Les Cahiers du GERAD ISSN: 0711–2440

Joint optimisation of server location and network dimension of optical grids or clouds

B. Jaumard, A. Shaikh,
C. Develder

G–2014–37

May 2014

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Joint optimisation of server location and network dimension of optical grids or clouds

Brigitte Jaumard[a]
Ali Shaikh[a]
Chris Develder[b]

[a] *GERAD & Department of Computer Science and Software Engineering, Concordia University, Montrál (Québec) Canada H3G 1M8*

[b] *Department of Information Technology – IBCN, Ghent University – iMinds, Ghent, Belgium*

bjaumard@cse.concordia.ca
al_shaikh@encs.concordia.ca
chris.develder@intec.ugent.be

May 2014

Les Cahiers du GERAD
G–2014–37

Abstract: In grid or cloud computing, the optimal location and capacity of data centers (hosting the servers executing the remote users' tasks) depends on the available network resources, esp. link bandwidths providing connectivity to the candidate data center locations. In this study, we propose to jointly optimize the link dimensioning and the location of the servers in an optical grid or cloud, where we exploit the anycast principle for resiliency against either link or node failures. (The latter implies that we will allow to use a different server location under failure conditions.) While the problem has some resemblance with either the classical p-center or k-means location problems, the resilience requirement exploiting anycast significantly increases complexity, since we need link/node disjoint paths to possibly two different locations for each request.

We propose three different decomposition schemes, each resulting in a different model, to address scalability issues encountered in traditional optimization models. We extensively compare the models with computational experiments. The first model is quite efficient and scalable, under the assumption of several backup paths for a given source node, i.e., more complexity at the control network level. If we only allow a single backup path to be followed for all requests from a given source node, then the second model is the most efficient. Our third model is a direct extension of the model proposed for the p-median problem, i.e., the closest classical problem of operation research, but it is the least efficient and scalable one.

Key Words: Network dimensioning, optical grids, cloud computing, ILP, column generation, anycast routing.

Acknowledgments: The first author has been supported by a Concordia University Research Chair (Tier I) and by an NSERC (Natural Sciences and Engineering Research Council of Canada) grant.

1 Introduction

Internet technology has been the major driver for communication traffic growth in recent decades. For a while, this was dominated by peer-to-peer traffic, i.e., file sharing between end user devices residing at those users' machines. But since 2008, most Internet traffic had its end point in a data center [1]. This has been largely driven by the development of distributed computing paradigms. Three widely explored parallel and distributed type computing paradigms are: Cluster Computing, Grid Computing, and Cloud Computing. *Cluster computing* refers to a collection of co-located servers under the control of a single administration entity (e.g., within a company/university/organisation) that are jointly used to tackle demanding computational tasks (possibly under the form of interacting components distributed across multiple physical servers). The idea of *grid computing* was to interconnect geographically distributed computing and storage infrastructure, where different locations could be under the control of different organizations. The motivation is to cost-efficiently share typically expensive infrastructure among, e.g., researchers.

A similar motivation of infrastructure sharing underlies *cloud computing*. To this end, the principle of virtualization is applied: e.g., multiple so-called virtual machines (VMs) are created, offering isolation between services/applications running on each of them, but the VMs can run on the same physical hardware (thus being more cost efficient to allocate a dedicated physical server). The idea is that such a service of offering, e.g., VMs, could be highly interactive, as reflected in the NIST definition of cloud computing as *"a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management error or service provider interaction"* [2]. Also, note that typically cloud computing is more commercially oriented (as opposed to generally more research oriented grid infrastructure). Hence, the notion of service level agreements (SLAs) between the service provider and its customers comes into play [3].

In terms of traffic volume, it is expected that by 2016, global data center traffic could reach 6.6 zettabytes (1 ZB = 10^{21} bytes), and nearly two thirds thereof will be cloud traffic [1]. This growing demand of traffic requires a reliable and high-bandwidth communication medium. For the core network (which we focus on in this paper), using optical network technology is highly cost-efficient, cf. the application of Dense Wavelength Division Multiplexing (DWDM) technology, i.e., running multiple wavelength carriers simultaneously over the same physical fiber to provide large bandwidth. Given the continually rising bandwidth demands, today's solutions can run 100Gb/s per wavelength (40-80 wavelengths on each fiber pair using DWDM). Currently, flexible grid networks are being considered: the flexible grid refers to adaptive transceivers and intelligent nodes, allowing service providers to increase the bandwidth without overhauling it [4]. This new paradigm is called Elastic Optical Networking (EON).

Applying such optical technologies for supporting modern distributed data center applications gave rise to the concept of optical grids and clouds [5]. Those optical grid/cloud networks target distributed applications, such as scientific applications (scientific computing, data-centric computing), business applications (transactional systems, multimedia, data mining), and consumer applications (personal content, gaming, interactive TV) [5]. A key concept in such applications, that distinguishes them from classical ones (such as voice communications), is the *anycast routing* principle. This boils down to the fact that users are unaware of, and do not care about, the exact location where their services are being offered. This implies a degree of freedom in deciding at which data center location to run a particular service. In technical terms, the latter amounts to the rise of a traffic vector (only specifying the data sources and the traffic volume they request), rather than the classical (source, destination)-based traffic matrix.

In this paper, we address the joint dimensioning of such optical grid/cloud networks and location optimization of the data centers: how to determine the resources necessary to meet a given traffic volume and where are the best data center locations. Dimensioning optical grid/cloud computing differs from dimensioning classical optical networks due to the aforementioned anycast routing principle – which complicates the dimensioning problem – and the fact that we are not just concerned with network infrastructure, but also with the location optimization of data centers (i.e., servers). Indeed, joint dimensioning of an optical grid/cloud network and data center location optimization require to determine (and typically minimize the total cost of): (i) the number of wavelengths on each link, (ii) the location of servers (data centers), and (iii) the number of CPUs (Central Processing Units) at each server location. In making this decision, we need to specify for all requests (with given source) at which data center it will be served.

In this paper, we address the joint optimization of steps (i) and (ii), in an uncapacitated network, i.e., we do not have a priori any limits on the maximal capacity per network link (nor per data center in terms of server capacity). The failures that we will consider are network links and/or nodes, and/or servers (see further). We note that including server capacity calculations for step (iii) is fairly straightforward, once steps (i)–(ii) have been taken (e.g., see [6] for an example on how to incorporate server capacity calculation in a unified network and data center model).

Compared to our previous work in this area, the *contributions* of this paper are the following:

- Opposed to [7, 8], which considers a routing and wavelength assignment problem for a capacitated network (with given locations of servers), i.e., maximize the requests that can be provisioned within a network of given size, we currently study the dimensioning problem of finding the minimal resources to satisfy all requests.
- In [9, 10] we addressed the issue of minimal network cost dimensioning when the location of the data centers is fixed, and given a priori (step (i)). In the current paper, we jointly optimize the dimensioning while making an optimal choice of the data center locations among a given set of candidates (i.e., combining steps (i)–(ii)).
- In previous work [6], we provided an algorithm to decide on both the locations of servers, as well as the dimensions of network (and server) resources, but we split the decision algorithm in two discrete, sequential steps. Where in [6] we had a separate step deciding on the server locations (making some simplifying assumptions with respect to the routing of requests), we currently integrate that location choice with the network sizing within a single, highly scalable, column generation algorithm.

This paper is an extension of the short conference paper [11], where we for the first time solved the combined server location and network dimensioning problem (i.e., steps (i)–(ii)). Beyond that, the current paper introduces:

- An entirely new Model III;
- Complete mathematical formulations for all three developed column generation models, and how to solve them;
- An elaborate comparison of the results of the three models on a European network topology, including optimality gaps, running time, and study of the routing characteristics of the found solutions.

The rest of the paper is organized as follows. In the next section, related work is reviewed. Three different mathematical models are presented in the next three sections. *Model I* (Section 4) is based on configurations associated with a single source node, and comprising a pair of one working and one protection path. *Model II* (Section 5) has configurations, still associated with a single request source node, but now composed of a set of multiple working and one protection paths. and *Model III* (Section 6) centers configurations of multiple paths (both working and backup) ending in a given data center location (expanding the previous p-center/p-median models relying on column generation formulations, see, e.g., [12]). Experimental and comparative results are described in Section 8. Therein, we first compare the quality of the solutions of the three models, and then investigate different scenarios for the layout of an optical grid in terms of the number of backup paths and of the number of servers. Conclusions and future work are outlined in Section 9.

2 Related Work

As one of the first works to study the network dimensioning problem in a grid computing setting (but only in failure-free conditions), Develder et al. [13] propose a four-step scheme that sequentially determines the location of servers, the number of CPUs, the resulting (source, destination)-based traffic and finally link capacity (using ILP, numeric calculations and simulations). Subsequent work [14] introduced the idea of exploiting relocation to cater for possible failures, and presented an initial phased approach to resiliently dimension the optical grid network. This was further elaborated by jointly dimensioning both the network and server capacities (but still with a separate step to first decide on the data center locations) in [15, 6]. Other works on anycast routing in optical networks, with a priori given data center locations, include [16, 17] (without resilience) and [18] (with resilience). (For a more elaborated overview of related works specifically on anycast routing in optical networks, both in an online setting and the offline dimensioning context studied here, we refer to [6, Section II].)

Finding the best server locations in grid/cloud networks has some resemblance with some classical facility location problems, see, e.g., [19]. Indeed, there are three classical facility location problems: median, covering, and center (also referred as p-median, set covering, and p-center) location problems, see, e.g., [20, 21]. The p-median (also called minisum) problem finds the location of p facilities with the objective of minimizing the overall distance between the facilities and demands, i.e., the sum of all distances between each (demand, facility) pair. This objective is used when demands have no hard constraint on the distance (or time to cover it) they may cover to reach a facility: e.g., network servers, warehouses, traveling. For some facility location problems, there is a hard time and/or constraint: e.g., in case of fire stations, ambulance centers the critical demands need to be served within a specified time. For such problems, the objective is then to minimize the total number of facilities (set covering) or to minimize the maximum distance (p-center or minimax). For a recent review, see, e.g., [19].

Larumbe and Sansò [22] investigated the optimal location of data centers, with the objective of minimizing the average network delay (convex objective), without taking into account any reliability concern or link dimensioning. Chakareski [23] also studied the location of data centers, with the aim of minimizing the overall operating cost of the network (again a convex objective), in a context of multi-service networks. They take the link and resource capacity into account, but do not worry about reliability.

However, survivability is a major issue in the area of data communications, where in case of link failure on the working path, the traffic is routed on the backup path. As indicated before, the same goes in grid/cloud environments, but exploiting the anycast routing principle, the primary and backup paths can end at different facilities – a feature that complicates the dimensioning of optical grid/cloud networks. Finding the best locations for server facilities (i.e., data centers) in such a setting is related to the p-median problem, for which a literature summary of solution methods is presented in [24].

While resiliency is now often studied in the context of location theory (e.g., [25, 26, 27],) these studies' assumptions vary and the resilient location models cannot be re-used as such in our context of optical cloud networks. For instance, studies in location theory have primarily focused on the identification of the most vulnerable and valuable system components as a tool to drive decisions where to direct funds for reinforcing systems. More recently, stochastic variants of network flow interdiction problems have appeared, and differ from the assumptions of resilient cloud grids.

3 Problem Statement

We focus on the core network of an optical grid/cloud, described as a directed graph $G = (V, L)$ where V is the node set (indexed by v) and L is the link set (indexed by ℓ). We assume that every pair of connected nodes has two directed links, one in each direction. We denote by $\omega^+(v)$ and $\omega^-(v)$ the set of outgoing and incoming links of v, respectively.

Traffic is defined by the number of requests, denoted by D_v, each originating from a source node $v \in V_{\text{D}} \subseteq V$, where V_{D} designates the subset of nodes of V with requests. Let $S \subseteq V$ denote the set of server nodes with $n_{\text{S}} = |S|$, i.e., the number of server nodes.

We are interested in resilient optical grids, and protect against link or node failures in the optical grid/cloud (i.e., the core network), as well as single server failures (i.e., data centers). More precisely, we will investigate three failure scenarios:

Failure Scenario 1. Protection against any single link failure.

Failure Scenario 2. Scenario 1, with the additional protection against any single (optical) node failure.

Failure Scenario 3. Scenario 2, with the additional protection against failure of a single server site (i.e., a data center).

4 CG-ILP Model I: Path Pair Based Configurations

We propose a decomposition model, based on so-called configurations, which we will solve using column generation. Our first model relies an extension of the model proposed in [10], with the addition of the selection of the best locations for the servers. The decomposition relies on a set of configurations, where each configuration is associated with a node v, and made of a pair of one working and one backup path, both originated at v, except for the nodes which are hosting a data center, where we assume that the traffic will be served locally and thus require no path. An example is sketched in Figure 1, where demands originating at source node v_1 can be accommodated by configurations c_2 and c_3 (either only one of them, or distributed over the two of them). Similarly, demands originating at node v_2 can be accommodated by configurations c_1 or c_4. Working paths are drawn with simple lines, while backup paths are associated with dotted lines.

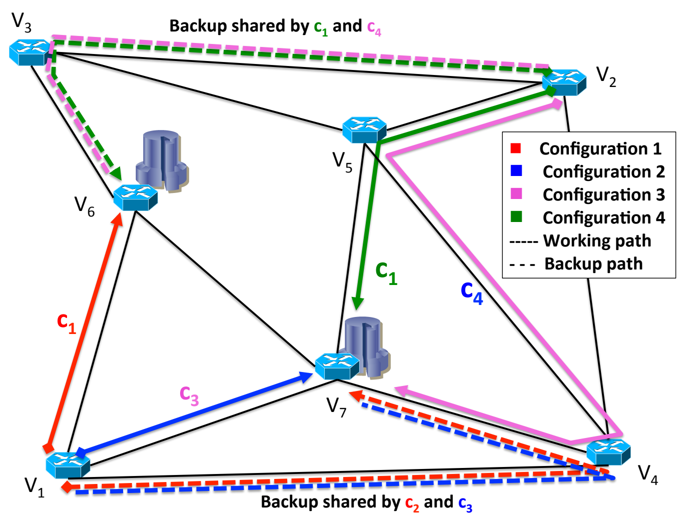


Figure 1: Several configuration examples for the single link failure scenario

The optimization model, called master problem, will assume these configurations to be given (or generated iteratively, see below, Section 4.2), and thus has to decide which ones to select (and how many times) to fulfil a given demand. We now define the configurations and the given parameters of Model I more formally:

- C_v Set of configurations associated with source node v.
- C Set of all configurations with $C = \bigcup_{v \in V_{\text{D}}} C_v$.
- c Generic index of a configuration. It is defined for a given source node $v \in V_{\text{D}}$ and comprises a pair of one working and one backup path, both originating from v, and each destined for a server $v_s \in S$ (not necessarily the same server node for the two paths).
- $p_\ell^{\text{W},c}$ $= 1$ if the working path goes through link ℓ in configuration c, 0 otherwise.
- $p_v^{\text{W},c}$ $= 1$ if the working path goes through node v in configuration c, 0 otherwise.
- $p_\ell^{\text{B},c}$ $= 1$ if link ℓ is used by the backup path of c, 0 otherwise.
- a_v^c $= 1$ if node v is selected as a server location either by the working or the backup path in configuration c, 0 otherwise.

