Anycast End-to-end Resilience for Cloud Services over Virtual Optical Networks

Minh Bui*, Brigitte Jaumard*, Chris Develder†
*CSE, Concordia University, Montreal (Qc) H3G 1M8 Canada
†INTEC - IBCN, Ghent University - iMinds, Ghent, Belgium

ABSTRACT

Optical networks are crucial to support increasingly demanding cloud services. Delivering the requested quality of service is key to successfully provisioning end-to-end services in clouds. Therefore, as for traditional optical network services, it is of utter importance to guarantee that clouds are resilient to any failure of either network infrastructure or data centers. A crucial concept in establishing cloud services is that of network virtualization: the physical infrastructure is logically partitioned in separate virtual networks. Also, combined control of the network and data center (IT) resources is exploited. To guarantee end-to-end resilience for cloud services in such a set-up, we need to simultaneously route the services and map the virtual network, while ensuring that an alternate routing is always available. Note that the anycast routing concept applies: assigning server resources requested by the customer to a particular (physical) data center can be done transparently. This paper investigates the design of scalable optimization models to perform the virtual network mapping resiliently (for single bidirectional link failures), thus supporting resilient anycast cloud virtual networks. We compare two resilience approaches: PIP-resilience maps each virtual link to two alternate physical routes, VNO-resilience provides alternate paths in the virtual topology (while enforcing physical link disjointness).

Keywords: Network Virtualization, End-to-End Resilience, Cloud Computing, Anycast Resilience.

1. INTRODUCTION

The evolution towards grid and cloud computing as observed for over a decennium illustrates the crucial role played by (optical) networks in supporting today's applications [1]. A core concept in cloud computing is that of virtualization: an extra layer of abstraction is provided, such that the same physical infrastructure can be simultaneously used by distinct entities, each running their own applications in a virtually isolated environment (i.e., a virtual machine). This allows more efficient use of the physical infrastructure, as well as flexible extension of capacity by adding more virtual machines (and distributing them among multiple physical machines). The same idea of virtualization is also applied in the networking domain [2]: physical infrastructure (i.e., fibers and optical cross-connects, OXCs) can be shared by multiple virtual network operators (VNOs), who only see their own resources in a virtual topology, and have full control over it. Combining both network and server virtualization in the optical cloud calls for appropriate joint provisioning mechanisms allocating both network and IT resources [3].

In this paper, we study two approaches to resiliently provision virtual networks (VNets) for cloud services. We consider the physical network and data center resources to be owned and operated by physical infrastructure providers (PIPs); note that the PIP for data center resources possibly is to be a different entity than the PIP for the optical network. The cloud services will be offered by a virtual network operator (VNO), who will run its VNet on top of the PIP resources. The problem we address is how to determine the VNet topology and its mapping to the physical infrastructure. We will consider two fundamental alternatives for realizing the VNet instantiation resiliently: the first option is to map a virtual network link resiliently by providing two alternate routes in the PIP underlay (PIP-resilience), the second is to take care of rerouting in the virtual network under control of the VNO (VNO-resilience). Since the latter offers more flexibility to the VNO, who also has a full picture of all the service requests it has to support, it could be that this approach can be implemented more efficiently in terms of resource capacity requirements. Yet, the VNO-resilience puts a bigger burden on the configuration of the VNet (i.e., it adds operational/management complexity) and requires that the VNO has insight in the physical network or at least is aware of common resource dependencies between (and hence potential joint failures of), e.g., virtual network links. In this paper however, we assess the potential advantage in network capacity cost that such a VNO-resilience approach might entail compared to the PIP-resilience scheme.

The contributions of this paper are to: *(i)* develop scalable methods for VNet planning for cloud services, resilient against physical network (single link) failures, and *(ii)* assess the physical network resource cost differences between PIP-resilience and VNO-resilience. Note that the models detailed in the current paper will focus in particular on single bidirectional link failures and data center failures, but can rather straightforwardly be extended to more general failure cases.

The remainder of this paper is structured as follows: in Section 2, we summarize an overview of related work. Next, Section 3 formally introduces the accurate problem statements, and Section 4 discusses the models we propose to solve both resilience models. A case study with numerical results is presented in Section 5.

