Resilience options for provisioning anycast cloud services with Virtual Optical Networks

Minh Bui, Brigitte Jaumard
Department of Computer Science and Software Engineering
Concordia University
Montreal, Canada
Email: {ng_bui,bjaumard}@cse.concordia.ca

Chris Develder
INTEC - IBCN
Ghent University - iMinds
Ghent, Belgium
Email: chris.develder@intec.ugent.be

***Abstract*—Optical networks are crucial to support increasingly demanding cloud services. Delivering the requested quality of services (in particular latency) 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 (links and/or nodes) 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. 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, in such a way that an alternate routing in case of physical resource failures is always available. Note that combined control of the network and data center resources is exploited, and the anycast routing concept applies: we can choose the data center to provide server resources requested by the customer to optimize resource usage and/or resiliency. This paper investigates the design of scalable optimization models to perform the virtual network mapping resiliently. We compare various resilience options, and analyze their compromise between bandwidth requirements and resiliency quality.**

I. Introduction

Today, cloud computing plays a crucial role in cost-efficiently supporting almost any application domain, an evolution which heavily relies on the advances in (optical) networking [1]. A core concept in the cloud domain, and one that has recently also been applied in the networking field itself, is that of virtualization. This boils down to providing an extra level of abstraction, such that the same underlying physical infrastructure can be used by different entities, each in a virtually isolated environment (e.g., a virtual machine in a data center). Similarly, physical networking infrastructure (i.e., fibers and switching equipment) can thus be shared by various *virtual network operators (VNOs)* [2]. The logical partition under the control of the VNO amounts to a virtual network topology, denoted as virtual network (VNet), operated in isolation from other VNOs. The physical network and data center infrastructures are then managed by typically different entities, the *physical infrastructure providers (PIPs)*. (In practice VNOs and PIPs could indeed be different companies.)

In this paper, we focus on the planning of the core network, in terms of the backbone network (e.g., the wavelength routing and OXCs) as well as allocation of server capacity at data centers. Both the assumed optical network and data centers are assumed to be virtualized, i.e., they will be partitioned into VNets: we consider a physical infrastructure (offered by a PIP) that will be shared to carry services offered by multiple VNOs. In particular, we will study how to resiliently provision VNets for cloud services: requests to be served by a VNO need to be allocated server capacity at a certain data center (DC) – whose physical location, i.e., mapping to a particular PIP's infrastructure, can be decided by the VNO – and obviously network connectivity from the VNO's customer to their assigned DC(s). We focus on a planning problem addressing multiple VNets simultaneously. In this paper, we propose new models for end-to-end cloud services with different quality in terms of recovery times and availabilities, under both network and DC failures. Our contributions are:

- Compared to earlier work by Barla *et al.* [3]–[5] (see Section II), our resilience approach explicitly includes the required network connectivity and associated bandwidth between a primary and backup data center.
- We introduce a comprehensive qualitative overview of the various resilience options in choosing the aforementioned synchronization path (beyond the single simple choice adopted in our initial short paper [6] on this topic).
- We provide full model details for four resilience approaches (not covered in [6]), and a large scale case study (beyond the small problem instances covered by e.g., Barla *et al.* [3]) for two of them on a US topology.

The remainder of this paper is structured as follows: Section II outlines related work. The two fundamental resilience strategies (VNO-resilience and PIP-resilience) are discussed in Section III, including details the various choices in the quality of the protection. The models, adopting a column generation approach, are detailed in the subsequent Section IV. Our case study results are presented in Section V, and we conclude in the final Section VI.

II. Related work

The focus of this work is the joint planning of multiple VNets, as introduced by Barla *et al.* in [3], which explains the two major resilience strategies (VNO- vs PIP-resilience) and focuses on delay minimization. Optimization of resource cost is treated by the same authors in [5], but there they do not account for resources for synchronization between

primary and secondary data centers (DC). 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. In the problem of survivable VNet embedding, [7] and [8] consider that the VNet is already designed and given, while in [9], [10], the authors build the most bandwidth efficient resilient VNet, under unicast traffic assumptions and using either single or multiple hop routing of requests in the virtual network. In proposing solutions for optimal server selection, as well as physical layer routing of anycast services for intra- and inter-DC networks, the resilience of the resulting virtual layer design is not considered by [11], [12]. It is important to note that we deal with a planning problem, jointly deciding on multiple VNets, and not an online VNet mapping that maps one VNet at a time (as in, e.g., [13]).