Note that we build configurations for provisioning job requests originating from any v, which in a particular configuration can be served at any of the possible server locations. Yet, we will retain and select only the ones associated with the selected server locations.

Next, we define the set of variables:

$z_c \in \mathbb{Z}^+$ equals the number of times we select configuration c.
$x^{\text{B}}_\ell \in \mathbb{Z}^+$ equals the number of shared backup bandwidth units on link ℓ.
$y_v \in \{0,1\}$ equals 1 if we set a server at node v, 0 otherwise.

4.1 Master Problem

We first present the master problem for the first failure scenario and then explain how to modify it for the second and third failure scenarios.

4.1.1 Single Link Failure Scenario

The objective function minimizes the total network capacity, with the first term representing shared backup capacity, and the second the capacity of working paths:

$$\min \sum_{\ell \in L} \left(x^{\text{B}}_\ell + \sum_{c \in C} p^{\text{W},c}_\ell \, z_c \right), \tag{1}$$

subject to:

$$D_v \, y_v \; + \sum_{c \in C_v} z_c \geq D_v \qquad v \in V_{\text{D}} \tag{2}$$

$$\sum_{c \in C} p^{\text{W},c}_\ell \, p^{\text{B},c}_{\ell'} \, z_c \leq x^{\text{B}}_{\ell'} \qquad \ell, \ell' \in L : \ell' \notin \{\ell, \text{OPP}(\ell)\} \tag{3}$$

$$\sum_{c \in C} a^c_v \, z_c \leq M \, y_v \qquad v \in V \tag{4}$$

$$\sum_{v \in V} y_v \leq n_{\text{S}} \tag{5}$$

$$z_c \in \mathbb{Z}^+ \qquad c \in C \tag{6}$$

$$x^{\text{B}}_\ell \in \mathbb{Z}^+ \qquad \ell \in L \tag{7}$$

$$y_v \in \{0,1\} \qquad v \in V. \tag{8}$$

Constraints (2) are the demand constraints, which ensure that all requests are assigned to and granted on a server. If node v hosts a server, we assume that all the job requests originating from v are readily served by v, unless the server is offline due to some failures, a case which is not considered in a single link failure scenario. Constraints (3) are used to compute the bandwidth requirement for link ℓ' in a backup path. Indeed, if ℓ' protects link ℓ, with ℓ belonging to several working paths (modelled here throughout the various configurations c associated with working paths containing ℓ), we must ensure ℓ' to have a large enough transport capacity. Constraints (3) are valid under the assumption that each configuration contains a single pair of working and protection paths, which are link or node disjoint, depending on the failure scenario under consideration. Constraints (4) ensure consistency among the variables y_v and z_c and the parameter a^c_v: we need to install servers ($y_v = 1$) if a configuration c is selected ($z_c > 0$) which has its endpoint in v ($a^c_v = 1$). The maximum number of data center loactions is controlled in (5). Constraints (6) to (8) define the domains of the variables.

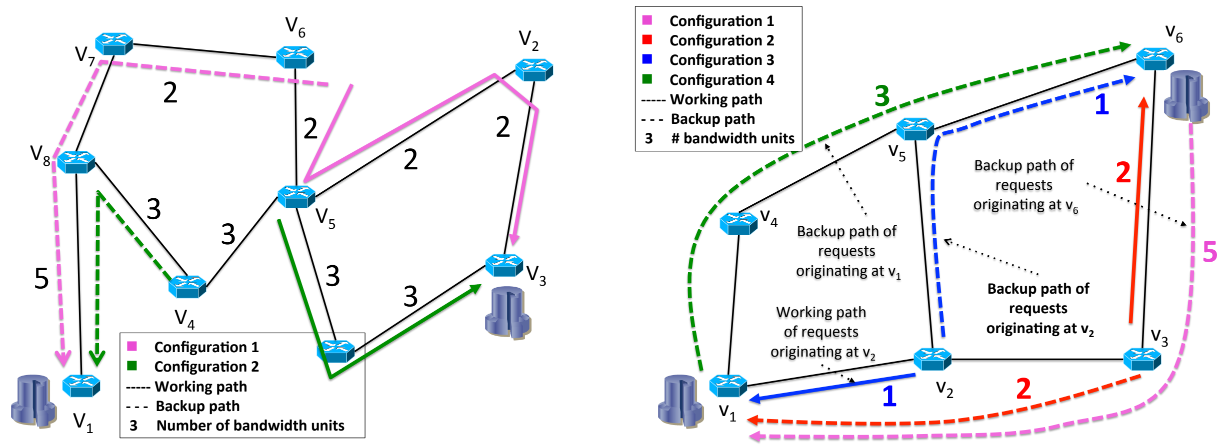
4.1.2 Single Link or Node Failure Scenario

The above model needs to be slightly modified in order to handle the second failure scenario, which adds the single node failures to the first scenario.

The objective function remains unchanged, except for the constraints taking into account the bandwidth requirements for the backup paths associated with working paths that are not node disjoint:

$$\sum_{c \in C} p^{\text{W},c}_v \, p^{\text{B},c}_\ell \, z_c \leq x^{\text{B}}_{\ell'} \qquad \ell \in L, v \in V. \tag{9}$$

An example of such a case is illustrated in Figure 2(a) with node v_5 belonging to two different working paths, while the two backup paths share two links. Thus, the bandwidth requirement on link $\ell = (v_8, v_9)$ must account for both affected paths under possible failure of node v_5.



(a) Single node protection (b) Single server protection

Figure 2: Model I: Configurations for single node or single server failure scenarios

4.1.3 Single Link or Node or Server Failure Scenario

When adding protection against server failure, a backup path is now needed for the requests originating at a node where a server is hosted. First, we need to modify the demand constraints (2), as we now must enforce the selection of a configuration with a backup path for the requests originating at a node hosting a server location:

$$\sum_{c \in C} z_c \geq D_v. \tag{2'}$$

Next, in the computation of the amount of required protection bandwidth, we need to include the bandwidth amount of the backup paths associated with requests originating or ending at a node hosting a server:

$$\sum_{c \in C} a_v^{\mathrm{W},c}\, p_\ell^{\mathrm{B},c}\, z_c \leq x_\ell^{\mathrm{B}} \qquad v \in V_{\mathrm{D}}, \ell \in L. \tag{10}$$

Such cases are illustrated in Figure 2(b). There, if the server located in v_1 fails, we need to reroute the requests originating at v_1, but also the requests which have been directed to the server located at v_1, i.e., the requests originating from v_2, so that the protection bandwidth requirement on (v_5, v_6) amounts to 4 units. Similarly, if the server located in v_6 fails, the protection bandwidth requirement on link (v_2, v_1) is 7 units.

4.2 Pricing Problem

In this section, we will describe the pricing problems associated with the formulations corresponding to each failure scenario. Indeed, we cannot solve the complete three master problems as described in Section 4.1 using all possible configurations, as they contain too many variables. One needs to recall that, in practice, the key idea of column generation techniques is to work only with a very small but meaningful subset of variables (or columns) of the master problem, forming the so-called restricted master problem (RMP). The role of the pricing problem is then to iteratively generate the most promising configuration(s) (variable(s)) in order to improve the optimal solution solution of the current RMP linear relaxation, until the optimality condition is fulfilled (i.e., the pricing problem does not contain any solution with a negative reduced cost).

Whatever the failure scenario, each pricing problem (PP for short) corresponds to the generation of one configuration for one source node, denoted by v_{s}, and outputs one backup path, and one working path if the source node does not host a server. Hence, in order to alleviate the notations, we will omit the c index, in the sections describing the pricing problems. We will denote by $\mathrm{PP}(v_\mathrm{s})$ each pricing problem associated with source node v_s.

4.2.1 Single Link Failures

While p_ℓ^{W} and p_ℓ^{B} designated parameters in the master problem, they now denote variables in the pricing problem, and are used to find a pair of working and backup paths. Similarly, $a_v = 1$ if a server (either working or backup) is located at node v, 0 otherwise. We next introduce two sets of variables: a_v^{W} and a_v^{B}, with $a_v = \max\{a_v^{\mathrm{W}}, a_v^{\mathrm{B}}\}$ such that $a_v^{\mathrm{W}} = 1$ (resp. a_v^{B}) if v is the location of the destination server of the working (resp. backup) path under construction in the configuration associated with the pricing problem.

The objective function of $\mathrm{PP}(v_s)$, i.e., the reduced cost of variable z_c, for the search of a configuration c associated with source node v_s, can be written as follows:

$$\overline{\mathrm{COST}}_1^{\,\mathrm{I}} = \sum_{\ell \in L} p_\ell^{\mathrm{W}} - u_{v_s}^{(2)} + \sum_{v \in V} u_{\ell\ell'}^{(4)} a_v + \sum_{\ell \in L} \sum_{\ell' \in L : \ell' \notin \{\ell, \mathrm{OPP}(\ell)\}} u_v^{(3)} \, p_\ell^{\mathrm{W},c} \, p_{\ell'}^{\mathrm{B},c} \tag{11}$$

where $u_v^{(2)} \geq 0$, $u_v^{(3)} \geq 0$ and $u_{\ell\ell'}^{(4)} \geq 0$ are the values of the dual variables associated with constraints (2), (3), and (4) respectively. Note that the dual values are indexed with the master problem's constraint numbers they are associated with.

The set of constraints of $\mathrm{PP}(v_s)$ can be written as follows:

$$\sum_{\ell \in \omega^+(v)} p_\ell^{\mathrm{W}} - \sum_{\ell \in \omega^-(v)} p_\ell^{\mathrm{W}} = \begin{cases} 1 - a_v^{\mathrm{W}} & \text{if } v = v_s \\ -\, a_v^{\mathrm{W}} & \text{otherwise} \end{cases} \qquad v \in V \tag{12}$$

$$\sum_{\ell \in \omega^+(v)} p_\ell^{\mathrm{B}} - \sum_{\ell \in \omega^-(v)} p_\ell^{\mathrm{B}} = \begin{cases} 1 - a_v^{\mathrm{B}} & \text{if } v = v_s \\ -\, a_v^{\mathrm{B}} & \text{otherwise} \end{cases} \qquad v \in V \tag{13}$$

$$p_\ell^{\mathrm{W}} + p_{\ell'}^{\mathrm{W}} + p_\ell^{\mathrm{B}} + p_{\ell'}^{\mathrm{B}} \leq 1 \qquad \ell, \ell' \in L, \ell \neq \ell', \ \ell \text{ and } \ell' \text{ are each other's opposite} \tag{14}$$

$$a_v \geq a_v^{\mathrm{W}} \qquad v \in V \tag{15}$$

$$a_v \geq a_v^{\mathrm{B}} \qquad v \in V \tag{16}$$

$$\sum_{v \in V} a_v^{\mathrm{W}} \leq 1 \tag{17}$$

$$\sum_{v \in V} a_v^{\mathrm{B}} \leq 1 \tag{18}$$

$$p_\ell^{\mathrm{W}}, p_\ell^{\mathrm{B}} \in \{0,1\} \qquad \ell \in L \tag{19}$$

$$a_v, a_v^{\mathrm{W}}, a_v^{\mathrm{B}} \in \{0,1\} \qquad v \in V. \tag{20}$$

Constraints (12) and (13) are flow conservation constraints in order to establish working and backup paths. When we solve a given pricing problem for a given source node v_s, we do not know yet whether v_s will host a server. In case it does, i.e., $a_{v_s}^{\mathrm{W}} = 1$, then we do not need to generate a working path, hence the term $1 - a_{v_s}^{\mathrm{W}}$ in the first part of the case in constraints (12). Similarly for the backup path, if v_s hosts a server, there is no need of generating a backup path in a single link failure scenario, hence the term $1 - a_{v_s}^{\mathrm{B}}$ in constraints (13). Note that, if $a_{v_s}^{\mathrm{B}} = 0$, a better solution can be found by setting $a_{v_s}^{\mathrm{B}} = 1$ (due to the expression of the reduced cost objective), hence a contradiction, therefore $a_{v_s}^{\mathrm{B}} = 1$ if $a_{v_s}^{\mathrm{W}} = 1$, and due to constraint (13), there is no backup path either, which is fine in the context of single link failure.

The disjointness of working and backup paths is ensured by constraints (14). The relationship $a_v = \max\{a_v^{\mathrm{W}}, a_v^{\mathrm{B}}\}$ is guaranteed by the combination of constraints (15) and (16) and the minimization of the third term of objective expression, see (11). Constraints (17) and (18) ensure that each path, working or backup, has exactly one destination server. The equality will be enforced through the flow constraints (12)–(13):

we cannot satisfy them without picking at least one server location. Constraints (19) and (20) define the domains of the variables.

In order to linearize the quadratic term of the reduced cost objective, i.e., $p_\ell^{\text{W}}\, p_{\ell'}^{\text{B}}$, we introduce the set of variables $p_{\ell\ell'}^{\text{W,B}} = p_\ell^{\text{W}}\, p_{\ell'}^{\text{B}}$ and the following set of constraints:
For all $\ell, \ell' \in L : \ell \neq \ell'$,

$$p_{\ell\ell'}^{\text{W,B}} \geq p_\ell^{\text{W}} + p_{\ell'}^{\text{B}} - 1 \tag{21}$$

$$p_\ell^{\text{W}} \geq p_{\ell\ell'}^{\text{W,B}} \tag{22}$$

$$p_{\ell'}^{\text{B}} \geq p_{\ell\ell'}^{\text{W,B}}. \tag{23}$$

Note that the last two sets of constraints (22) and (23) are actually redundant, taking into account the minimization objective (i.e., the reduced cost) of the pricing problem.

4.2.2 Single Link or Node Failures

If we want to protect also against single network node (i.e., OXC) failures, the expression of the reduced cost is modified due to the addition of the set of constraints (9) in the master problem:

$$\overline{\text{COST}}_2^{\text{I}} = \overline{\text{COST}}_1^{\text{I}} + \sum_{\ell \in L} \sum_{v \in V} u_{\ell v}^{(9)} p_v^{\text{W}} p_\ell^{\text{B}} \tag{24}$$

where $u^{(9)} = \left(u_{\ell v}^{(9)}\right)$ is the dual value vector associated with constraints (9). We can linearize the quadratic terms $p_v^{\text{W}} p_{\ell'}^{\text{B}}$ in (24), in the same way as the quadratic terms $p_\ell^{\text{W},c}\, p_{\ell'}^{\text{B},c}$ from (11).