2. RELATED WORK

The problem of dimensioning optical clouds/grids basically involves finding the amount of resources (network and servers), to meet a set of given cloud services (i.e., traffic requests). The main complication herein stems from the anycast principle: in a cloud scenario, we have a certain flexibility in choosing an appropriate data center among a given set of possible locations to serve the cloud traffic. Thus, the classical notion of a (source,destination)-based traffic matrix disappears [4]. We previously developed scalable methods, based on the column generation technique to solve the resilient dimensioning problem: finding working and backup paths for a set of requests as to always be able to reach an operational data center location [5], even including the sizing of the data center capacity [6]. However, this previous work did not consider any resource to accommodate synchronization between distinct working and backup data center locations (as opposed to the current paper).

In the current paper, we address VNet planning. This builds on the work of Barla *et al.* who in [7] discuss the VNet planning problem and explain the two major options of addressing it while providing resilience, using mixed integer linear programming (MILP), focusing on delay minimization. The same authors also consider resource cost optimization in [8], yet resources for synchronization between backup and working data center (DC) locations are not accounted for. Furthermore, those authors also point out that other work treats optimization of *(i)* routing cloud service requests and *(ii)* mapping a VNet to the physical infrastructure separately. (For example, [9, 10] offer solutions for survivable VNet embedding, but consider that the VNet is already designed and given.) Work on optimal server selection and routing of anycast services in the physical layer for intra- and inter-DC networks [11, 12] do not consider resilient network design in the virtual layer.

The current paper explicitly addresses solving the VNet design and mapping problem with simultaneous routing of the requests, where we also account for synchronization connectivity (and associated bandwidth) between alternate data centers.

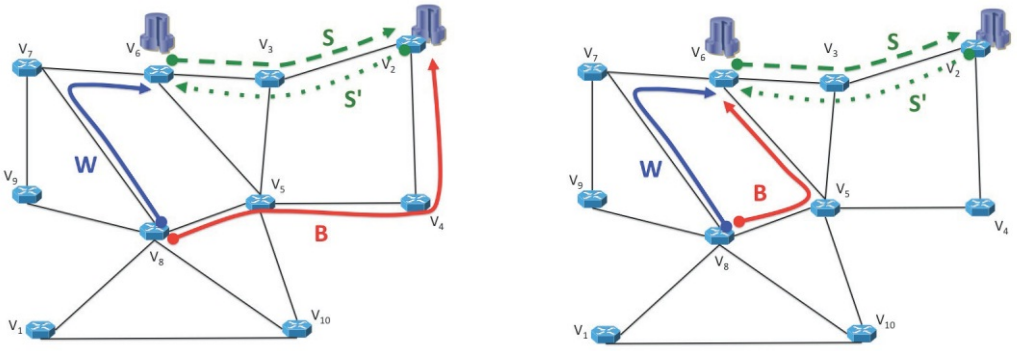
3. PROBLEM STATEMENT

The cloud network is described by 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 (v_1, v_2) has two opposite directed links ($v_1 \to v_2$ and $v_2 \to v_1$). Let $V_{\mathrm{D}} \subseteq V$ denote the set of nodes hosting a data center with $n_{\mathrm{D}} = |V_{\mathrm{D}}|$, i.e., the number of data center nodes. Each link ℓ has a transport capacity W_ℓ, i.e., limit on the number of wavelengths or ports at the endpoints of the link. 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 demands (services), originating from a set of source nodes $V_{\mathrm{S}} \subseteq V$.Let K be the set of services, indexed by k. Each service k is characterized by *(i)* v_k, its source (origin), and *(ii)* Δ_k, its bandwidth requirement. Each service request is ensured by anycast routing, with a primary and a backup data center. Requests originating at a node hosting a data center are assumed to be served by that particular data center.

We propose the design of a resilient cloud network that can survive single link failures, as well as single data center failures. In order to do so, we assume some synchronization paths between the primary and the backup data centers. Therefore, for each service, we assume one third characteristic, δ_k, the fraction of Δ_k that is required for synchronization and a soft migration between the primary and the backup data centers.