The current paper explicitly addresses solving the resilient VNet design and mapping problem with simultaneous routing of the requests. This is undeniably related to the general problem of dimensioning optical clouds/grids: how to find the (minimal) amount of network and DC resources, to meet a set of given cloud service requests? A major complexity arises from the anycast principle: we have flexibility in choosing a DC among a given set of possible locations. Hence, the classical notion of a (source,destination)-based traffic matrix disappears [14]. While we previously developed scalable methods solve the resilient anycast dimensioning problem [15]–[17], that work did not consider synchronization between distinct working and backup data center locations (as opposed to the current paper). We believe this is the first work to discuss this in depth: previously we only sketched initial ideas in [6].

III. VNO vs PIP Resilience

Cloud service requests that we consider a VNO to support, are assumed to have a given origin s (i.e., the location of customer of the VNO), and need to be served at a data center d (where server capacity should be allocated) and requires network connectivity between the (s, d) pair. Assuming anycast, d can be chosen out of a set of given locations (i.e., where the VNO can rely on a PIP's infrastructure). We will design the VNet such that requests can survive single failures, which can each affect either the physical network or data center infrastructure. We will now discuss the two fundamental options in doing so: VNO-resilience and PIP-resilience. They are illustrated in Fig. 1, where both approaches rely on two disjoint DCs ($d1$ and $d2$) to protect against data center failures. Further, we assume there is an automatic switch-back to the original network path or DC once a fault is repaired, and therefore will allow reusing the same network/DC capacity to protect against other failures: backup capacity is shared.

In the VNO-resilience model, 1:1 protection routing is provided in the VNet for network failures, where the working and protection paths of a service have to be physically link/node disjoint: the working path W routes the services towards the primary DC, the protection path B towards the backup DC, and W and B will be disjoint in their physical layer mapping. In addition, one (or two, see further, Section III-C) synchronization paths S are established in order to handle migration and failure routing requirements when a DC failure occurs: services then need to be rerouted from the primary $d1$ to the backup DC $d2$. Thus, the resulting VNet for the request from source s comprises three links, mapped to resp. the physical W, B and S paths. Note that both W and B need to carry the full traffic (but B only when W or $d1$ are affected by a failure), but S possibly only a fraction thereof, only to keep the state at the backup location $d2$ synchronized with that of $d1$ to allow smooth handover upon $d1$ failure.

In PIP-resilience, services are routed on single paths in the VNet layer, where each virtual link is mapped on two link/node disjoint physical paths in the physical layer. Thus, there will be a single virtual link connecting the source s to the primary data center $d1$, which in the physical layer will be supported by the two disjoint paths W and B. In addition, to cater for DC failures, a second location $d2$ will be chosen, and connectivity between $d1$ and $d2$ will be provided along the physical path S. Thus, the VNet will comprise only two virtual links. In terms of capacity, it is important to note that under PIP-resilience the S path needs to carry not just synchronization traffic but also the full traffic bandwidth (hence the additional red line in Fig. 1) in case of $d1$ failure.

From the discussion above, it is clear that the W and B paths need to be disjoint (for both VNO- and PIP-resilience). However, depending on the recovery time requirements, we can have different disjointness requirements for S or even choose to have two disjoint synchronization paths S and S′, as argued below. For the sake of clarity, we will discuss in detail the various failure scenarios and how they are dealt with in the two fundamental resilience schemes.

A. VNO-resilience

Let us first consider a single link failure, say of link ℓ, and then the single DC failure:

(i) If $\ell \in$ W fails, then the request will be rerouted to the backup data center $d2$, using the backup path B (which is disjoint from W, thus $\ell \notin$ B). If it happens that $\ell \in$ S $\cap$ W, then it means that as long as the failure is not restored, the primary data center $d1$ can not be kept in sync with the now operational $d2$. Thus, right after the repair of ℓ, the primary

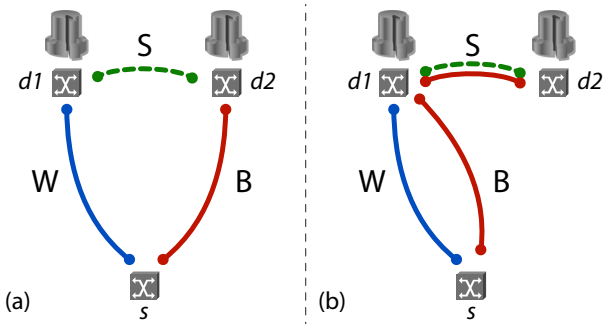

Fig. 1. Two protection schemes: (a) VNO-resilience, (b) PIP-resilience.