Furthermore, the PP constraints (14) need to be replaced by constraints (25) and constraints (26) must be added:

$$\sum_{\ell \in \omega^-(v)} p_\ell^{\text{W}} + \sum_{\ell \in \omega^-(v)} p_\ell^{\text{B}} \leq 1 \qquad v \neq v_s, v \in V \tag{25}$$

$$a_v^{\text{W}} + a_v^{\text{B}} \leq 1 \qquad v \neq v_s, v \in V. \tag{26}$$

Constraints (25) ensure that, for each node, there is at most one incoming link. Note that, except for the source, constraints (25) also ensure protection against a single server failure, except if there is a server located at the source node. This last case will be dealt with in the next section (Section 4.2.3). Constraints (26) ensure that different node locations are selected for the working and the backup servers.

4.2.3 Single Link or Node or Server Failures

To also protect against data center failures, the expression of the reduced cost is modified due to the addition of master problem constraints (10):

$$\overline{\text{COST}}_3^{\text{I}} = \overline{\text{COST}}_2^{\text{I}} + \sum_{\ell \in L} \sum_{v \in V_{\text{D}}} u_{\ell v}^{(10)} a_v^{\text{W}} p_\ell^{\text{B}} \tag{27}$$

where $u^{(10)} = \left(u_{\ell v}^{(10)}\right)$, with $u_{\ell v}^{(10)} \geq 0$, is the value of the dual vector associated with constraints (10).

In addition to constraints (25) and (26), constraints (13) needs to be modified as follows:

$$\sum_{\ell \in \omega^+(v)} p_\ell^{\text{B}} - \sum_{\ell \in \omega^-(v)} p_\ell^{\text{B}} = \begin{cases} 1 & \text{if } v = v_s \\ -\, a_v^{\text{B}} & \text{if otherwise} \end{cases} \qquad v \in V. \tag{13'}$$

Indeed, in the case of a data center failure at v, it is mandatory to have a backup path towards an alternate destination for the requests originating at that data center v.

5 CG-ILP Model II: Source Based Configurations

In Model II, each configuration is again related with a single source node, but now the entire source node demand is provisioned using exactly one configuration. Each configuration may select one or more working paths and a single backup path. The idea is to reduce the number of configurations to be generated, and to define a decomposition where the difficulties of the master and pricing problems are more balanced. Examples of configurations for Model II are depicted in Figure 3. The impact of allowing a single backup path will be discussed in Section 8.

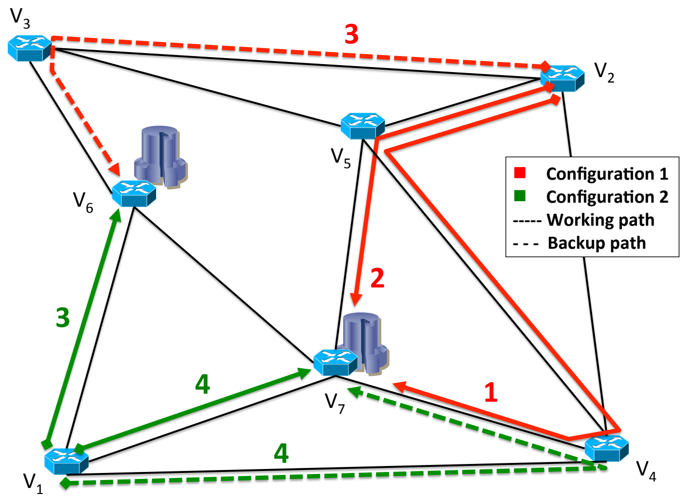


Figure 3: Configuration examples – Model II

Next, we define the overall set of variables that are needed for the three failure scenarios. (Note that not all variables will appear in all models.) The variables (x^{B}_{ℓ}) and parameters (a^{c}_{v}) have the same definition as in the Model I of Section 4. The new parameters are:

$\delta^{c}_{\ell\ell'}$ $\in \mathbb{Z}^{+}$ denotes the number of required backup units in order for link ℓ to be protected by ℓ' in configuration c, 0 otherwise.

$\varphi^{\text{W},c}_{\ell}$ $\in \mathbb{Z}^{+}$ is equal to the number of primary working units on link ℓ.

While the z_c are now binary decision variables:

z_c $= 1$ is configuration c is selected in the optimal solution, 0 otherwise.

There are also two more sets of variables to compute the backup bandwidth requirements:

$x^{v_s}_{\ell\ell'}$ is the bandwidth required on ℓ' in order to protect against a failing ℓ with respect to demand originating from $v_s \in V_{\text{D}}$.

$x^{v_s}_{v\ell'}$ is the bandwidth required on ℓ' in order to protect against a failure of $v \in V$ with respect to demand originating from $v_s \in V_{\text{D}}$ with $v \neq v_s$.

5.1 Master Problem

We first present the master problem for the first failure scenario and then explain how to modify it for the second and third failure scenarios.

5.1.1 Single Link Failure Scenario

The objective is the same as in Model I, but its mathematical expression differs. It is written as follows:

$$\min \quad \sum_{\ell \in L} \left(x_\ell^{\text{B}} + \sum_{c \in C} \varphi_\ell^{\text{W},c} \, z_c \right) \tag{28}$$

subject to:

$$y_v + \sum_{c \in C_v} z_c = 1 \qquad v \in V_{\text{D}} \tag{29}$$

$$\sum_{c \in C_v} \varphi_\ell^{\text{W},c} \, p_{\ell'}^{\text{B},c} \, z_c \leq x_{\ell\ell'}^v \qquad v \in V_{\text{D}}, \; \ell, \ell' \in L : \ell \notin \{\ell', \textsc{opp}(\ell')\} \tag{30}$$

$$\sum_{v \in V_{\text{D}}} x_{\ell\ell'}^v \leq x_{\ell'}^{\text{B}} \qquad \ell, \ell' \in L : \ell \notin \{\ell', \textsc{opp}(\ell')\} \tag{31}$$

$$\sum_{c \in C} a_v^c \, z_c \leq M \, y_v \qquad v \in V \tag{32}$$

$$\sum_{v \in V} y_v \leq n_{\text{S}} \tag{33}$$

$$z_c \in \{0,1\} \qquad c \in C \tag{34}$$

$$x_\ell^{\text{B}} \in \mathbb{Z}^+ \qquad \ell \in L \tag{35}$$

$$x_{\ell\ell'}^v \in \mathbb{Z}^+ \qquad v \in V, \; \ell, \ell' \in L : \ell \notin \{\ell', \textsc{opp}(\ell')\} \tag{36}$$

$$y_v \in \{0,1\} \qquad v \in V. \tag{37}$$

Constraints (29) ensure that a single configuration is selected for each source node (we consider maximal configurations only). Note that if a source node is selected as one of the server nodes, then there is no need to accept any configuration regarding the source node: this is ensured by adding first term, i.e., y_v. The backup bandwidth sharing is controlled in constraints (30) and (31). The last two sets of constraints, (32) and (33), define the selection of the server nodes and of the limit on the server number respectively. They are identical to constraints (4) and (5). Constraints (34) to (37) define the domains of the variables.

5.1.2 Single Link or Node Failure Scenario

Constraints (30) and (31) need to be replaced by the following two sets of constraints, as node failure protection implies link protection:

$$\sum_{c \in C} \varphi_v^{\text{W},c} p_{\ell'}^{B,c} z_c \leq x_{v\ell'}^{v_s} \qquad \ell' \in L; v \in V, v_s \in V_{\text{D}} : v \neq v_s \tag{38}$$

$$\sum_{v_s \in V_{\text{D}} : v \neq v_s} x_{v\ell'}^{v_s} \leq x_{\ell'}^{\text{B}} \qquad \ell' \in L, v \in V \tag{39}$$

$$x_{v\ell'}^{v_s} \in \mathbb{Z}^+ \qquad v \in V; \ell \in L; v_s \in V_{\text{D}}. \tag{40}$$

where

$$\varphi_v^{\text{W},c} = \sum_{\ell \in \omega^+(v)} \varphi_\ell^{\text{W},c}.$$

5.1.3 Single Link or Node or Server Failure Scenario

Constraints (29) need to be changed to:

$$\sum_{c \in C} z_c = 1 \qquad v \in V_{\text{D}} \tag{29$'$}$$

In addition to constraints (38) and (39), we need to add the following constraints to take care of the single server failures:

$$\underbrace{\sum_{c \in C} \varphi_v^{\text{W},c} \, a_{v_s}^{\text{W},c} \, p_{\ell'}^{\text{B},c} \, z_c}_{\substack{\text{Requests ending} \\ \text{at a location with a failing server}}} + \underbrace{\sum_{c \in C} D_v \, a_{v_s}^{\text{W},c} \, p_{\ell'}^{\text{B},c} \, z_c}_{\substack{\text{Requests originating} \\ \text{at a location with a failing server}}} \leq x_{\ell' v}^{v_s} \qquad \ell' \in L, v_s \in V_{\text{D}}. \tag{41}$$

where

$$\varphi_v^{\text{W},c} = \sum_{\ell \in \omega^-(v)} \varphi_\ell^{\text{W},\ell}.$$

5.2 Pricing Problem

Recall that each pricing problem corresponds to a single source, and that a configuration may comprise multiple working paths but just a single backup path.

5.2.1 Single Link Failures

The reduced cost expression is similar to the previous model, except for the change of φ_ℓ^{W} in the second term:

$$\overline{\text{COST}}_1^{\text{II}} = \sum_{\ell \in L} \varphi_\ell^{\text{W}} - u_{v_s}^{(29)} + \sum_{\ell \in L} \sum_{\ell' \in L : \ell \neq \ell'} u_{\ell\ell'}^{(30)} \, p_{\ell'}^{\text{B}} \varphi_\ell^{\text{W}} + \sum_{v \in V} u_v^{(32)} \, a_v \tag{42}$$

where $u^{(29)} \gtrless 0$, $u_{\ell\ell'}^{(30)} \geq 0$ and $u_v^{(32)} \geq 0$ are the values of the dual variables associated with the master problem constraints (29), (30) and (32), respectively.

Next we describe the set of pricing problem constraints:

$$\sum_{\ell \in \omega^+(v)} \varphi_\ell^{\text{W}} - \sum_{\ell \in \omega^-(v)} \varphi_\ell^{\text{W}} = \begin{cases} D_v(1 - a_{v_s}^{\text{W}}) & \text{if } v = v_s \\ -d_v & \text{otherwise} \end{cases} \qquad v \in V \tag{43}$$

$$\sum_{\ell \in \omega^+(v)} p_\ell^{\text{B}} - \sum_{\ell \in \omega^-(v)} p_\ell^{\text{B}} = \begin{cases} 1 - a_v^{\text{B}} & \text{if } v = v_s \\ -a_v^{\text{B}} & \text{otherwise} \end{cases} \qquad v \in V \tag{44}$$

$$\varphi_\ell^{W} \leq M \, p_\ell^{\text{W}} \qquad \ell \in L \tag{45}$$

$$p_\ell^{\text{W}} + p_{\ell'}^{\text{W}} + p_\ell^{\text{B}} + p_{\ell'}^{\text{B}} \leq 1 \qquad \ell, \ell' \in L : \ell = \text{OPP}(\ell') \tag{46}$$

$$a_v^{\text{B}} \leq a_v \qquad v \in V \tag{47}$$

$$d_v \leq M a_v^{\text{W}} \qquad v \in V \tag{48}$$

$$\sum_{v \in V} a_v^{\text{B}} \leq 1 \qquad v \in V \tag{49}$$

$$\sum_{v \in V} d_v \leq D_{v_s} \qquad v_s \in V \tag{50}$$

$$\varphi_\ell^{\text{W}} \in \mathbb{Z}^+ \qquad \ell \in L \tag{51}$$

$$p_\ell^{\text{W}}, p_\ell^{\text{B}} \in \{0, 1\} \qquad \ell \in L \tag{52}$$

$$d_v \in \mathbb{Z}^+ \qquad v \in V \tag{53}$$

$$a_v^{\text{W}}, a_v^{\text{B}}, \, a_v \in \{0, 1\} \qquad v \in V. \tag{54}$$

The demand contributions of each working path are recorded by flow constraints (43), while the overall demand is satisfied if node v_s hosts a server. These paths may be towards different server nodes. Multiple paths are indirectly enforced by dual values, if it increases the backup sharing. Next, constraints (44) select a single backup path. Note that the same comments apply for those flow constraints as for the flow constraints of the pricing problems of Model I, we will not repeat them again here. Constraints (45) set the value of p_ℓ^{W}

— equalling 1 if working path p^{W} is provisioned on link ℓ — that is subsequently used for link disjoint paths and linearization in other constraints. Constraints (46), (47), and (48) are the same or the equivalent of (14), (15) and (16) in the pricing problem of Model I presented in Section 4. Required demands D_{v_s} are ensured by constraints (50), summing up all demand contributions. The last four sets of constraints, (51)–(54), set domains of the variables.

In order to linearize the third term of objective, we introduce the variable $\delta_{\ell\ell'} \in \mathbb{Z}^+$, where $\delta_{\ell\ell'} = \varphi_\ell^{\text{W}}\, p_\ell^{\text{B}}$, and add the following three sets of constraints:

For $\ell, \ell' \in L : \ell' \notin \{\ell, \text{OPP}(\ell)\}$:

$$\delta_{\ell\ell'} \leq D_{v_s}\, p_{\ell'}^{\text{B}} \tag{55}$$

$$\delta_{\ell\ell'} \leq \varphi_\ell^{\text{W}} \tag{56}$$

$$\delta_{\ell\ell'} \geq \varphi_\ell^{\text{W}} - D_{v_s}\,(1 - p_{\ell'}^{\text{B}}). \tag{57}$$

After linearization, the reduced cost expression can be written as follows:

$$\overline{\text{COST}}_1^{\text{II}} = \sum_{\ell \in L} \varphi_\ell^{\text{W}} - u_{v_s}^{(29)} + \sum_{\ell \in L} \sum_{\ell' \in L : \ell \neq \ell'} u_{\ell\ell'}^{(30)}\, \delta_{\ell\ell'} + \sum_{v \in V} u_v^{(32)}\, a_v. \tag{58}$$

5.2.2 Single Link or Node Failures

In the case of the scenario with protection against single link or node failures, the expression of the reduced cost is as follows:

$$\overline{\text{COST}_2}^{\text{II}} = \overline{\text{COST}}_1^{\text{II}} - \sum_{\ell \in L} \sum_{\ell' \in L : \ell \neq \ell'} u_{\ell\ell'}^{(30)}\, \delta_{v\ell'} + \sum_{v \in V} \sum_{\ell' \in L : \ell \neq \ell'} u_{v\ell'}^{(38)}\, \delta_{v\ell'}, \tag{59}$$

where $(u_{v\ell'}^{(38)})_{v\ell'}$ define the values of the dual vector associated with master problem constraints (38) and $\delta_{v\ell'} = \varphi_v^{\text{W}}\, p_\ell^{\text{B}}$, with the following three sets of linearization constraints:

For $v \in V, \ell' \in L$:

$$\delta_{v\ell'} \leq D_{v_s}\, p_{\ell'}^{\text{B}} \tag{60}$$

$$\delta_{v\ell'} \leq \varphi_\ell^{\text{W}} \tag{61}$$

$$\delta_{v\ell'} \geq \varphi_\ell^{\text{W}} - D_{v_s}\,(1 - p_{\ell'}^{\text{B}}). \tag{62}$$

For a single link or node (i.e., OXC failure), constraints (46) need to be replaced by constraints (63), and constraints (64) must be added:

$$\sum_{\ell \in \omega^-(v)} p_\ell^{\text{W}} + \sum_{\ell \in \omega^-(v)} p_\ell^{\text{B}} \leq 1 \qquad v \neq v_s, v \in V \tag{63}$$

$$a_v^{\text{W}} + a_v^{\text{B}} \leq 1 \qquad v \neq v_s, v \in V. \tag{64}$$

Constraints (63) ensure that, for each node, there is at most one incoming link (i.e., a pair of working/backup paths cannot share a node). Note that, except for the source, the constraints (63) also ensure protection against a single server failure except if there is a server located at the source node. Constraints (64) force the selection of two different locations, one for the working server location, a distinct one for the backup server location.