We investigate and compare two protection schemes. The first one, called VNO-resilience, see Fig. 1(a), assumes that VNO designs a resilient virtual network against single link and single DC failures. The second one, PIP-resilience, see Fig. 1(b), is such that the virtual network is designed without any protection against single DC failures, and in case of a DC failure, services are redirected to a second data center using a soft migration thanks to the synchronization path. Upon a single link failure on the working path, the traffic is switched on to the backup path in both protection schemes, where the service(s) is (are) still handled by the primary data center in the PIP-resilience scheme. Upon a failure of the primary data center, a soft migration is performed from the primary to the backup data center in both protection schemes, with the service(s) routed on the backup path in the first protection scheme, but on the primary path in the second protection scheme.



(a) Scheme 1: VNO-resilience

(b) Scheme 2: PIP-resilience

Fig. 1: Two Resilience Schemes

4. MODELS

In order to investigate and compare the two resilience schemes, we propose two large scale optimization models that adopt a column generation (CG) formulation. Such a formulation relies on a decomposition of the original problem into the so-called master and pricing problems, and on a set of configurations. Each configuration consists in a set of four paths for a given service node: a working and a backup path both originating at the same service source node, and two synchronization paths, one in each direction, between the primary and the backup data centers, as well as the set of services protected by the set of four paths. Note that two services originating at the same source node may be protected by two different configurations.

The master and pricing problems are alternately solved as illustrated in Fig. 2, with the master problem selecting the best subset of generated configurations (at least one for each service source node), while the pricing problem generates promising configurations. The column generation technique offers a solution scheme for the linear relaxation of the optimization model, such that, at each iteration, the linear relaxation of the master problem is solved and its optimal dual values are transferred to the pricing problem, which, in turn, generates a new "improving" configuration. The latter is a configuration such that, if added to the current set of configurations, allows an improvement of the incumbent value of the objective function of the master problem (see, e.g., [13] for more details on column generation techniques). This iterative process stops when no improving configurations can be generated, meaning that we have found the optimal linear programming (LP) solution. We then generate an integer solution, by solving the linear integer program made of the configurations generated in order to reach the optimal solution of the linear relaxation.

Next, we describe the column generation formulation for both proposed resilience schemes.

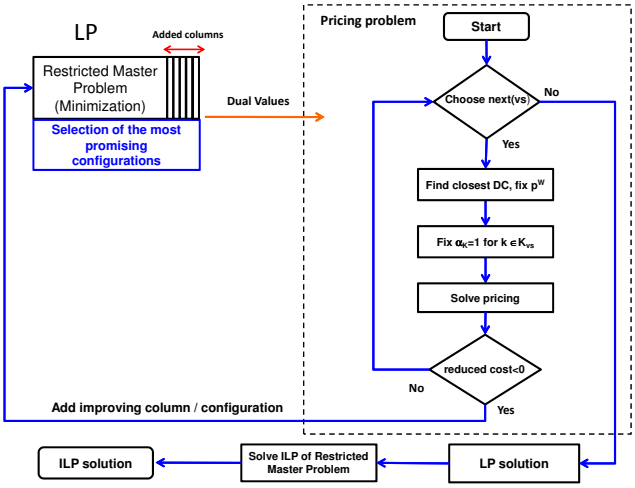


Fig. 2: Decomposition flow chart

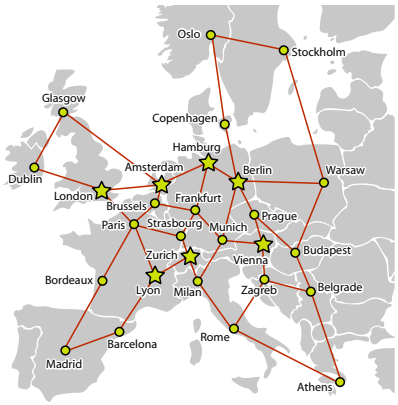


Fig. 3: NobelEU network with all possible data center (DC) locations indicated with a star symbol. Only four DCs are actually chosen in each experiment.