$d1$ will have stale state, and hence switching back to $d1$ will either suffer from this stale state or need to wait some extra time to receive the requests again. The remedy is of course to enforce $\text{W} \cap \text{S} = \emptyset$. (Yet, note that the same problem of a non-synchronized primary $d1$ clearly also occurs after the repair of a $d1$ that failed itself.)

(ii) If $\ell \in \text{S} \setminus \text{W}$ fails, this does not immediately pose a problem. Yet, if shortly after ℓ's repair the working path W fails, the switchover to the backup $d2$ (via path B) will suffer from stale state since the failing S will have interrupted the synchronization between primary and backup DCs. This can only be remedied by providing a second synchronization path S$'$ disjoint from S.

(iii) If $\ell \in \text{B}$ fails, again no immediate problem arises (since this means that W will be operational, given $\text{W} \cap \text{B} = \emptyset$). However, if $\ell \in \text{S} \cap \text{B}$ and shortly after ℓ's repair the primary path W (or $d1$) fails – meaning that now B will be followed towards $d2$ – the secondary data center $d2$ might not be fully sync'ed yet. Clearly, this can be remedied by choosing $\text{B} \cap \text{S} = \emptyset$. Yet, essentially the problem is exactly the same as for case (ii), which obviously remains, even if we take $\text{S} \cap \text{B} = \emptyset$.

(iv) If the primary DC at $d1$ fails, the requests will be rerouted to the backup $d2$ via the B path. Clearly, the failing $d1$ cannot be kept in sync with the now operational backup $d2$. Thus, we might need to wait some time after $d1$'s repair to switch back requests via W. Any failure that would occur shortly after $d1$'s repair and which would prevent services to remain being served at $d2$ clearly could imply service degradation because of the unsync'ed $d1$: (1) failure of S, (2) failure of B, or (3) failure of $d2$. This can however not be remedied without extra DC resources or extra paths.

B. PIP-resilience

(i) If $\ell \in \text{W}$ fails, requests will keep being served at primary $d1$, but now follow the B path to get there. If $\ell \in \text{S} \cap \text{W}$, then it means the secondary DC $d2$ will not be synchronized as long as ℓ is not repaired: if ℓ's repair is followed closely by a subsequent failure of the primary DC $d1$, then $d2$ will not be fully sync'ed yet, potentially resulting in temporary service degradation. This can be easily remedied by choosing $\text{S} \cap \text{W} = \emptyset$.

(ii) If $\ell \in \text{S} \setminus \text{W}$ fails, it means that $d2$ is no longer reachable and remains unsynchronized. As in the VNO-resilience case, the only remedy is a second, disjoint, synchronization path S$'$.

(iii) If $\ell \in \text{B}$ fails, this poses no immediate problem. Yet, if $\ell \in \text{S} \cap \text{B}$, and shortly after ℓ's repair the primary data center $d1$ fails, the backup $d2$ will not be fully sync'ed yet. A possible remedy is choosing $\text{S} \cap \text{B} = \emptyset$, but again, the same problem still occurs under failure of S alone (case (ii)).

(iv) If the primary DC at $d1$ fails, traffic is deflected to $d2$ (using the W + S route). Obviously, during its failure, $d1$ remains unsynchronized with the now operational $d2$. This means we might have to wait for this synchronization to be completed (via S) before switching back to a repaired $d1$. Clearly, a subsequent failure of S will obstruct that. This can be remedied by a second synchronization path S$'$, disjoint from S. Yet, as in the VNO case, the same problem of switch-back to a non-sync'ed $d1$ can occur if the repair of $d1$ is followed by a failing $d2$.

C. Resilience quality options

To wrap up the previous discussion, if we choose $\text{S} \cap \text{W} = \emptyset$, this guarantees a prompt switchback to the primary $d1$ in the VNO-resilience case upon clearance of a W failure. For the PIP-resilience case, it helps smooth switching to the secondary DC upon a primary DC failure following a repaired W (even though the problem remains for a cleared S failure followed by a primary DC failure). The benefit of choosing $\text{S} \cap \text{B} = \emptyset$ seems limited, since problems stemming from jointly failing B and S are largely the same as those due to failing just S.