5.2.3 Single Link or Node or Server Failures

The reduced cost needs to be updated to incorporate master problem constraints (41):

$$\overline{\text{COST}}_3^{\text{II}} = \overline{\text{COST}}_2^{\text{II}} + \sum_{v \in V} \sum_{\ell' \in L} u^{(41)}\, a_{v_s}^{\text{W},c}\, p_{\ell'}^{\text{B},c}\, (\varphi_v^{\text{W},c} + D_v)\,, \tag{65}$$

where $u^{(41)}$ are the values of the dual variables associated with constraints (41). The nonlinear terms can be easily linearized as before.

As for Model I, the flow constraints for computing the backup paths need to be modified:

$$\sum_{\ell \in \omega^+(v)} p_\ell^{\text{B}} - \sum_{\ell \in \omega^-(v)} p_\ell^{\text{B}} = \begin{cases} 1 & \text{if } v = v_s \\ -a_v^{\text{B}} & \text{otherwise} \end{cases} \qquad v \in V. \tag{44'}$$

6 CG-ILP Model III: Server Based Configurations

6.1 Generalities

In this third model, configurations are centered around potential locations of servers. For a given potential server location at node v, a configuration c is characterized by the set of primary and backup paths for routing demands satisfied by the server located at v, see Figure 4 for an illustration. Therein, there are 2 configurations, one centered at v_6, the other one at v_7. Primary paths are represented by plain lines, while backup paths are represented by dotted lines. The first configuration comprises one backup path (for the demand originating at v_2), and one primary path (for part of the demand originating from v_1). The second configuration comprises one backup path (for the demand from v_1), and three primary paths (one for part of the demand from v_1, and two for the demand from v_2). Let C_v be the set of all configurations centered at location $v \in S$, and C be the overall set of configurations. We have: $C = \bigcup_{v \in S} C_v$.

6.2 Assumption

As for Model II, there is a unique backup path for the demands originating at a given source node v, while the demand can be served using several primary paths. Splitting the demand over several primary paths may allow saving some bandwidth for ensuring protection. For instance, in the example of Figure 4, dividing the demand originating at v_1 over two node disjoint paths (except for the source and destination nodes) allows restricting the backup bandwidth requirement to 4 units, while the overall primary demand is 7 units.

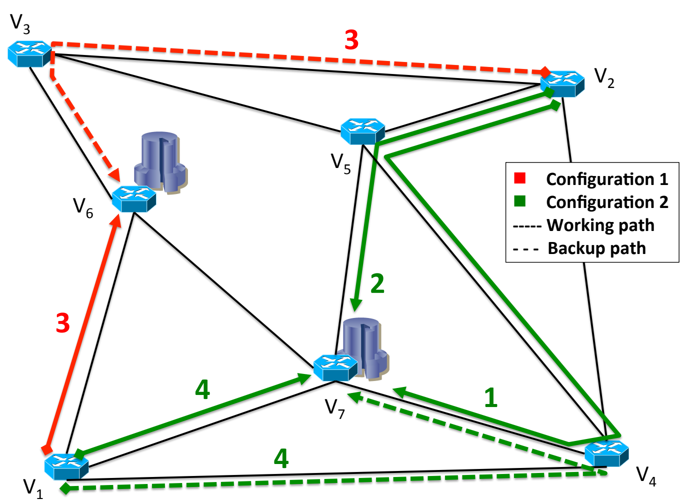


Figure 4: Configuration examples – Model III

6.3 Variables

There are two decision vectors $(z_c)_{c \in C}$ and $(y_v)_{v \in V}$ such that:

z_c $= 1$ if configuration c is selected, 0 otherwise.

y_v $= 1$ if there is a server located at node $v \in V$, 0 otherwise.

In addition, there are three integer vectors $(x^{v_s}_{\ell\ell'})_{\ell,\ell' \in L; v_s \in V_{\mathrm{D}}}$, $(x^{v_s}_{v\ell'})_{v \in V; \ell' \in L; v_s \in V_{\mathrm{D}}}$ and $(x^{\mathrm{B}}_{\ell})_{\ell \in L}$ such that:

$x^{v_s}_{\ell\ell'}$ $\in \mathbb{Z}^+$ equals the number of backup bandwidth units required on link ℓ' in order to protect link ℓ with respect to the primary paths originating from v_s.

$x^{v_s}_{v\ell'}$ $\in \mathbb{Z}^+$ equals the number of backup bandwidth units required on link ℓ' in order to protect node v with respect to the primary paths originating from v_s.

$x^{\mathbf{B}}_{\ell}$ $\in \mathbb{Z}^+$ equals the number of backup units required on link ℓ.

6.4 Parameters

$d^{\mathbf{W},c}_v$ $\in \mathbb{Z}^+$ equals the number of requests originating from v, which are accommodated (primary paths) by the server node v associated with c.

$b^{\mathbf{W},c}_{\ell v_s}$ $\in \mathbb{Z}^+$ equals the number of required working bandwidth units on link ℓ for provisioning primary paths from source node v_s to the server location s_c of configuration c.

$b^{\mathbf{W},c}_{vv_s}$ $\in \mathbb{Z}^+$ is the number working bandwidth units traversing node v required for provisioning primary paths from source node v_s to the server location s_c of configuration c.

$p^{\mathbf{B},c}_{\ell v_s}$ $\in \{0,1\}$, where $p^{\mathrm{B},c}_{\ell v_s} = 1$ if link ℓ is used by the unique backup path from v_s (from where the demand originates) to the server location s_c of configuration c, 0 otherwise.

$p^{\mathbf{B},c}_{vv_s}$ $\in \{0,1\}$, where $p^{\mathrm{B},c}_{vv_s} = 1$ if node v is used by the unique backup path from v_s (from where the demand originates) to the server location s_c of configuration c, 0 otherwise.

$p^{\mathbf{B},c}_{v}$ $\in \{0,1\}$, with $p^{\mathrm{B},c}_{v} = 1$ if there exists one backup path (there cannot be more than one) from v (from where the demand originates) to the server location s_c of configuration c, 0 otherwise. Also, $p^{B,c}_v = 1$ if $v = s_c$, where v is the server node associated with configuration c.

6.5 Master Problem

Recall that we use column generation, i.e., a decomposition into the master problem and several pricing problems, one for each potential server location.

6.5.1 Single Link Failure Scenario

We first present the master problem for the first failure scenario and then explain how to modify it for the second and third failure scenarios.

Objective function:

$$\min \quad \sum_{\ell \in L} \left(x^{\mathrm{B}}_{\ell} + \sum_{c \in C} \sum_{v \in V_{\mathrm{D}}} b^{\mathrm{W},c}_{\ell v}\, z_c \right) \tag{66}$$

Constraints

$$\sum_{v \in V} y_v \leq n_s \tag{67}$$

$$\sum_{c \in C_v} z_c \leq y_v \qquad v \in V \tag{68}$$

$$D_v y_v + \sum_{c \in C \setminus C_v} d^{\mathrm{W},c}_v\, z_c \geq D_v \qquad v \in V_{\mathrm{D}} \tag{69}$$

$$\sum_{c \in C} p_v^{\text{B},c}\, z_c = 1 \qquad v \in V_{\text{D}} \tag{70}$$

$$\sum_{c \in C} b_{\ell v}^{\text{W},c} z_c \leq M \left(1 - \sum_{c \in C} p_{\ell v}^{\text{B},c} z_c\right) \qquad \ell \in L, v \in V_{\text{D}} \tag{71}$$

$$\sum_{c \in C_v} b_{\ell v}^{\text{W},c} z_c \leq\ x_{\ell\ell'}^v + M \left(1 - \sum_{c \in C} p_{\ell v}^{\text{B},c} z_c\right) \qquad \ell, \ell' \in L : \ell' \notin \{\ell, \text{OPP}(\ell)\},\ v \in V_{\text{D}} \tag{72}$$

$$\sum_{v \in V_{\text{D}}} x_{\ell\ell'}^v \leq x_{\ell'}^{\text{B}} \qquad \ell, \ell' \in L : \ell' \notin \{\ell, \text{OPP}(\ell)\} \tag{73}$$

$$\sum_{c \in C_v} \sum_{\ell \in \omega^+(v)} p_{\ell v}^{\text{B},c} z_c = 1 - y_v \qquad v \in V_{\text{D}} \tag{74}$$

$$z_c \in \{0,1\} \qquad c \in C \tag{75}$$

$$y_v \in \{0,1\} \qquad v \in V \tag{76}$$

$$x_{\ell\ell'}^v \in \mathbb{Z}^+ \qquad \ell \in L, \ell' \in L : \ell' \notin \{\ell, \text{OPP}(\ell)\},\ v \in V \tag{77}$$

$$y_v \in \{0,1\} \qquad v \in V \tag{78}$$

$$x_\ell^{\text{B}} \in \mathbb{Z}^+ \qquad \ell \in L. \tag{79}$$

Constraint (67) is equivalent to:

$$\sum_{c \in C} z_c \leq n_s,$$

and expresses that we select at most n_s server locations.

Constraint (68) imposes that, if v is selected as a server location, then at most one configuration must be selected for that server location (we assume the configurations to be maximal in terms of demands satisfied by a given configuration). On the other hand, if v is not selected as a server location, then no configuration centered at v is selected. Constraints (69) are the demand constraints. Note that the summation is over $C \setminus C_v$ and not C_v as in Model I, due to the different definitions of the configurations in those two models. Note that if v is the data center location of configuration c, then demand of v is trivially satisfied (as v hosts a server). Otherwise, primary paths need to be established in order to grant the job requests of source node v. Constraints (70) enforce that, for a given source node, there is a unique backup path (but possibly several working paths). They can be equivalently written:

$$\sum_{c \in C_v} z_c + \sum_{c \in C \setminus C_v} p_v^{B,c}\, z_c \leq 1 \qquad v \in V_{\text{D}},$$

with the convention (see definition of the parameters) that $p_v^{\text{B},c} = 1$ if $v = s_c$, and 0 otherwise.

Constraints (71) ensure that, for a given source (demand), primary and backup paths are link disjoint. However, there can be several primary paths going through the same given link. In other words,

$$\sum_{c \in C} p_{\ell v_s}^{\text{B},c} z_c \in \{0,1\} \quad \text{in the optimal solution,}$$

while

$$\sum_{c \in C} b_{\ell v_s}^{\text{W},c} z_c \in \mathbb{Z}^+ \quad \text{in the optimal solution.}$$

Note that the M constant value should be set to be at least equal to the demand of the source node D_v in constraint (71).

In order to compute the protection bandwidth requirements, we need to proceed in two steps. First, with constraints (72), we compute the protection bandwidth requirements related to a given source node. Second, since we only allow one backup path, any link belonging to the backup path should account for the protection of any link belonging to a working path, as expressed in (73). Thus, the justification of constraints (71) is as follows:

1. ℓ' does not belong to any potential backup path. In such a case, we then have

$$\sum_{c \in C} p^{\mathrm{B},c}_{\ell' v}\, z_c = 0,$$

and therefore constraint (72) can be rewritten

$$\sum_{c \in C} b^{\mathrm{W},c}_{\ell v} z_c \leq x^{v}_{\ell \ell'} + M$$

i.e., it is a redundant constraint. As variable x^{B}_{ℓ} is minimized in the objective function of the pricing problem, due to constraints (73), $x^{v_s}_{\ell \ell'}$ will be set to 0 in the optimal solution.
2. ℓ' belongs to some potential backup paths, which are not selected in the solution. Similar to the previous case.
3. ℓ' belongs to a potential backup path, which is selected in the solution. In such a case, we have

$$\sum_{c \in C} p^{\mathrm{B},c}_{\ell' v}\, z_c = 1,$$

and link ℓ' should have enough capacity to carry all working traffic of any any link ℓ belonging to a working path as there is a unique backup path. This is done through (72) with the value of $x^{v}_{\ell \ell'}$ being minimized thanks to constraints (73) and the minimization of x^{B}_{ℓ} in the objective function

Constraints (73) can be equivalently rewritten:

$$\max_{\ell \in L} \sum_{v \in V_{\mathrm{D}}} x^{v}_{\ell \ell'} \leq x^{\mathrm{B}}_{\ell'} \qquad \ell' \in L. \tag{80}$$

They are identical to the constraints (39) of Model II. Their purpose is to compute the backup bandwidth requirements on ℓ' taking into account the bandwidth sharing.

6.5.2 Single Link or Node Failure Scenario

When we add the protection against a single node failure, we need to modify the computation of the protection bandwidth in order to deal with the working paths which are link disjoint, but not node disjoint. We therefore need to replace constraints (71), (72) and (73) by the following:

$$\sum_{c \in C} b^{\mathrm{W},c}_{v\, v_s} z_c \leq M \left(1 - \sum_{c \in C} p^{\mathrm{B},c}_{v\, v_s} z_c\right) \qquad v \in V, v_s \in V_{\mathrm{D}} : v \neq v_s \tag{81}$$

$$\sum_{c \in C} b^{\mathrm{W},c}_{v v_s} z_c \leq\ x^{v_s}_{v \ell'} \qquad \ell \in L, \quad v_s \in V_{\mathrm{D}}, v \in V : v \neq v_s \tag{82}$$

$$\sum_{v_s \in V_{\mathrm{D}} : v_s \neq v} x^{v_s}_{v \ell'} \leq x^{\mathrm{B}}_{\ell'} \qquad v \in V, \ell \in L. \tag{83}$$

Constraints (81) ensure that, for a given source (demand), primary and backup paths are node disjoint. Constraints (82) compute, for node $v \in V$ on a given working path, and for a given source node v_s from which the working path originates, how much bandwidth is needed to protect it on link ℓ'. Constraints (83) compute, for a given node $v \in V$, how much bandwidth is needed to protect against its failure on link ℓ' taking into account all possible source nodes for the working paths.