4.1. Master Problem of the PIP Resilience Model

Let C be the overall set of configurations: $C = \bigcup_{v \in V_S} C_v$, where C_v is the set of configurations associated with source node $v \in V_S$. We define a configuration $c \in C_v$ by: *(i)* a set of four paths, one primary path ($p^{W,c}$) originating at v_s toward a primary data center DC^W, one backup path ($p^{B,c}$) originating at v_s toward a primary data center DC^B, and two synchronization paths ($p^{S,c}$ from DC^W to DC^B) and ($p^{S',c}$ from DC^B to DC^W) between the primary and the backup data centers, as well as *(ii)* the service request routed and protected by this set of four routes.

More formally, a configuration is characterized by:

$p_\ell^{W,c}$ $= 1$ if link ℓ is used by working path $p^{W,c}$ of configuration c, 0 otherwise.

$p_\ell^{B,c}$ $= 1$ if link ℓ is used by backup path $p^{B,c}$ of configuration c, 0 otherwise.

$p_\ell^{S,c}$ $= 1$ if link ℓ is used by synchronization path $p^{S,c}$ of c from the primary data center to the backup data center, 0 otherwise.

$p_\ell^{S',c}$ $= 1$ if link ℓ is used by synchronization path $p^{S',c}$ of c from the backup data center to the primary data center, 0 otherwise.

α_k^c $= 1$ if service k is routed and protected by configuration c, 0 otherwise.

In addition, we have:

$b_\ell^{W,c}$: working bandwidth requirement on ℓ, with ℓ being a link of $p^{W,c}$,

$b_\ell^{S,c}$: primary-to-backup synchronization bandwidth requirement on ℓ, with ℓ being a link of $p^{S,c}$,

which both only depend on the services associated with c. Note that these bandwidths are not shared.

For each link ℓ, let β_ℓ^W and β_ℓ^B be the working and backup bandwidth on ℓ, respectively.

Let $\mathcal{F}$ be the set of failure sets, indexed by F, where F contains links that fail at the same time: for the single physical link failures that we will consider, F comprises the two opposite directed links that connect the same two nodes.

The objective function is to minimize the overall (working, backup, including synchronization) bandwidth requirements:

$$\min \quad \sum_{\ell \in L} \left(\beta_\ell^{\mathrm{W}} \; + \; \beta_\ell^{\mathrm{B}}\right) \tag{1}$$

subject to:

$$\beta_\ell^{\mathrm{W}} + \beta_\ell^{\mathrm{B}} \leq W_\ell \qquad \ell \in L \tag{2}$$

$$\beta_\ell^{\mathrm{W}} = \sum_{c \in C} \left(b_\ell^{\mathrm{W},c} + b_\ell^{\mathrm{S},c}\right) z_c \qquad \ell \in L \tag{3}$$

$$\sum_{c \in C} \alpha_k^c \, z_c \geq 1 \qquad k \in K \tag{4}$$

$$\sum_{v \in V} \sum_{c \in C_v} \left(b_\ell^{\mathrm{W},c} \, p_{\ell'}^{\mathrm{B},c} + \sum_{k \in K_v} \Delta_k \, \alpha_k^c \, p_\ell^{\mathrm{W},c} \delta_k p_{\ell'}^{\mathrm{S}',c} \right) z_c \; \leq \beta_{\ell'}^{\mathrm{B}} \qquad \ell \in F, F \in \mathcal{F}, \ell' \in L \setminus F \tag{5}$$

$$z_c \in \{0,1\} \qquad c \in C \tag{6}$$

$$\beta_\ell^{\mathrm{W}} \in \mathbb{R} \qquad \ell \in L \tag{7}$$

Constraints (2) are capacity constraints on each link of the optical grid. Constraints (3) compute the overall working (regular + synchronization) bandwidth requirements on link ℓ. Constraints (4) are the demand constraints, and ensure that each service k is granted. Constraints (5) compute the overall backup (regular + synchronization) bandwidth requirements on link ℓ'.