The models discussed in the next Section IV will cover these cases, starting with just the disjoint W and B conditions, and indicating what constraints to add to ensure the optional disjointness for S (with W and possibly B).

To ensure continuous synchronization between both data centers, and hence quick recovery and switchback times upon repairs, one can opt for protecting the synchronization path S by a failure disjoint S$'$. This option is deferred to future work.

IV. Models for a single synchronization path

We will adopt a column generation (CG) approach, as this tends to be a highly scalable solution methodology (e.g., its application in [15], [17]). That means that we will divide the model into a Restricted Master Problem (RMP) and a Pricing Problem (PP). The RMP will take as input a set of given configurations (of W, B and S paths, see further), and decide which ones to use to achieve minimal cost. The PP will be responsible for finding such suitable configurations. PP and RMP will be solved alternately until the optimality condition (no more configurations with a negative reduced cost are found by the PP) is satisfied. An integer solution is obtained by solving the last generated RMP, see the flowchart in Fig. 2. Scalability is achieved because this set of PP configurations will be only a fraction of all possible ones. For details on column generation method, we refer to, e.g., [18].

We focus on a core network, comprising optical links and cross-connects as well as data centers, that will be modeled by an undirected graph $G = (V, L)$ where V is the node set (indexed by v) and L is the link set (indexed by ℓ), for which $\omega(v)$ denotes the set of links adjacent to v. Further, the set of data center (DC) nodes will be denoted as $V_{\text{D}} \subseteq V$, with $n_{\text{D}} = |V_{\text{D}}|$ the number of DC nodes. Note that in our setting, a single DC node $v \in V_{\text{D}}$ represents the whole of all real-world data centers that are connected to the same core network node (i.e., an OXC).

Traffic is defined by the number of services (demands), originating from a set of source/service nodes $V_{\text{S}} \subseteq V$, with generic index v_{S}. Let K be the set of services, indexed by k. Each service k is characterized by its bandwidth requirement Δ_k, its source (or origin) v_k, and δ_k (with $0 \leq \delta_k \leq 1$), representing the fraction of Δ_k that is required for synchronization between the primary and the backup data center.

A. *Master problem:* WB-VNO *resilience*

In our CG approach, a configuration is associated with a source node (v_s) where some services are requested. 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 3 paths, one primary path p^W originating at v_s towards a primary data center DC^W, one backup path p^B originating at v_s towards a primary data center DC^B, and one synchronization paths (p^S) between the primary and the backup data center, as well as *(ii)* the services routed and protected by this set of 3 routes. We will protect against single link failures as well as single data center failures. (Extension to generic failures modeled as shared risk groups is fairly trivial, e.g., using a similar approach as [17].)

More formally, a configuration is characterized by the following given parameters[1]:

- $p^W_{\ell,c}$ $= 1$ if link ℓ is used by the working path of configuration c, 0 otherwise;
- $p^B_{\ell,c}$ $= 1$ if link ℓ is used by the backup path of configuration c, 0 otherwise;
- $p^S_{\ell,c}$ $= 1$ if link ℓ is used by the synchronization path of c between the primary data center and the backup data center, 0 otherwise;
- α^c_k $= 1$ if service k is routed and protected by configuration c, 0 otherwise;
- $a^W_{v,c}$ $= 1$ if node v is selected as the primary data center, 0 otherwise;
- $a^{B,c}_v$ $= 1$ if node v is selected as the backup data center, 0 otherwise.