6.5.3 Single Link or Node or Server Failure Scenario

Again, demand constraints need to be modified in order to allow the selection of configurations with a backup path for the requests originating a location hosting a server:

$$\sum_{c \in C} d_v^{\text{W},c} \, z_c \geq D_v \tag{69$'$}$$

6.6 Pricing Problem

Each pricing problem outputs one promising configuration c, i.e., for a potential s_c server location. Indeed, it identifies the source nodes v (fully or partially or not at all) served (primary paths) by s_c and the source nodes v (fully or partially or not at all) served (backup paths) by s_c.

6.6.1 Single Link Failures

The objective of the pricing problem is the reduced cost of variable $z_c, c \in C_{s_c}$.

Let $u_v^{(68)} \leq 0$, $u_v^{(69)} \geq 0$, $u_v^{(70)} \lesseqgtr 0$, $u_{\ell v}^{(71)} \leq 0$, $u_{\ell\ell' v}^{(72)} \leq 0$ be the (real) values of the dual variables associated with master problem constraints (68), (69), (70), (71) (72), respectively.

For $c \in C_{s_c}$ (c is omitted in the expression of the reduced cost for the simplicity of the notations), we have:

$$\overline{\text{COST}}_1^{\text{III}} = \sum_{\ell \in L} \sum_{v \in V_{\text{D}}} b_{\ell v}^{\text{W}} - u_{s_c}^{(68)} - D_{s_c} u_{s_c}^{(69)} - \sum_{v \in V_{\text{D}} \setminus \{s_c\}} u_v^{(69)} \, d_v^{\text{W}} - \sum_{v \in V_{\text{D}}} u_v^{(70)} \, p_v^{\text{B}} - \sum_{\ell \in L} \sum_{v \in V_{\text{D}}} u_{\ell v}^{(71)} \, (b_{\ell v}^{\text{W}} + M p_{\ell v}^{\text{B}}) - \sum_{\ell \in L} \sum_{\ell' : \ell' \notin \{\ell, \text{OPP}(\ell)\}} \sum_{v \in V_{\text{D}}} u_{\ell\ell' v}^{(72)} \, b_{\ell v}^{\text{W}}. \tag{84}$$

If, instead of (69), we use (69′), the expression of the reduced cost becomes:

$$\overline{\text{COST}}_1^{\text{III}} = \sum_{\ell \in L} \sum_{v \in V_{\text{D}}} b_{\ell v}^{\text{W}} - u_{s_c}^{(68)} - \sum_{v \in V_{\text{D}} \setminus \{s_c\}} u_v^{(69')} \, d_v^{\text{W}} - \sum_{v \in V_{\text{D}}} u_v^{(70)} \, p_v^{\text{B}} - \sum_{\ell \in L} \sum_{v \in V_{\text{D}}} u_{\ell v}^{(71)} \, (b_{\ell v}^{\text{W}} + M p_{\ell v}^{\text{B}}) - \sum_{\ell \in L} \sum_{\ell' : \ell' \notin \{\ell, \text{OPP}(\ell)\}} \sum_{v \in V_{\text{D}}} u_{\ell\ell' v}^{(72)} \, b_{\ell v}^{\text{W}}. \tag{85}$$

After some algebraic manipulations, we get:

$$\overline{\text{COST}}_1^{\text{III}} = \overbrace{-u_{s_c}^{(68)} - D_{s_c} u_{s_c}^{(69)}}^{\text{constant}} - \sum_{v \in V_{\text{D}} \setminus \{s_c\}} u_v^{(69)} \, d_v^{\text{W}} + \sum_{\ell \in L} \sum_{v \in V_{\text{D}}} \left(1 - u_{\ell v}^{(71)} - \sum_{\ell' \in L : \ell' \notin \{\ell, \text{OPP}(\ell)\}} u_{\ell\ell' v}^{(72)} \right) b_{\ell v}^{\text{W}} - \sum_{v \in V_{\text{D}}} u_v^{(70)} \, p_v^{\text{B}} - M \sum_{v \in V_{\text{D}}} \sum_{\ell \in L} u_{\ell v}^{(71)} \, p_{\ell v}^{\text{B}}. \tag{86}$$

Again, if instead of (69), we use (69′), the expression of the reduced cost becomes:

$$\overline{\text{COST}}_1^{\text{III}} = \overbrace{-u_{s_c}^{(68)}}^{\text{constant}} - \sum_{v \in V_{\text{D}} \setminus \{s_c\}} u_v^{(69)} \, d_v^{\text{W}} + \sum_{\ell \in L} \sum_{v \in V_{\text{D}}} \left(1 - u_{\ell v}^{(71)} - \sum_{\ell' \in L : \ell' \notin \{\ell, \text{OPP}(\ell)\}} u_{\ell\ell' v}^{(72)} \right) b_{\ell v}^{\text{W}} - \sum_{v \in V_{\text{D}}} u_v^{(70)} \, p_v^{\text{B}} - M \sum_{v \in V_{\text{D}}} \sum_{\ell \in \omega^+(v)} u_{\ell v}^{(71)} \, p_{\ell v}^{\text{B}}. \tag{87}$$

Now, for the pricing constraints: instead of flows originating from a single source node, in Model III we have multi-flows, with possibly one flow originating from each source node.

$$\sum_{\ell \in \omega^-(v)} b^{\text{W}}_{\ell v} = 0 \qquad v \in V_{\text{D}} \tag{88}$$

$$\sum_{\ell \in \omega^+(s_c)} b^{\text{W}}_{\ell s_c} = 0 \tag{89}$$

$$\sum_{\ell \in \omega^+(v')} b^{\text{W}}_{\ell v} - \sum_{\ell \in \omega^-(v')} b^{\text{W}}_{\ell v} = \begin{cases} d^{\text{W}}_v & \text{if } v' = v \neq s_c \\ -d^{\text{W}}_v & \text{if } v' = s_c \neq v \\ 0 & \text{otherwise} \end{cases} \qquad v \in V_{\text{D}}, v' \in V \tag{90}$$

$$d^{\text{W}}_v \leq D_v \qquad v \in V_{\text{D}} \tag{91}$$

$$\sum_{\ell \in \omega^-(v)} p^{\text{B}}_{\ell v} = 0 \qquad v \in V_{\text{D}} \tag{92}$$

$$\sum_{\ell \in \omega^+(s_c)} p^{\text{B}}_{\ell s_c} = 0 \tag{93}$$

$$\sum_{\ell \in \omega^+(v')} p^{\text{B}}_{\ell v} - \sum_{\ell \in \omega^-(v')} p^{\text{B}}_{\ell v} = \begin{cases} p^{\text{B}}_v & \text{if } v' = v \neq s_c \\ -p^{\text{B}}_v & \text{if } v' = s_c \neq v \\ 0 & \text{otherwise} \end{cases} \qquad v \in V_{\text{D}}, v' \in V \tag{94}$$

$$b^{\text{W}}_{\ell v} \in \mathbb{Z}^+ \qquad \ell \in L, v \in V_{\text{D}} \tag{95}$$

$$p^{\text{B}}_{\ell v} \in \{0, 1\} \qquad \ell \in L, v \in V_{\text{D}} \tag{96}$$

$$d^{\text{W}}_v \in \mathbb{Z}^+ \qquad v \in V_{\text{D}} \tag{97}$$

$$p^{\text{B}}_v \in \{0, 1\} \qquad v \in V_{\text{D}} \tag{98}$$

There is no need to explicitly forbid primary and backup paths to use different links, as such a requirement is indirectly reinforced in constraints (71), (72) and (73) of the master problem, where we compute the requirements for the backup bandwidth. In addition, it is likely that the working path and the backup path of a given source will not belong to the same configuration.

Constraints (88) forbid incoming working paths at the source nodes, while constraints (89) forbid outgoing working paths from the server location. Similarly, constraints (92) forbid incoming backup links at the source node of a request and constraints (93) forbid outgoing backup links at the destination node of a request.

While constraints (91) are not absolutely required — there is no reason to overprovision the demand while we aim at minimizing the bandwidth requirement — they will be satisfied in any optimal solution.

6.6.2 Single Link or Node Failures

The reduced cost needs to be updated following the addition of constraints (81) and (82):

$$\overline{\text{COST}}_2^{\text{III}} = \sum_{\ell \in L} \sum_{v \in V_{\text{D}}} b^{\text{W}}_{\ell v} - u^{(68)}_{s_c} - D_{s_c} u^{(69)}_{s_c} - \sum_{v \in V_{\text{D}} \setminus \{s_c\}} u^{(69)}_v d^{\text{W}}_v - \sum_{v \in V_{\text{D}}} u^{(70)}_v p^{\text{B}}_v - \sum_{v \in V} \sum_{v \in V_{\text{D}}} u^{(81)}_{vv_s} \left(b^{\text{W}}_{v_s v} + M p^{\text{B}}_{vv_s}\right) - \sum_{\ell \in L} \sum_{v \in V} \sum_{v \in V_{\text{D}}} u^{(82)}_{\ell v_s v} b^{\text{W}}_{\ell v}. \tag{99}$$

6.6.3 Single Link or Node or Server Failures

The reduced cost remains identical:

$$\overline{\text{COST}}_3^{\text{III}} = \overline{\text{COST}}_2^{\text{III}}. \tag{100}$$

Constraints (94) are changed as follows:

$$\sum_{\ell \in \omega^+(v')} p^{\mathrm{B}}_{\ell v} - \sum_{\ell \in \omega^-(v')} p^{\mathrm{B}}_{\ell v} = \begin{cases} p^{\mathrm{B}}_v & \text{if } v' = v \\ -p^{\mathrm{B}}_v & \text{if } v' = s_c \\ 0 & \text{otherwise} \end{cases} \qquad v \in V_{\mathrm{D}}, v' \in V. \tag{94'}$$

7 Solution Process

For all three models, we use the solution process described in Figure 5.

We solve the continuous relaxation of the models using the column generation method until either the optimality condition is satisfied, or until an ε-optimality condition is satisfied when the convergence is too slow. Indeed, for Model III, we observed a very slow convergence in practice, which is not due to degeneracy. We then define a stopping condition in order to allow reaching an optimized LP value within reasonable computing times: if the LP objective value is not improved by more than 0.05% (Model III(a)) or 0.01% (Model III(b)) during the last 50 iterations, we stop the LP solution process. While this implies we can no longer guarantee to reach the optimal LP solution, it allows reaching optimized LP solutions within a reasonable amount of time. The choice of the numerical values used in the stopping criteria are justified by additional experiments presented in Appendix A. We also investigated the generation of initial configurations using the ILP solution of Model II (as Models II and III both assume a single backup path per source), leading to the so-called Model III(c). Therein, we use the stopping criterion of no more than 0.5% improvement of the value of the objective function during the last 50 iterations. We denote by $z^\star_{\mathrm{LP}}$ and $\tilde{z}_{\mathrm{LP}}$ the optimal and ε-optimal values of the linear relaxation, respectively.

Rather than developing a costly branch-and-cut algorithm, we next solve the incumbent integer restricted master problem (RMP) (i.e., the restricted master problem made of the columns generated until the linear relaxation was (ε-)optimally solved) using the MILP and then the ILP solver of CPLEX [28]. Rather than

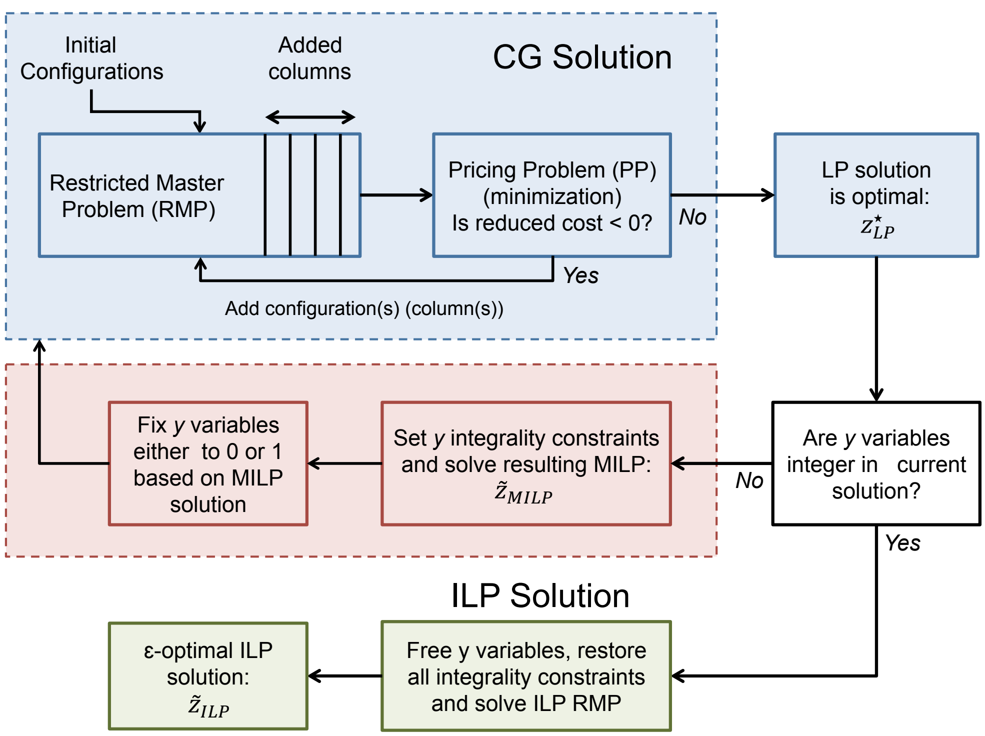


Figure 5: Flowchart of the solution process

solving directly the integer RMP, we first only restore the integrality conditions for the y_v variables, solve the corresponding MILP RMP, fix the y_v variables to their value in the MILP optimal solution (denoted by $\tilde{z}_{\text{MILP}}$ since, while it is the optimal value of the incumbent MILP RMP, it is not necessarily the optimal MILP solution of the original model), and then again generate improving columns using the remaining continuous variables (i.e., the $x_{\square}$ and the z_c variables). Once the (ε-)optimality condition is again satisfied for the LP relaxation of the RMP with binary valued y_v variables, we release the y_v variables and solve the resulting incumbent ILP RMP using the ILP CPLEX solver. Let $\tilde{z}_{\text{ILP}}$ denote the value of the resulting ILP solution.