4.2. Pricing Problems of the PIP Resilience Model

There is one pricing problem per source node, and the role of each pricing problem is to generate a set of four paths $(p^{\mathrm{W}}, p^{\mathrm{B}}, p^{\mathrm{S}}, p^{\mathrm{S}'})$ in order to route, provision and protect all or a subset of the services originating at that source node.

The objective of the pricing problem associated with a given source node corresponds to the reduced cost (see, e.g., Chvatal [13] if not familiar with linear programming):

$$\begin{aligned}
\overline{\mathrm{COST}} &= \sum_{\ell \in L} u_\ell^{(3)} \; (b_\ell^{\mathrm{W}} + b_\ell^{\mathrm{S}}) - \sum_{k \in K} u_k^{(4)} \, \alpha_k - \sum_{F \in \mathcal{F}} \sum_{\ell \in F} \sum_{\ell' \in L : \ell \neq \ell'} \sum_{k \in K_{v_{\mathrm{S}}}} u_{\ell\ell' F}^{(5)} \, \Delta_k \, \alpha_k \, p_\ell^{\mathrm{W}} \, \left(p_{\ell'}^{\mathrm{B}} + \delta_k \, p_{\ell'}^{\mathrm{S}'}\right) \\
&= \sum_{\ell \in L} u_\ell^{(3)} \; (b_\ell^{\mathrm{W}} + b_\ell^{\mathrm{S}}) - \sum_{k \in K} u_k^{(4)} \, \alpha_k \quad - \sum_{F \in \mathcal{F}} \sum_{\ell \in F} \sum_{\ell' \in L : \ell \neq \ell'} u_{\ell\ell' F}^{(5)} \, b_\ell^{\mathrm{W}} \, p_{\ell'}^{\mathrm{B}} \\
&\qquad - \sum_{F \in \mathcal{F}} \sum_{\ell \in F} \sum_{\ell' \in L : \ell \neq \ell'} \sum_{k \in K_{v_{\mathrm{S}}}} u_{\ell\ell' F}^{(5)} \, \Delta_k \, \delta_k \, \alpha_k \, p_\ell^{\mathrm{W}} \, p_{\ell'}^{\mathrm{S}'}
\end{aligned}$$

where $u^{(3)}$, $u_v^{(4)}$, $u_{\ell\ell' F}^{(5)}$ are the values of the dual variables associated with constraints (3), (4), (5) respectively.

We observe that the reduced cost contains cubic and quadratic terms. Although linearization is possible, it is very costly in terms of additional variables and constraints. Therefore, we propose to pre-select the working path as the shortest path to the closest data center. By doing so, the reduced cost is simplified and becomes linear.

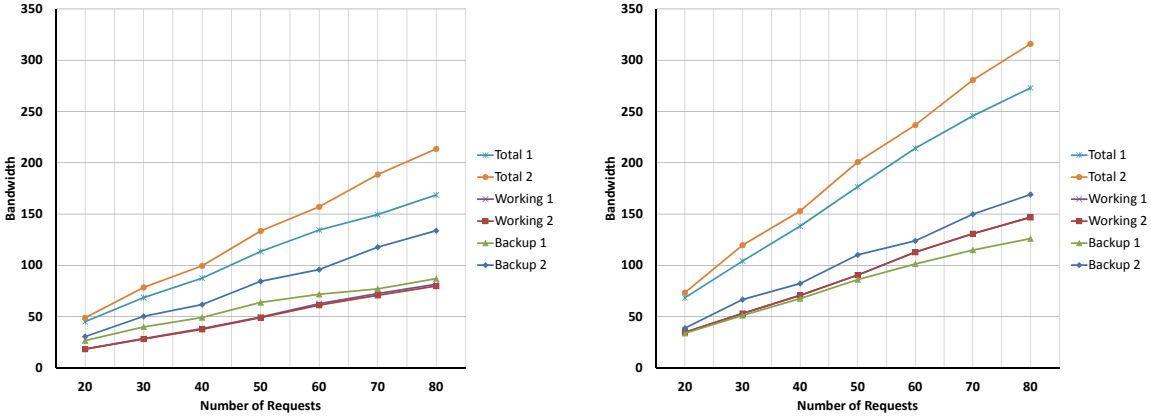
We here omit the expression of the constraints of the pricing problems as they correspond to classical flow constraints in order to identify the remaining paths, with some link disjointness conditions for the working and backup paths.