The master problem will determine which configurations to use, using binary decision variables z_c (0 if configuration c is not used). For each link ℓ, let β^W_ℓ be the working bandwidth on ℓ, and β^B_ℓ the backup bandwidth on ℓ. The objective function is to minimize the overall (working + backup) bandwidth requirements, where $\|\ell\|$ denotes the length of link ℓ:

$$\min \quad \sum_{\ell \in L} (\beta^W_\ell + \beta^B_\ell) \cdot \|\ell\|, \tag{1}$$

subject to:

$$\sum_{c \in C} \alpha^c_k \, z_c \geq 1 \qquad k \in K \tag{2}$$

$$\sum_{c \in C} \Delta_{k_c} \left(p^W_{\ell,c} + \delta_k \, p^S_{\ell,c}\right) z_c = \beta^W_\ell \qquad \ell \in L \tag{3}$$

$$\sum_{c \in C} \Delta_{k_c} \, p^W_{\ell',c} \, p^B_{\ell,c} \, z_c \leq \beta^B_\ell \qquad \ell' \in L, \ell \in L \setminus \{\ell'\} \tag{4}$$

$$\sum_{c \in C} \Delta_{k_c} \, a^W_{v,c} \, p^B_{\ell,c} \, z_c \leq \beta^B_\ell \qquad v \in V_D, \ell \in L \tag{5}$$

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

$$\beta^W_\ell, \, \beta^B_\ell \in \mathbb{R} \qquad \ell \in L. \tag{7}$$

Constraints (2) are the demand constraints, and ensure that each service k is granted. Constraints (3) compute the overall bandwidth requirements on link ℓ under failure-free conditions: this is the sum of the working path (W) and synchronization path (S) bandwidths, where the latter only is a fraction δ_k of the former. Constraints (4) ensure sufficient backup bandwidth requirements on link ℓ to cover a failure of any other link ℓ'. Constraints (5) guarantee sufficient backup bandwidth ℓ to handle any data center failure.

Note that in our experiments, we will not consider any network capacity constraints. However, should one want to pose capacity limits on the links, this can be accommodated by adding the following constraints (using BW_ℓ to denote the capacity of link ℓ):

$$\beta^W_\ell + \beta^B_\ell \leq BW_\ell \qquad \ell \in L. \tag{8}$$

B. *Master problem:* WB-PIP *resilience*

For PIP-resilience, we need to replace constraints (5) with (9). Remark that S will need to support the full request bandwidth when a node failure occurs at the primary data center (but it can be shared among different failure cases):

$$\sum_{c \in C} \Delta_{k_c} \, a^W_{v,c} \, p^S_{\ell,c} \, z_c \leq \beta^B_\ell \qquad v \in V_D, \ell \in L. \tag{9}$$

Note that the synchronization bandwidth on the S path will be reserved on top of that (see (3) in the master problem). Since the backup capacity on S is only required when the primary DC fails, we then cannot synchronize and hence one could argue that we should actually add a factor $1 - \delta_k$ in (9). Yet, upon restoration of the primary DC failure, we will need to synchronize it and thus do need the synchronization bandwidth in addition to the full traffic bandwidth along the path S.

[1]From the master problem's perspective, these are indeed given parameters. However, in the pricing problem they will become decision variables.

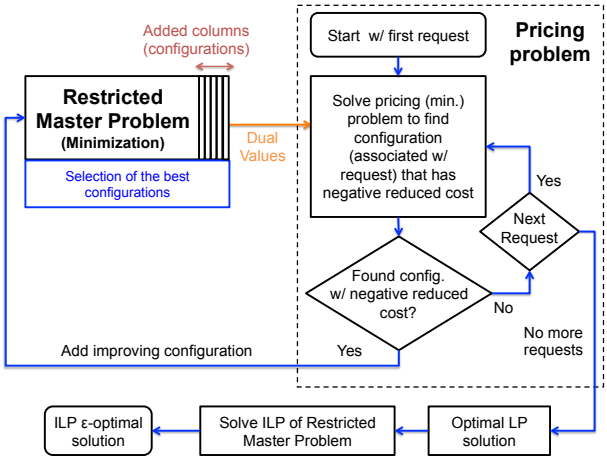

Fig. 2. Flowchart of the CG ILP Approach.

C. Pricing problem: WB-VNO resilience

Recall that the pricing problem (PP) will determine useful configurations, i.e., routes for W, B and S paths. Each PP is written for a given source node v_{S} and for a given set of requests originating there. The given parameters Δ_k and δ_k retain their definition for a request k as in the RMP.