The two-phase process was validated in [11], where we compared the quality of the solutions and the efficiency of the solution scheme using *(i)* a one-phase ILP solution (solving directly the RMP associated with the first optimal LP solution with the ILP Cplex solver), and *(ii)* a two-phase ILP solution process. We found that the two-phase solution process generated ILP solutions that were around 10% better than those with the one-phase solution process. For this reason, we here only consider the two phase ILP solution process.

For all models, we generated an initial solution, i.e., an initial set of configurations, as follows: we use the pricing problems, without changing the set of constraints, but with replacing the reduced cost objective by the minimization of the overall working and backup bandwidth requirements instead. This way of proceeding is not necessarily the fastest way to generate initial configurations, but a very practical one, as it does not require developing new models or algorithms.

The characteristics of our three models and their solutions have been summarized in Table 1.

Table 1: Summary of the three models and their solution variants

Model	Configuration	Variables	Solution
I	Associated with 1 source: 1 working, 1 backup path ≥ 1 config. used per source	$z_c \in \mathbb{Z}^+$ $x^{\text{B}}_{\ell} \in \mathbb{Z}^+; x^v_{\ell\ell'} \in \mathbb{Z}^+$ $y_v \in \{0,1\}$	See flowchart in Figure 5 LP solved optimally $\rightsquigarrow z^{\star,\text{I}}_{\text{LP}}$ ILP solved heuristically $\rightsquigarrow \tilde{z}^{\text{I}}_{\text{ILP}} \geq z^{\star,\text{I}}_{\text{ILP}}$
II	Associated with 1 source: ≥ 1 working, 1 backup paths exactly 1 config. per source only 1 backup path per source	$z_c \in \mathbb{Z}^+$ $x^{\text{B}}_{\ell} \in \mathbb{Z}^+; x^v_{\ell\ell'} \in \mathbb{Z}^+$ $y_v \in \{0,1\}$ 2^{nd} and 3^{rd} scenarios: $x^{v_s}_{v\ell'} \in \mathbb{Z}^+$	See flowchart in Figure 5 LP solved optimally $\rightsquigarrow z^{\star,\text{II}}_{\text{LP}}$ ILP solved heuristically $\rightsquigarrow \tilde{z}^{\text{II}}_{\text{ILP}} \geq z^{\star,\text{II}}_{\text{ILP}}$
III	Associated with 1 server: ≥ 1 working, ≥ 1 backup paths exactly 1 config. per server only 1 backup path per source	$z_c \in \mathbb{Z}^+$ $x^{\text{B}}_{\ell} \in \mathbb{Z}^+$ $y_v \in \{0,1\}$ 2^{nd} and 3^{rd} scenarios: $x^{v_s}_{v\ell'} \in \mathbb{Z}^+$	III(a) Stop CG if during last 50 iterations, LP cost reduction was $< 0.05\%$ III(b) Stop CG if last during 50 iterations, LP cost reduction was $< 0.01\%$ III(c) Start from ILP solution of Model II, and stop CG if during 50 iterations LP cost reduction was $< 0.5\%$ LP solved heuristically $\rightsquigarrow \tilde{z}^{\text{III}}_{\text{LP}} \geq z^{\star,\text{III}}_{\text{LP}}$ ILP solved heuristically $\rightsquigarrow \tilde{z}^{\text{III}}_{\text{ILP}} \geq z^{\star,\text{III}}_{\text{ILP}}$

8 Numerical Results

8.1 Data instances

We considered the network topology of the European backbone network taken from [29], with 37 nodes and 57 links, depicted in Figure 6. We conducted experiments with $n_{\text{s}} = 3$ and 5 data center locations hosting servers.

Different demand sets have been randomly generated as sets of unit requests, assuming them to be uniformly distributed among the source nodes. A unit request calls for a full wavelength to be provisioned between the source and a destination to be chosen among the server nodes (using anycast). Experiments have been conducted with 100 up to 1,000 unit requests, where for each demand size we have repeated the experiment for 10 instances. Also, note that when increasing the request set for a particular random seed from D_1 to D_2 (thus with $|D_2| > |D_1|$ unit requests), we have $D_2 \supset D_1$.

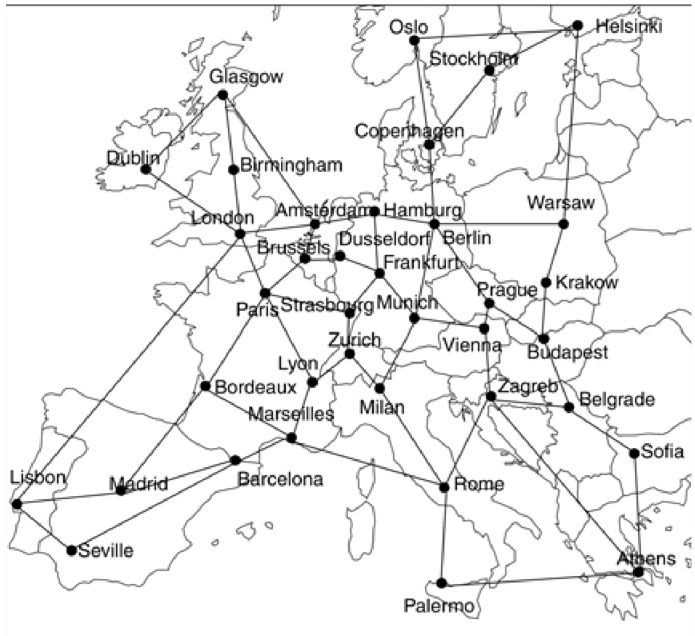


Figure 6: European network

8.2 Accuracy of the solutions of the three models

In this section, we compare the quality of the solutions as output by the three CG models, as described in Sections 4 to 6, in a scenario with $n_{\text{s}} = 5$ server locations and a single link failure protection scheme.

We looked at the optimal (Models I and II) or optimized (Model III) values of the LP relaxations (see Table 2), the optimal MILP values (see Figure 7), and the best ILP values (see Table 3), together with the computing times (for reaching the optimal/optimized LP values in Table 2 and the overall computing times for reaching the final ILP values in Table 3) and the number of generated/selected configurations. All values correspond to averages over 10 randomly generated instances of the respective demand size (# requests).

Highlighted values correspond to the best values between Models II and III and the different solutions (III(a,b,c)) of Model III. Model I is not included in that comparison due to different assumption with respect to the number of backup paths (several vs. only one for Models II and III).

The comparisons of the various lower (optimal values of the LP relaxations) and upper (ILP values) bounds lead to the following observations. Note that the LP values of Model III are valid lower bounds due to the stopping condition described in Section 7. However, all ILP values are upper bounds on the optimal ILP values.

The lowest LP (Table 2) and ILP (Table 3) values are obtained by Model I, as expected, since it is the only model in which multiple backup paths are allowed. In Section 8.5, we will further investigate the protection schemes generated by Model I and the average number of backup paths. Under the restricted assumption of a single backup path, Model III provides, by far, the best lower bounds, except for 800 and 900 jobs, but then the upper bounds (ILP values) of Model II and Model III(b) or (c) are very close under the assumptions of a single backup path.

While Model III(b) gives the best lower bounds among the three variants of Model III, i.e., when the stopping criteria is no improvement larger than 0.01% during the last 50 iterations, the computing times are much larger than Model III(a), without providing significantly better ILP solutions. Observe that the LP values of Model III(c) are worse than those of Model II, since the stopping condition for Model III(c) does not guarantee reaching the optimal LP value.

Table 2: Optimal (Models I and II) and optimized (Model III) LP values (5 servers)

# requests	$z^\star_{\text{LP}}$ I	$z^\star_{\text{LP}}$ II	$\tilde{z}_{\text{LP}}$ III(a)	$\tilde{z}_{\text{LP}}$ III(b)	$\tilde{z}_{\text{LP}}$ III(c)	CPU$_{\text{LP}}$ (seconds) I	CPU$_{\text{LP}}$ II	CPU$_{\text{LP}}$ III(a)	CPU$_{\text{LP}}$ III(b)	CPU$_{\text{LP}}$ III(c)	Number of generated configurations I	II	III(a)	III(b)	III(c)
100	127	181	160	**152**	**152**	332	8,878	5,194	28,223	4,153	134	986	552	1,349	589
200	265	330	341	**326**	348	322	8,477	3,361	14,253	2,665	133	1011	448	1,064	463
300	405	483	491	**469**	482	364	5,855	6,348	23,828	3,774	136	754	585	1,266	545
400	549	648	654	**631**	707	310	6,135	5,559	17,679	4,107	132	742	555	1,175	543
500	695	798	831	803	**784**	303	8,537	7,049	22,874	4,688	133	1012	633	1,287	591
600	842	970	981	**947**	962	332	6,463	7,651	21,237	4,638	133	781	624	1,129	568
700	995	1,127	1,153	1,152	1,139	298	5,794	8,628	13,350	5,092	130	761	619	1,081	508
800	1,120	**1,280**	1,320	1,295	1,293	328	5,905	5,920	17,498	4,756	132	767	565	1,220	543
900	1,274	**1,434**	1,476	1,439	1,508	333	6,505	5,890	16,940	4,028	131	790	563	1,159	473
1,000	1,428	1,628	1,663	**1,590**	1,594	286	5,883	7,656	20,281	3,807	130	700	569	1,292	544

Note that the computing times of generating the initial configurations, i.e., solving Model II, are not concluded in the computing times of Method III(c)

Table 3: Best ILP values (5 server locations)

# requests	$\tilde{z}_{\text{ILP}}$ I	$\tilde{z}_{\text{ILP}}$ II	$\tilde{z}_{\text{ILP}}$ III(a)	$\tilde{z}_{\text{ILP}}$ III(b)	$\tilde{z}_{\text{ILP}}$ III(c)	CPU$_{\text{OVERALL}}$ (seconds) (includes all solution phases) I	II	III(a)	III(b)	III(c)	Number of generated configurations I	II	III(a)	III(b)	III(c)	# selected config. I
100	281	**306**	309	309	**306**	496	**9,361**	23,144	88,682	21,120	217	1,019	697	1,512	657	35
200	557	606	**604**	**604**	606	491	**8,888**	14,967	44,118	12,542	218	1,041	579	1,237	526	38
300	822	902	**896**	**896**	902	509	**6,508**	21,610	62,974	16,893	214	798	678	1,339	641	37
400	1,099	1,199	**1,192**	**1,192**	1,199	465	**6,741**	23,912	56,626	19,127	216	784	679	1,370	654	43
500	1,384	**1,480**	1,490	1,490	**1,480**	463	**9,378**	26,622	64,373	20,768	209	1,069	798	1,434	723	41
600	1,637	**1,762**	1,780	1,787	**1,762**	499	**7,843**	28,821	64,216	22,264	221	865	711	1,329	682	43
700	1,993	**2,037**	2,079	2,079	**2,037**	408	**7,420**	31,583	47,127	19,269	178	862	710	1,261	598	35
800	2,309	**2,309**	2,354	2,354	**2,309**	465	**8,078**	23,723	54,991	18,023	178	904	680	1,333	693	30
900	2,450	**2,591**	2,658	2,658	**2,591**	503	**9,003**	25,496	53,568	14,341	219	950	640	1,328	527	40
1,000	2,730	**2,892**	2,922	2,922	**2,892**	448	**9,152**	31,251	68,404	21,260	218	900	719	1,493	783	40

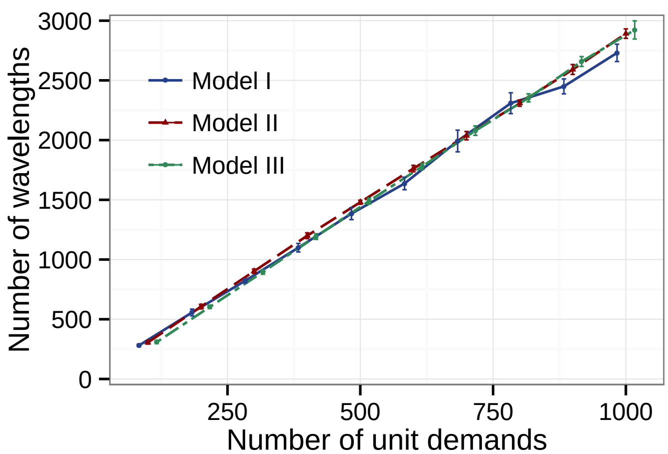


Figure 7: Comparison of all three models with respect to their bandwidth requirements

We now look at the empirical lower bounds as given by the $\tilde{z}_{\text{MILP}}$ values. In Figure 7, we provide the curves of the number of unit requests vs. $\tilde{z}_{\text{MILP}}$, together with the confidence intervals associated with the values $\tilde{z}_{\text{MILP}}$ over a set of 5 randomly generated problems. We then observe that the standard deviation values are rather small, and that all three models generate similar results, except for Model I, especially when the number of unit requests increases. Indeed, since Model I allows several backup paths, it achieves a smaller overall bandwidth requirement.

As a conclusion on the comparison of the lower and upper bounds for all three models, we observe that, while those gap values are not very small, they are, to some extent, very similar to the gap values observed for the p-center problem [30], which is a priori simpler as there is no backup path to compute.

The computing times (CPU) (up to the LP) for Model I are always the lowest ones, while under the assumption of a single backup path, Model III(c) has the lowest ones, thanks to the "warm" start and the relaxed optimality condition. However, if we take into account the computing times of generating an initial solution, Model III(a) is the most economical one. We conclude that it would be worth designing a fast heuristic in order to provide an initial solution, as the complete solution of Model II, even if it provides a good initial solution, is computationally quite expensive. However, while providing a "warm" start helps to speed up the LP solution of Model III(c), it does not help to speed up the ILP solution: its overall computing times are smaller than Models III(a) and III(b), but much higher than those of Model II (or even Model I).

We have also analyzed the number of generated configurations before reaching the (ε-)optimal LP and ILP solutions. Model I generates by far the lowest number of configurations, while, under the assumption of a single backup path, it is Model III(c), whether it is for reaching the (ε-)optimal LP values or the best ILP values.

8.3 Server selection

In order to further investigate the different ILP values of the three models, we examined the selection of server nodes made by all three models under the single link failure protection scheme. Results are depicted in Figure 8. Along the vertical axis, we put the average number of times (expressed as a percentage) a given server location was selected by a given model. We observe that the three models mostly select the same set of server nodes, i.e., London, Lyon, Zurich, Berlin, Vienna. Only Model I selects Madrid as a server node in very few instances.

