crossing link $(k,l)$

$$\Theta^\phi_{(i,j),(k,l)} + 1 \geq R^\phi_{(i,j)} + P^\phi_{(k,l)}$$  \hspace{1cm} (7)

$$\forall \phi \in \beta, \forall (i,j),(k,l) \in E$$

Finally we have to bound the variables which express the total number of wavelengths used for back-up paths on a specified link.

$$\pi_{(i,j)} \geq \sum_{\phi} \Theta^\phi_{(i,j),(k,l)}$$  \hspace{1cm} (8)

$$\forall (i,j) \neq (k,l) \in E, \forall (k,l) \in E$$

C. Shared Path Protection with relocation

When we allow the backup path to end in a resource other the the primary resource, we achieve shared path protection with relocation. We keep every equation and only change the flow-demand constraints for the backup paths from Eq. (5) to the new (10), and add constrains (10) and (11). We also have to introduce variable $b^\phi_\delta$ which has the same purpose as $m^\phi_\delta$ but then for the relocation server.

$$\sum_{i:(i,j) \in E} R^\phi_{(i,j)} - \sum_{p:(j,p) \in E} R^\phi_{(j,p)} = \begin{cases} -1 : j = s \\ b^\phi_\delta : \text{else} \end{cases}$$  \hspace{1cm} (9)

$$\forall j \in V, \forall \phi \in \beta$$

$$\sum_{\delta \in \Delta} b^\phi_\delta = 1, \forall \phi \in \beta$$  \hspace{1cm} (10)

$$b^\phi_\delta = 0, \forall \delta \notin \Delta,$$  \hspace{1cm} (11)

IV. HEURISTIC

A. Heuristic overview

While the ILP formulation presented above allows to find the optimal solution, it is well-known not to be scalable to large problem instances (leading to high complexity in terms of memory utilisation and execution time). Hence, in order to evaluate the relocation strategy on a larger scale, we also propose a heuristic solution of the network dimensioning problem. To this end, we started from a heuristic [5] which tries to minimize the total resource usage by minimizing the resources for the primary connections as well by maximizing the sharing among the backup resources. We extended this heuristic to the Grid case (cf. anycast principle, choosing also job destination) and the relocation scheme. The heuristic also works in four stages:

1) For every connection $\phi$ try find a resource $\delta \in \Delta$ which minimizes its pair of link disjoint paths from the fixed source to that resource.

2) For every connection $\phi$, assign the best possible choice for the primary- and back up path ($P_\phi, B_\phi$) from the discovered link disjoint pair of paths.

3) For every connection $\phi$, maximize the sharing among the backup paths by rerouting its back up path $B_\phi$.

4) For every connection $\phi$, minimize the resources needed for its primary path $P_\phi$ by rerouting it.

In the next paragraphs we will explain each step in detail.

B. Step 1: compute two link disjoint paths

In this step we have to calculate the two initial, edge disjoint paths from the source node to one of the available resources in the network. A naive way of calculating these paths could be to find the shortest path and then finding the shortest path in the same graph, but with the edges of the first shortest path deleted from the graph. But as shown in [6] this technique can generate suboptimal solutions and can even fail to generate pairs of paths when such paths actually exist. Therefore the heuristic uses Suurballe’s algorithm[6] which will find a pair of edge-disjoint paths from vertex $s$ to vertex $d$ such that the total cost of the two paths is minimal among all such path pairs.

For the normal shared path protection heuristic we compute for every connection $\phi$ the set $\Psi_\phi$ of pairs of paths to all resources and we take the pair with the least amount of edges.

$$\forall \phi \in \beta : \Psi_\phi = \forall \delta \in \Delta : \{ P^1_{s \to \delta}, P^2_{s \to \delta} \}$$

$$\forall \phi \in \beta : \{ P^1_{\phi}, P^2_{\phi} \} = \min_{\delta \in \Delta} (P^1_{s \to \delta} + P^2_{s \to \delta})$$

This exhaustive search is feasible, since we assume a reasonably small set $\Delta$ of resource sites. This choice is motivated by
which shows that a small number of resource sites suffices and allows to minimize overall network load.

In the relocation case, an artificial transformation of the original network topology is required as shown in figure 2. First we introduce a single virtual resource which will serve as the single destination for the Suurballe algorithm. Secondly, the real resources are connected to the virtual resource by introducing two virtual links between each resource and the single virtual resource. If we now find a pair of edge disjoint paths from the source to the virtual resource, we can find the primary server and relocation server as the hop before the virtual resource on both paths. In the next paragraph, the destination for the normal protection strategy is the primary server found in this step while in the relocation case the destination denotes the virtual server.