4.3. VNO Resilience Model

The master problem of the VNO-resilience model is identical to the one of the PIP-resilience model. Differences arise in the path configurations, i.e., in the destination of the backup path, while link disjointness constraints are identical.

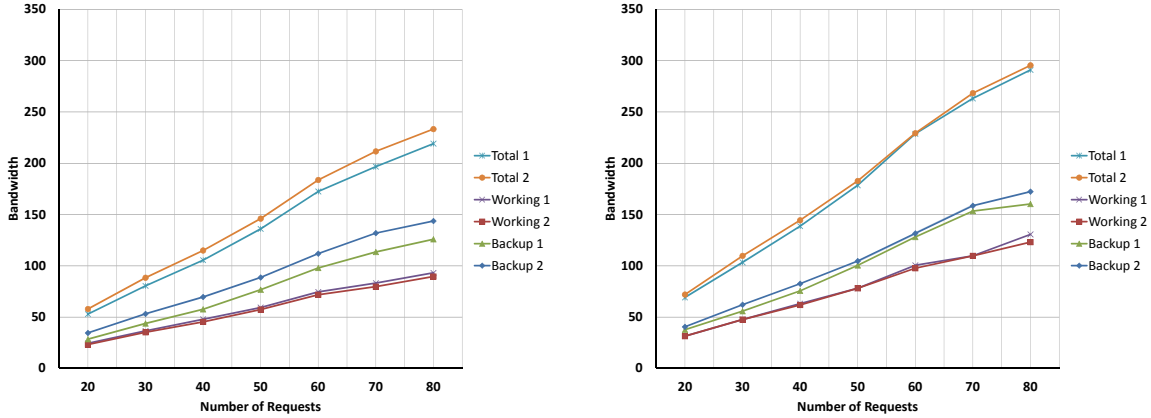
5. NUMERICAL RESULTS

We conduct our experiments on NobelEU network which has 28 nodes, 82 directed links (see Fig. 3), and uncapacitated links (i.e., we omit constraints (2)). We randomly generated between 20 and 80 requests, with bandwidth per request in $[0.1, 0.9]$, and assumed that all requests originating from the same source will adopt the same routing (i.e., use the same configuration in our previously explained model). To study the effect of the data center (DC) locations on the bandwidth requirements, we ran experiments for two choices of DC locations. The first experiment has DCs in Lyon, Berlin, London, and Vienna, thus scattered rather evenly across the central network nodes. The second experiment has two pairs of neighboring DC locations: Amsterdam, Hamburg, Lyon, and Zurich. For both, we compare the performance of the two protection schemes (VNO-resilience vs PIP-resilience), for two parameter choices for the relative synchronisation bandwidth: $\delta_k = 0.1$ and $\delta_k = 0.9$. The LP/ILP programs have been implemented in OPL and solved using IBM ILOG CPLEX 12.2, running on a 4-core 2.2 GHz AMD Opteron 64-bit processor.



(a) $\delta_k = 0.1$ (b) $\delta_k = 0.9$

Fig. 4: Experiment 1: DCs in Lyon, Berlin, London, and Vienna. Model 1: VNO-resilience, Model 2: PIP-resilience.



(a) $\delta_k = 0.1$ (b) $\delta_k = 0.9$

Fig. 5: Experiment 2: DCs in Amsterdam, Hamburg, Lyon, and Zurich. Model 1: VNO-resilience, Model 2: PIP-resilience.

Comparing the protection schemes (recall Fig. 1), we intuitively expect VNO-resilience to perform better, since the backup paths are expected to be shorter on average (this is the benefit of exploiting relocation, as shown previously in other settings [6]), at least if we momentarily disregard synchronisation bandwidth (the S and S' paths from Fig. 1). Thus, for low values of δ_k, we expect a clear relative advantage of VNO-resilience (model 1) over PIP-resilience (model 2). This is confirmed by the results plotted in Fig. 4. The relative advantage diminishes for increasing δ_k value.