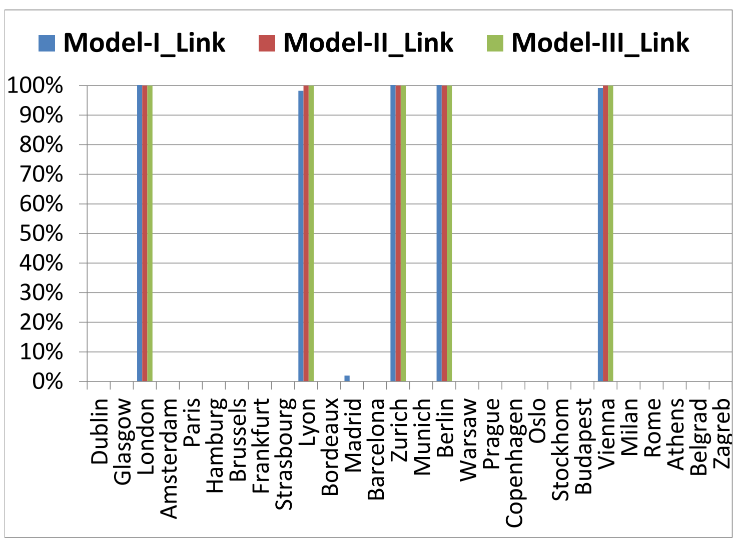
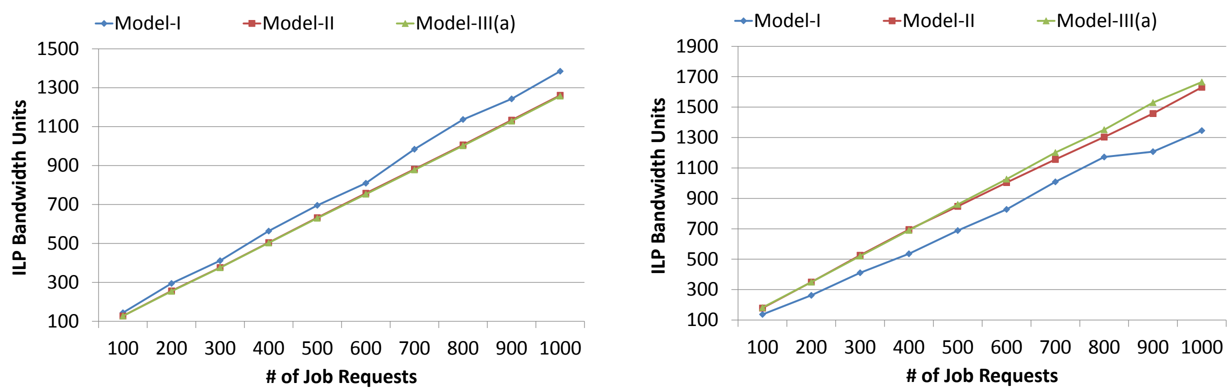


Figure 8: Selection of server nodes under link protection scheme

Although the models do select the same server locations, they do not necessarily select the same paths, nor the same number of backup paths (one for Models II–III, several for Model I). Figure 9 plots the amount of working and backup capacity for each of the three models. The bandwidth requirements are very similar for Models II and III(a) (see also Table 3). For Model I, the working bandwidth requirements are larger, while its backup requirements are lower than for Models II and III, but the sum of those bandwidth requirements are similar to those of Models II and III. This is mainly explained by the path structure: only one backup path for Models II and III, while there could be several for Model I.

8.4 Bandwidth requirements

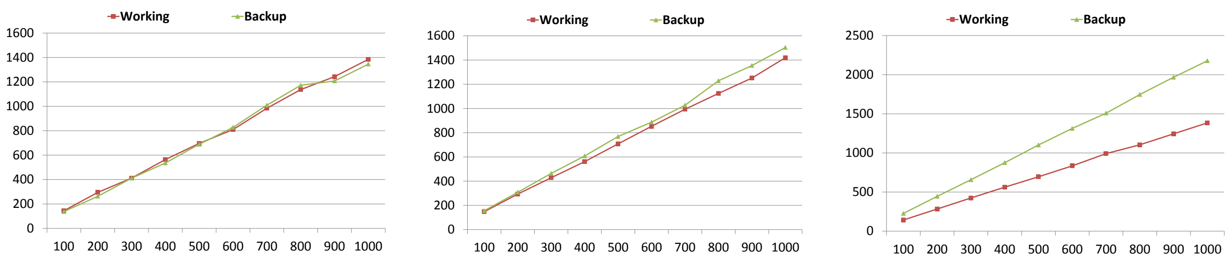
Comparative bandwidth requirements for the three failure scenarios are depicted in Figure 10. We only explicitly present results for Model I, since the qualitative trends are the same for all three models, even if



(a) Working bandwidth requirements

(b) Backup bandwidth requirements

Figure 9: Comparison of the bandwidth requirements of the three models



(a) Link

(b) Link+Node

(c) Link+Node+Server

Figure 10: ILP solution for working and backup bandwidth units under Model I (5 Servers)

the numerical values are not exactly the same. For the bandwidth requirements, we distinguish the working and the backup bandwidth requirements, as a function of the number of unit requests.

We observe that, for Model I, for the first two failure scenarios, the total backup bandwidth is comparable to the working. Indeed, as the backup paths are shared among disjoint working paths, they do not require more bandwidth, even if they may be longer. This observation is independent of the number of requests. However, for the third failure scenario, i.e., where we protect against link/node/server failures, backup path capacity always exceeds the working capacity, and the more so as the number of requests increases. This can be explained by the fact that when a data center collocated with network node v_s fails, also the requests originating at that node need to be rerouted, whereas in the case of only network node failure, we only need to reroute traffic that was sent to the DC at v_s coming from other source nodes $v \neq v_s$.

Figure 11 plots the overall bandwidth requirements for each failure scenario, where heights of each bar are averages over 10 randomly generated instances with the same number of unit requests. As expected from the analysis of the results in Figure 10, we observe that the single link or node failure scenario only uses slightly more bandwidth than the single link failure scenario: on average, there is a 6% difference. However, relative differences, i.e., $\left(\tilde{z}^{\text{III}}_{\text{ILP}}(\text{link/node/server}) - \tilde{z}^{\text{III}}_{\text{ILP}}(\text{link})\right) / \tilde{z}^{\text{III}}_{\text{ILP}}(\text{link})$, increases to around 30% for the comparison with the single link/node/server failure scenario.

We next present some results for 3 server nodes, again only for Model I, in Figure 12. Therein, the single link or node (resp. the single link or node or server) failure scenarios require an average of 7% (resp. 29%) extra bandwidth units compared to the single link failure scenario.

Comparison between 3 and 5 server locations shows that the case with 3 server locations uses around 22%, 23%, and 22% extra bandwidth compared to 5 server locations respectively in the single link, single link/node, single link/node/server failure scenario.

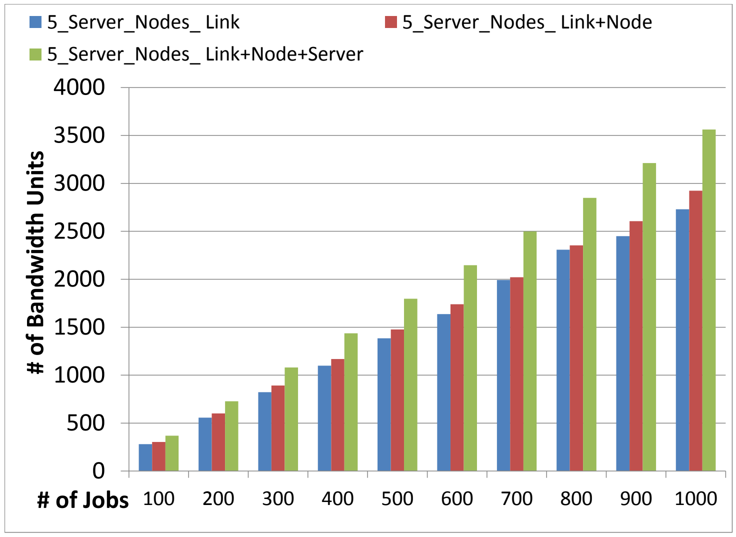


Figure 11: Different protection schemes – Model I – 5 server nodes

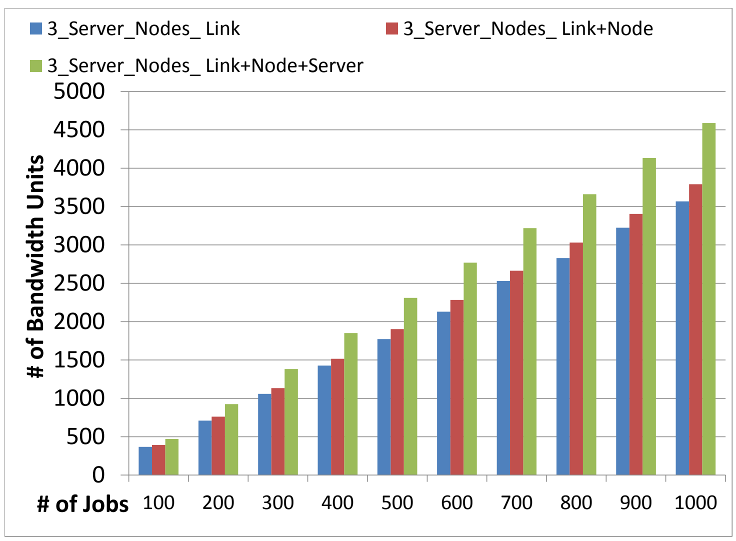


Figure 12: Different protection schemes – Model I – 3 server nodes

8.5 One vs. multiple backup paths

8.5.1 All Models

In order to compare the best ILP solutions obtained by the different models, we computed the following gaps, where we use variant III(a) as solution for Model III:

$$\text{GAP}_{\text{ II vs. I}} = \frac{\tilde{z}^{\text{II}}_{\text{ILP}} - \tilde{z}^{\text{I}}_{\text{ILP}}}{\tilde{z}^{\text{I}}_{\text{ILP}}} \qquad \text{GAP}_{\text{ III(a) vs. I}} = \frac{\tilde{z}^{\text{III(a)}}_{\text{ILP}} - \tilde{z}^{\text{I}}_{\text{ILP}}}{\tilde{z}^{\text{I}}_{\text{ILP}}}.$$

Numerical values are described in Table 6, from which we deduce that the ILP value varies from 0.02% to 9.85% depending on the number of traffic requests, leading to an average value of 6.74 % for Model II vs. Model I, and of 7.38 % for Model III vs. Model I. It indicates an estimate of the advantage of one backup path vs. multiple ones. Differences between the Model II vs. Model III based average values result from the solution (in)accuracies.

Table 4: Model I – Working paths – Link or node or server node or server protection scheme (5 servers). A configuration with 0 paths implies that requests are served at a data center collocated with the request source

(a) Working Paths under Single Link Failure

# of destinations	1	1	1	1	2	2	3	2	3	4	Σ
# of paths	1	0	2	3	2	3	3	4	4	4	
Size	%										
100	58.3	17.9	11.5	1.2	6.3	2.8	0.0	2.0	0.0	0.0	100
200	51.2	17.9	12.3	2.0	10.3	6.0	0.0	0.4	0.0	0.0	100
300	53.2	17.9	15.0	1.8	9.6	2.1	0.0	0.4	0.0	0.0	100
400	42.5	17.9	16.3	5.2	8.7	6.3	0.4	2.0	0.4	0.4	100
500	42.1	17.9	19.4	4.4	9.1	4.0	0.8	0.8	0.8	0.8	100
600	45.7	17.9	16.8	4.3	7.9	5.0	0.0	1.8	0.4	0.4	100
700	50.4	17.9	15.6	0.9	9.4	3.6	0.4	1.8	0.0	0.0	100
800	62.5	17.9	9.5	0.0	5.4	2.4	0.0	1.2	0.6	0.6	100
900	49.3	17.9	15.7	1.4	9.6	3.6	0.4	1.4	0.4	0.4	100
1000	31.1	17.9	8.6	0.4	23.6	6.8	4.6	0.7	3.2	3.2	100
AVERAGE	48.6	17.9	14.1	2.1	10.0	4.3	0.7	1.2	0.6	0.6	100

(b) Working Paths under Single Link or Node Failure

# of destinations	1	1	1	1	2	2	3	2	3	4	Σ
# of paths	1	0	2	3	2	3	3	4	4	4	
Size	%										
100	54.3	17.9	7.9	2.1	10.4	6.1	0.4	1.1	0.0	0.0	100
200	47.9	17.9	11.4	2.9	11.1	6.8	0.4	1.8	0.0	0.0	100
300	45.4	17.9	15.4	1.8	9.3	6.1	0.7	3.6	0.0	0.0	100
400	45.4	17.9	12.1	1.8	12.5	7.1	0.7	1.8	0.4	0.4	100
500	42.9	17.9	16.1	3.9	11.8	6.1	0.0	0.7	0.4	0.4	100
600	42.5	17.9	12.1	2.9	15.4	7.5	0.4	0.7	0.4	0.4	100
700	40.4	17.9	14.3	5.0	13.6	5.0	1.1	2.9	0.0	0.0	100
800	41.3	17.9	16.7	2.4	12.7	7.1	0.8	0.4	0.4	0.4	100
900	43.2	17.9	12.9	2.1	13.2	6.8	0.4	2.1	0.7	0.7	100
1000	37.7	17.9	8.1	0.0	25.3	5.8	2.6	0.6	1.0	1.0	100
AVERAGE	44.1	17.9	12.7	2.5	13.5	6.4	0.7	1.6	0.3	0.3	100

(c) Working Paths under Single Link or Node or Server Failure

# of destinations	1	1	1	1	2	2	3	2	3	4	Σ
# of paths	1	0	2	3	2	3	3	4	4	4	
Size	%										
100	64.6	17.9	10.7	0.4	5.7	0.7	0.0	0.0	0.0	0.0	100
200	61.8	17.9	13.6	0.7	5.7	0.4	0.0	0.0	0.0	0.0	100
300	55.7	17.9	12.9	1.1	10.4	2.1	0.0	0.0	0.0	0.0	100
400	57.5	17.9	12.1	0.7	7.5	3.2	0.4	0.0	0.4	0.4	100
500	55.0	17.9	13.6	0.7	10.7	1.4	0.4	0.4	0.0	0.0	100
600	58.2	17.9	13.9	0.7	8.6	0.0	0.4	0.4	0.0	0.0	100
700	56.8	17.9	16.8	0.0	7.1	1.4	0.0	0.0	0.0	0.0	100
800	59.1	17.9	13.5	0.0	7.1	1.6	0.4	0.4	0.0	0.0	100
900	59.3	17.9	13.9	0.4	7.1	1.4	0.0	0.0	0.0	0.0	100
1000	49.0	17.9	5.5	0.3	24.0	0.6	1.9	0.0	0.3	0.3	100
AVERAGE	57.7	17.9	12.7	0.5	9.4	1.3	0.3	0.1	0.1	0.1	100

8.5.2 Model I

Recall that Model I allows multiple working and multiple backup paths for each source node. Thus, we analyze the impact of allowing multiple paths to the same or different destination nodes under different protection schemes.