C. Step 2: Choose the primary and backup path

In this stage of the heuristic we have found two link disjoint paths to one server in case for the normal protection (or two paths to two possible different servers for the relocation case). Now we have to decide for every connection which path to use as primary path and which one for the backup path. This choice is important because it fixes the search tree in which the solution must be found. We have opted to use a greedy attack and make the best possible choice at this point in the heuristic. So for every connection we try each permutation and we choose the one which minimizes the total number of wavelengths (II ) at that moment.

\[ \forall \phi : (P_\phi, B_\phi) = \begin{cases} (P_\phi^1, P_\phi^2) & \text{if this minimizes } \Pi \\ (P_\phi^2, P_\phi^1) & \text{if this minimizes } \Pi \end{cases} \]

D. Step 3: increase the sharing among backup paths

For every connection we try to optimize the sharing of backup wavelengths, as follows. Consider connection \( \phi = (P_\phi, B_\phi) \). We remove \( P_\phi \) from the network topology and we find the set of edges (\( \Gamma \)) from the backup paths which protect a primary path which is link disjoint with \( P_\phi \). We give all the edges from \( \Gamma \) weight 0 and all other edges which have not been recognized weight 1 after which we find the shortest path \( \hat{B}_\phi \) from the source to the destination. This \( \hat{B}_\phi \) is a valid backup path for connection \( \phi \) with a cost not larger than the cost of \( B_\phi \). Note that if \( \hat{B}_\phi \) contains an edge from \( \Gamma \), it means that we have increased the sharing ratio among the backup paths.

We calculate this alternate backup path for every connection, but we only reroute the backup path from that connection which minimizes the total number of wavelengths for the whole network. We then repeat this process (of changing one backup path leading to the greatest cost reduction) until it converges, i.e. the new backup path does not have a smaller cost compared with the previous one.

E. Step 4: decrease the number of primary resources

In this step we aim at reducing the amount of wavelengths for primary paths, similarly as before for backup wavelengths. In each iteration one primary path is changed. The new primary path for a connection \( \phi \) is calculated as follows. We remove the backup path \( B_\phi \) from connection \( \phi \) and find the set of all edges \( \Omega \) which are part of some primary path which is protected by a backup path which is not link disjoint with \( B_\phi \). We remove all these edges from the network while the remaining edges receive weight 1. We find the shortest path from source to destination which is a valid primary path for \( \phi \) with a cost not larger than the original primary path. Again, we calculate this potential cost reduction for all connections, but change only the one leading to maximal cost savings. We then repeat the calculation for the new setup, until no further savings are possible.

V. Complexity

The complexity of an ILP can be measured by the number of variables which need to be created and by the number of constraints which are being imposed on them. The shared protection scheme uses \( L + 2 \times N \times |\beta| + N \times |\beta| + L^2 \times |\beta| \) variables and the relocation ILP \( L + 2 \times L \times |\beta| + 2 \times N \times |\beta| + L^2 \times |\beta| \times |\Delta| \) variables.

For the heuristics however, the number of iterations is an important factor. The algorithm of finding the pair of link disjoint paths of minimal cost is an \( O(N^2 \log N) \) algorithm [8]. Choosing the appropriate path for the primary and backup paths is \( |\beta| \times 2 \) and the rerouting steps take in the worst case a predefined number of steps (worst case is when there is no convergence). Therefore, the whole algorithm is an \( O(N^2 \log N + |\beta| \times 2 + 2 \times \text{bound}) \) algorithm.

VI. Case Study

The network we have considered is one constructed by joint effort of the LION[9] and COST ACTION 266 projects which resulted in a pan-European network, as show in 3. For an arbitrary demand vector we have run the ILP (as described in [2]) to find the best possible server locations. These locations are:

- Dublin
- Paris
- Zurich
- Munich
- Berlin

To run all these jobs we have used the Hight Performance Computing (HPC) cluster provided by Ghent University. The HPC cluster consists out of 196 IBM HS 21 XM blades with
each blade containing a dual core Intel Xeon L5420 processor running at 2.5 GHz.