Reflecting on the previous statement, and looking at the results for the second experiment in Fig. 5, we observe that the difference between PIP-resilience and VNO-resilience almost disappears. Thus, the capacity advantage that VNO-resilience might theoretically bring, seems to disappear if data centers are not distributed reasonably uniformly across the topology, but rather occur in pairs of nearby nodes. Indeed, in that case the difference in lengths of backup paths to either the same, or a neigboring DC (compared to that of the working), becomes negligible, especially for large δ.

ACKNOWLEDGMENT

B. Jaumard was supported by NSERC (Natural Sciences and Engineering Research Council of Canada) and by a Concordia University Research Chair (Tier I) on the Optimization of Communication Networks.

REFERENCES

[1] C. Develder, M. Leenheer, B. Dhoedt, M. Pickavet, D. Colle, F. Turck, and P. Demeester, "Optical networks for grid and cloud computing applications," *Proceedings of the IEEE*, vol. 100, pp. 1149–1167, May 2012.

[2] N. M. K. Chowdhury and R. Boutaba, "A survey of network virtualization," *Comput. Netw.*, vol. 54, p. 862876, Apr. 2010.

[3] P. Vicat-Blanc and *et al.*, "Bringing optical networks to the cloud: an architecture for a sustainable future internet," in *The Future Internet* (J. Domingue and *et al.*, eds.), vol. 6656/2011 of *LCNS*, pp. 307–320, Springer, 2011.

[4] C. Develder, B. Mukherjee, B. Dhoedt, and P. Demeester, "On dimensioning optical grids and the impact of scheduling," *Photonic Netw. Commun.*, vol. 17, pp. 255–265, Jun. 2009.

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

[6] C. Develder, J. Buysse, M. De Leenheer, B. Jaumard, and B. Dhoedt, "Resilient network dimensioning for optical grid/clouds using relocation (invited paper)," in *Proc. Workshop on New Trends in Optical Networks Survivability, at IEEE Int. Conf. on Commun. (ICC 2012)*, (Ottawa, Ontario, Canada), 11 Jun. 2012.

[7] I. B. Barla, D. A. Schupke, and G. Carle, "Resilient virtual network design for end-to-end cloud services," in *Proc. 11th Int. Conf. Networking (Networking 2012)*, (Prague, Czech Republic), pp. 161–174, Springer-Verlag, May 2012.

[8] I. B. Barla, D. A. ScSchupke, M. Hoffmann, and G. Carle, "Optimal design of virtual networks for resilient cloud services," in *Proc. 9th Int. Conf. Design of Reliable Commun. Netw. (DRCN 2013)*, (Budapest, Hungary), 4–7 Mar. 2013.

[9] K. Lee and E. Modiano, "Cross-layer survivability in WDM-Based networks," in *Proc. 28th IEEE Conf. Computer Commun. (INFOCOM 2009)*, (Rio de Janeiro, Brazil), pp. 1017–1025, Apr. 2009.

[10] H. Yu, C. Qiao, V. Anand, X. Liu, H. Di, and G. Sun, "Survivable virtual infrastructure mapping in a federated computing and networking system under single regional failures," in *Proc. IEEE Global Telecommun. Conf. (Globecom 2010)*, (Miami, FL, USA), pp. 1–6, Dec. 2010.

[11] J. Jiang, T. Lan, S. Ha, M. Chen, and M. Chiang, "Joint VM placement and routing for data center traffic engineering," in *Proc. 31th IEEE Conf. Computer Commun. (INFOCOM 2012)*, (Orlando, FL, USA), pp. 2876–2880, Mar. 2012.

[12] M. Alicherry and T. V. Lakshman, "Network aware resource allocation in distributed clouds," in *Proc. 31th IEEE Conf. Computer Commun. (INFOCOM 2012)Proc. 31th IEEE Conf. Computer Commun. (INFOCOM 2012)*, (Orlando, FL, USA), pp. 963–971, Mar. 2012.

[13] V. Chvatal, *Linear Programming*. Freeman, 1983.