Results corresponding to averages over 10 randomly generated instances for each number of requests, are shown in Table 4 and Table 5. Each sub-table (a, b, c) corresponds to a specific failure scenario. Each column is associated with a number of distinct destinations for the working (or the backup) paths and the number of distinct working or backup paths (see the heading of the columns). For instance, 2 destinations and 4 paths in Table 4 means that there are some source nodes with 4 different working paths and there are two distinct destinations (i.e., server nodes) for those four paths. In each column, i.e., a path pattern for a given source node, we indicate the percentage in terms of demand served by this path pattern. The last column is a check that all requests have been accounted for, i.e., the sum of all percentages should be equal to 100%. The last

Table 5: Model I – Backup paths – Link or node or server node or server protection scheme (5 servers)

(a) Backup Paths under Single Link Failure

# of destinations	1	1	1	1	2	2	3	2	3	4	Σ
# of paths	1	0	2	3	2	3	3	4	4	4	
Size	%										
100	49.3	17.9	10.0	0.4	3.6	3.6	1.8	1.1	1.4	1.4	100
200	42.1	17.9	10.7	0.0	17.1	5.6	3.2	1.2	1.2	1.2	100
300	42.9	17.9	10.0	0.7	19.6	6.8	1.8	0.4	0.0	0.0	100
400	27.4	17.9	9.1	1.6	19.4	8.7	5.6	0.8	4.8	4.8	100
500	32.5	17.9	7.9	0.0	23.4	7.1	3.6	2.8	2.4	2.4	100
600	33.9	17.9	9.3	0.4	19.3	7.1	3.9	1.8	3.2	3.2	100
700	46.0	17.9	6.7	0.8	19.0	4.4	1.6	2.0	0.8	0.8	100
800	59.9	17.9	7.5	1.2	9.1	0.8	1.2	0.8	0.8	0.8	100
900	35.0	17.9	12.1	1.1	23.9	3.2	2.5	0.7	1.8	1.8	100
1000	31.3	17.9	9.9	0.0	23.0	6.7	4.4	1.2	2.8	2.8	100
AVERAGE	40.0	17.9	9.3	0.6	17.8	5.4	2.9	1.3	1.9	1.9	100

(b) Backup Paths under Single Link or Node Failure

# of destinations	1	1	1	1	2	2	3	2	3	4	Σ
# of paths	1	0	2	3	2	3	3	4	4	4	
Size	%										
100	48.2	17.9	5.7	0.4	17.1	4.3	5.0	0.0	0.7	0.7	100
200	36.1	17.9	6.7	0.4	22.6	4.8	6.0	0.8	2.4	2.4	100
300	37.5	17.9	6.4	0.7	22.1	3.9	6.4	0.7	2.1	2.1	100
400	39.6	17.9	4.3	0.7	22.1	4.3	6.4	1.1	1.8	1.8	100
500	36.8	17.9	6.8	0.4	24.6	5.4	6.8	0.7	0.4	0.4	100
600	36.4	17.9	5.4	0.0	25.4	5.0	7.9	0.0	1.1	1.1	100
700	32.1	17.9	6.4	0.7	22.9	7.1	6.1	1.8	2.5	2.5	100
800	33.9	17.9	5.7	0.0	25.0	8.2	5.7	0.7	1.4	1.4	100
900	34.3	17.9	7.5	0.0	24.6	2.9	5.7	0.7	3.2	3.2	100
1000	37.9	17.9	7.5	0.0	24.6	6.4	2.9	0.7	1.1	1.1	100
AVERAGE	37.3	17.9	6.2	0.3	23.1	5.2	5.9	0.7	1.7	1.7	100

(c) Backup Paths under Single Link or Node or Server Failure

# of destinations	1	1	1	1	2	2	3	2	3	4	Σ
# of paths	1	0	2	3	2	3	3	4	4	4	
Size	%										
100	82.5	0.0	10.7	0.4	5.7	0.7	0.0	0.0	0.0	0.0	100
200	79.6	0.0	13.6	0.7	5.7	0.4	0.0	0.0	0.0	0.0	100
300	73.6	0.0	12.9	1.1	10.4	2.1	0.0	0.0	0.0	0.0	100
400	75.4	0.0	12.1	0.7	7.5	3.2	0.4	0.0	0.4	0.4	100
500	72.9	0.0	13.6	0.7	10.7	1.4	0.4	0.4	0.0	0.0	100
600	76.1	0.0	13.9	0.7	8.6	0.0	0.4	0.4	0.0	0.0	100
700	74.6	0.0	16.8	0.0	7.1	1.4	0.0	0.0	0.0	0.0	100
800	77.0	0.0	13.5	0.0	7.1	1.6	0.4	0.4	0.0	0.0	100
900	77.1	0.0	13.9	0.4	7.1	1.4	0.0	0.0	0.0	0.0	100
1000	66.9	0.0	5.5	0.3	24.0	0.6	1.9	0.0	0.3	0.3	100
AVERAGE	75.6	0.0	12.7	0.5	9.4	1.3	0.3	0.1	0.1	0.1	100

row of each sub-table contains the average of each path pattern over the various sets of traffic instances (i.e., 100 instances).

As we can see in the different tables, if we exclude the basic path pattern with a single path, both for the working and the backup path patterns the largest percentages correspond to: (i) the number of working paths with two different paths toward an identical destination and (ii) the number of backup paths with two different paths toward different (two) destinations,

It is also observed that the single link and the single link/node failure scenarios do not differ much with respect to the number of bandwidth units (6%, see Figure 11) as well as with respect to the ratio of multiple paths (3%, see Tables 4 and 5). On the other hand, the single link/node/server failure scenario requires more bandwidth units (30%, see Figure 11) and has more multiple paths (6% more for working and 10% for backup, see again Tables 4 and 5).

Table 6: Some ILP gaps

request size	GAP II vs. I	GAP III vs. I
100	8.75%	9.85%
200	8.74%	8.35%
300	9.70%	8.96%
400	9.11%	8.47%
500	6.91%	7.63%
600	7.64%	8.72%
700	2.24%	4.34%
800	0.02%	1.93%
900	5.77%	8.50%
1,000	5.91%	7.03%
average	6.74	7.38

9 Conclusion and Future Work

We proposed three different decomposition models for the joint optimization of the location servers and of the provisioning of the working/backup paths within the any cast framework in an optical grid. The decomposition model with the most simple set of configurations, i.e., a pair of a working and a backup path associated with the same source node, proved to be the most scalable. In addition, it offers the largest flexibility with potentially multiple working or backup paths. Consequently, the most efficient decomposition differs from the one used in the p-center problem, which is the one extended in Model III. Future work will include the development of heuristics and the use of the stabilization techniques [31] in Model III to overcome the slow convergence, and potentially take advantage of the lowest LP bounds provided by Model III.

The best location of the servers is independent of the number of backup paths (i.e., we observed the same choices being made by Model I vs Models II–III), and offers a compromise between the distribution of the server locations and the availability of short paths and highly sharable backup paths. Allowing more than one backup path leads to a reduced overall bandwidth requirement (about 7%), but comes at the expense of a higher management complexity (to set up and switch to the correct backup path).

Appendix A Selection of the parameters for the ε-optimality conditions

In Figure 13, we did additional intensive testing for finding out the best compromise between the parameters of the stopping condition and the accuracy of the last output LP solution. Therein, we see that if we only require 30 instead of 50 iterations, the accuracy of the LP bound is definitively deteriorating.

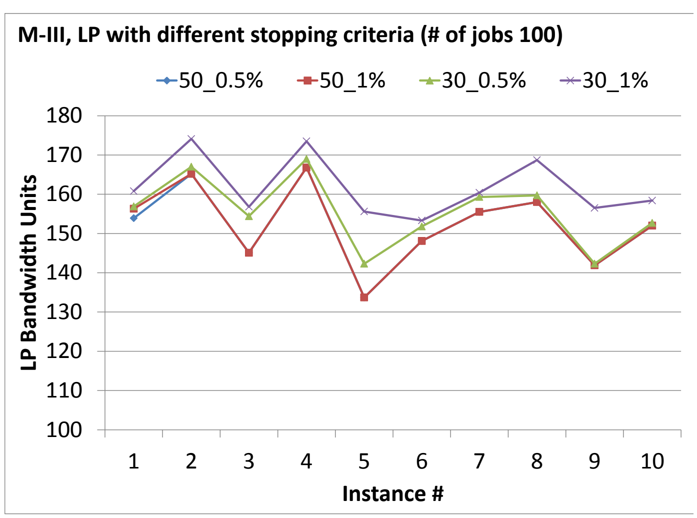


Figure 13: Model III – 100 requests – LP bound accuracy with different stopping criteria on Model III

References

[1] Cisco global cloud index: Forecast and methodology, 2010–2015, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns1175/ (2012).

[2] M. Hogan, F. Liu, A. Sokol, J. Tong, NIST cloud computing standards roadmap working roadmap, Tech. Rep. Spec. Publ. 500-291, NIST – National Institute of Standards and Technology, U.S. Department of Commerce, version 1.0, 83 pages (July 2011).

[3] R. Buyya, C. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility, Future Generation Computer Systems 25 (2009) 599–616.

[4] O. Gerstel, M. Jinno, A. Lord, S. B. Yoo, Elastic optical networking: A new dawn for the optical layer?, IEEE Journal of Communications Magazine (2012) 512–520.

[5] C. Develder, M. Leenheer, B. Dhoedt, M. Pickavet, D. Colle, F. Turck, P. Demeester, Optical networks for grid and cloud computing applications, in: Proceedings of the IEEE, Vol. 100, 2012, pp. 1149–1167.

[6] C. Develder, J. Buysse, B. Dhoedt, B. Jaumard, Joint dimensioning of server and network infrastructure for resilient optical grids/clouds, IEEE/ACM Transactions on Networking (2013) 1–16.

[7] B. Jaumard, A. Shaikh, Maximizing access to IT services on resilient optical grids, in: Workshop on Reliable Networks Design and Modeling – RNDM, 2011, pp. 1–6.

[8] A. Shaikh, B. Jaumard, Optimized dimensioning of resilient optical grids with respect to grade of it services, Telecommunication Systems (2012) to appear.

[9] B. Jaumard, J. Buysse, A. Shaikh, M. Leenheer, C. Develder, Column generation for dimensioning resilient optical grid networks exploiting relocation, in: IEEE Global Telecommunications Conference – GLOBECOM, 2010, pp. 1–5.

[10] A. Shaikh, J. Buysse, B. Jaumard, C. Develder, Anycast routing for survivable optical grids: Scalable solution methods and the impact of relocation, Journal of Optical Communications and Networking 3 (2011) 767–779.

[11] B. Jaumard, A. Shaikh, C. Develder, Selecting the best locations for data centers in resilient optical grid/cloud dimensioning (invited paper), in: Int. Conf. Transparent Optical Netw. (ICTON), Coventry, UK, 2012, pp. 1–4.

[12] L. Lorena, E. Senne, A column generation approach to capacitated p-median problems, Computers & Operations Research 31 (2004) 863–876.

[13] C. Develder, B. Mukherjee, B. Dhoedt, P. Demeester, On dimensioning optical grids and the impact of scheduling, Photonic Network Communications 17 (3) (2009) 255–265.

[14] J. Buysse, M. D. Leenheer, C. Develder, B. Dhoedt, Exploiting relocation to reduce network dimensions of resilient optical grids, in: Proceedings of IEEE/VDE Workshop on Design of Reliable Communication Networks – DRCN, 2009, pp. 100–106.

[15] C. Develder, J. Buysse, M. D. Leenheer, B. Jaumard, B. Dhoedt, Resilient network dimensioning for optical grid/clouds using relocation, in: IEEE International Conference on Communications – ICC, Ottawa, Ontario, Canada, 2012, pp. 1–5.

[16] E. Hyytiä, Heuristic algorithms for the generalized routing and wavelength assignment problem, in: Proc. 17th Nordic Teletraffic Seminar (NTS-17), Fornebu, Norway, 2004, pp. 373–386.

[17] K. Walkowiak, Anycasting in connection-oriented computer networks: Models, algorithms and results, Journal of Applied Mathematics and Computer Science 20 (1) (2010) 207–220.

[18] K. Walkowiak, J. Rak, Shared backup path protection for anycast and unicast flows using the node-link notation, in: IEEE International Conference on Communications – ICC, Kyoto, Japan, 2011, pp. 1–6.

[19] M. Melo, S. Nickel, F. S. da Gama, Facility location and supply chain management – A review, European Journal of Operational Research 196 (2) (2009) 401–412.

[20] S. Owen, M. Daskin, Strategic facility location: A review, European Journal of Operational Research 111 (3) (1998) 423–447.

[21] M. Daskin, What you should know about location modeling, Naval Research Logistics 55 (2008) 283–294.

[22] F. Larumbe, B. Sansò, Optimal location of data centers and software components in cloud computing network design, in: 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing – CCGrid, 2012, pp. 841–844.

[23] J. Chakareski, Joint optimization of flow allocation and data center placement in multi-service networks, in: 19th International Packet Video Workshop (PV), 2012, pp. 101–106.

[24] J. Reese, Solution methods for the p-median problem: An annotated bibliography, European Journal of Operational Research 48 (3) (2006) 125–142.

[25] W. Klibi, A. Martel, A. Guitoni, The design of robust value-creating supply chain networks: A critical review, European Journal of Operational Research 203 (2010) 283–293.

[26] C. Losada, M. Scaparra, J. O'Hanley, Optimizing system resilience: A facility protection model with recovery time, European Journal of Operational Research 217 (2010) 519–530.

[27] P. Cappanera, M. Scaparra, Optimal allocation of protective resources in shortest-path networks, Transportation Science 45 (2013) 64–80.

[28] IBM, IBM ILOG CPLEX 12.0 Optimization Studio (2011).

[29] S. D. Maesschalck, D. Colle, I. Lievens, M. Pickavet, P. Demeester, C. Mauz, M. Jaeger, R. Inkret, B. Mikac, J. Derkacz, Pan-European optical transport networks: An availability-based comparison, Photonic Network Communications 4 (2003) 203–225.

[30] D. Chen, R. Chen, New relaxation-based algorithms for the optimal solution of the continuous and discrete p-center problems, Computers & Operations Research 36 (5) (2009) 1646–1655.

[31] O. du Merle, D. Villeneuve, J. Desrosiers, P. Hansen, Stabilized Column Generation, Discrete Mathematics 194 (1) (1999) 229–248.