A. Validation of the Heuristic

For every demand size (the total number of requested connections) \( n \in [5 - 15] \) we have created 10 random demand vectors which serve as input for both the ILP and the heuristics. We have also run the protection ILPs described in [2]. As outlined before, the main difference between these ILPs and the current work is that the former assume a fully specified traffic matrix as given, i.e. the destination of Grid jobs is a priori given. In the paper currently at hand however, we optimize network dimensions by appropriately choosing the destination site and hence start from a demand vector specifying just the source. However, we used the ILPs from [2] to determine how good the heuristic achieves the minimal wavelength requirement for the destination choice made. Therefore, we performed the calculations as follows:

1) We start from a demand vector \( \alpha \) and run the heuristics to find a RWA solution for the total number of necessary wavelengths \( (N^\text{prot}_h, N^\text{reloc}_h) \).
2) The solutions from the heuristics also render two static demand matrices \( \sigma^\text{prot} \) and \( \sigma^\text{reloc} \).
3) We run the 'local optimum ILP' starting from \( \sigma^\text{prot} \) and \( \sigma^\text{reloc} \) and find its solution \( (N^\text{prot}_l, N^\text{reloc}_l) \).
4) We run the 'global optimum ILP' starting from \( \alpha \) and find its solution \( (N^\text{prot}_g, N^\text{reloc}_g) \).

The results for these numbers are shown in Fig. 4. We see that the heuristic approximates the global- and local ILP very well. On average, the total number of wavelengths for the shared path protection heuristic only differs 4.5% from the number of wavelengths for the global ILP and this is even less for the relocation heuristic (3, 7%). This gives us a fairly good indication of the effectiveness of our heuristic.

On the matter of shared path protection versus relocation, we refer to figure 5. In this graph we show the the rates \( \eta_h = \frac{N^\text{prot}_h}{N^\text{reloc}_h}, \eta_l = \frac{N^\text{prot}_l}{N^\text{reloc}_l}, \eta_g = \frac{N^\text{prot}_g}{N^\text{reloc}_g} \) which are indicators for the reduction in wavelengths caused by relocation. On average we see that the relocation strategy only needs about 85% of the total wavelengths needed when using the traditional protection strategy. This (average) observation is the same for the heuristic, the local ILP and the global ILP. Compared to the results we gathered in [2] we see that this is 5% less. This reason for this is that we have used another network and we believe the savings brought down by relocation is dependent on the network type.

B. Network with a larger demand

Due to the scalability problems of the ILPs, these programs start to have memory issues when running them for large problem instances. Because of this observation we are required to turn to the heuristics. We have worked the same way as before: for every demand size we have made 10 random demand vectors and ran the heuristics for which we show the average results per demand size. These results can be found...
in figure 6. We see that the trend which has been spotted in VI-A continues. The average ratio $\eta_h$ is 84% which means relocation imposes a reduction of 16% on the total number of wavelengths compared to the traditional protection scheme.

VII. CONCLUSION

In this paper we have described an alternative method for path protection against single link failures in an optical Grid scenario. We have exploited the grid specific anycast principle into a traditional protection scheme. This principle states that a Grid system is free in choosing an appropriate destination for a job to execute it (since indeed Grid users in general do not care where exactly the jobs they submit end up being executed). Therefore, in case of a network failure, we allow to relocate the job to another possible resource, in order to minimize the bandwidth which needs to be allocated for the backup path.

We have created two ILP programs which will compute the optimal solution for both RWA problems, starting from a so-called demand vector specifying only the connection sources (i.e. job origins). The job destinations are chosen such that network capacity is minimized. Due to scalability problems with ILP, we have also presented two accompanying heuristics that are practical for solving larger problem instances. The effectiveness of the heuristics has been validated through comparison with their corresponding ILP: results on small enough cases (to allow for ILP solution) showed that the heuristics approximated the ILP solutions very well (discrepancy of only a few percent).

Our study case pointed out that by exploiting relocation we can achieve a reduction of the total number of necessary wavelengths of about 15%, both for a small and a larger number of connections in the demand vector. The price paid for this network capacity reduction is that of the increased load on Grid resources at the sites where jobs are relocated to in case of failures. However, we note that the resulting increase in resource capacity needed to cater for this extra load would also need to be present to protect against Grid resource failures.

Material for future work includes careful investigation of that extra Grid resource load. We plan to consider a dimensioning study for both network and computational resources, using relocation as protection mechanism to protect also the Grid resources.

ACKNOWLEDGMENT

The work described in this paper was carried out with the support of the BONE-project (Building the Future Optical Network in Europe), a Network of Excellence funded by the European Commission through the 7th ICT-Framework Programme, as well as the IST Phosphorus-project. The Flemish government partly funded this work through the Research Foundation (FWO). C. Develder is a post-doctoral fellow of the FWO. Jens Buysse is supported by the research institute IWT.

REFERENCES