The sets of variables are as follows:

- α_k $= 1$ if service k is granted in the configuration under construction, 0 otherwise;
- p_ℓ^{W} $= 1$ if link ℓ is used by the working path of the configuration under construction, 0 otherwise;
- p_ℓ^{B} $= 1$ if link ℓ is used by the backup path of the configuration under construction, 0 otherwise;
- p_ℓ^{S} $= 1$ if link ℓ is used by the synchronization path of the configuration under construction between the primary data center and the backup data center, 0 otherwise;
- a_v^{W} $= 1$ if node v is selected as a data center location by the working path in the configuration under construction, 0 otherwise;
- a_v^{B} $= 1$ if node v is selected as a data center location by the backup path in the configuration under construction, 0 otherwise;
- d_v^{W} $= 1$ if node v is on the working path in the configuration under construction, 0 otherwise;
- d_v^{B} $= 1$ if node v is on the backup path in the configuration under construction, 0 otherwise;
- d_v^{S} $= 1$ if node v is on the synchronization path between the primary data center and the backup data center in the configuration under construction, 0 otherwise.

The objective of the PP is to minimize the reduced cost as obtained from the RMP, defined as:

$$\overline{\mathrm{COST}} = 0 - \sum_{\ell\in L} u_\ell^{(3)}\,\Delta_{k_c}\,\left(p_{\ell,c}^{\mathrm{W}} + \delta_k\, p_{\ell,c}^{\mathrm{S}}\right) - \sum_{k\in K} u_k^{(2)}\,\alpha_k - \sum_{\ell\in L}\sum_{\ell'\in L\setminus\{\ell\}} u_{\ell\ell'}^{(4)}\,\Delta_k p_\ell^{\mathrm{W}}\,p_{\ell'}^{\mathrm{B}} - \sum_{v\in V_{\mathrm{D}}}\sum_{\ell\in L} u_{v\ell}^{(5)}\Delta_k a_v^{\mathrm{W}} p_\ell^{\mathrm{B}} \quad (10)$$

where $u^{(3)}$, $u_v^{(2)}$, $u_{\ell\ell' F}^{(4)}$, $u_{v\ell}^{(5)}$ are the values of the dual variables associated with constraints (3), (2), (4), (5), respectively. (Note that the first explicit 0 term stems from the RMP objective, which does not contain the configuration variable z_c.)

The path and data center variables have to obey:

$$\sum_{\ell\in\omega(v)} p_\ell^{\mathrm{W}} = \begin{cases} 1 & \text{if } v = v_s \\ 2\,d_v^{\mathrm{W}} - a_v^{\mathrm{W}} & \text{otherwise} \end{cases} \qquad v\in V \quad (11)$$

$$\sum_{\ell\in\omega(v)} p_\ell^{\mathrm{B}} = \begin{cases} 1 & \text{if } v = v_s \\ 2\,d_v^{\mathrm{B}} - a_v^{\mathrm{B}} & \text{otherwise} \end{cases} \qquad v\in V \quad (12)$$

$$\sum_{\ell\in\omega(v)} p_\ell^{\mathrm{S}} = 2\,d_v^{\mathrm{S}} - a_v^{\mathrm{W}} - a_v^{\mathrm{B}} \qquad v\in V \quad (13)$$

$$p_\ell^{\mathrm{W}} + p_\ell^{\mathrm{B}} \le 1 \qquad \ell\in L \quad (14)$$

$$\sum_{v\in V_{\mathrm{D}}} a_v^{\mathrm{W}} = 1\,;\ \sum_{v\in V_{\mathrm{D}}} a_v^{\mathrm{B}} = 1\,;\ \sum_{v\notin V_{\mathrm{D}}} a_v^{\mathrm{W}} + a_v^{\mathrm{B}} = 0 \quad (15)$$

$$a_v^{\mathrm{W}} + a_v^{\mathrm{B}} \le 1 \qquad v\in V_{\mathrm{D}} \quad (16)$$

$$a_v^{\mathrm{W}}, a_v^{\mathrm{B}} \in \{0,1\} \qquad v\in V \quad (17)$$

$$p_\ell^{\mathrm{W}}, p_\ell^{\mathrm{B}}, p_\ell^{\mathrm{S}} \in \{0,1\} \qquad \ell\in L. \quad (18)$$

Constraints (11)–(13) are the conventional flow constraints for working, backup and synchronization paths. Constraints (14) force p_{W} and p_{B} to be disjoint[2]. Constraints (15) ensure that each configuration has exactly one primary and one back up data center, while constraints (16) coerce them to be different. Constraints (17)–(18) define the domains of the variables.

D. Pricing problem: WB-PIP resilience

The objective of the PP for the PIP-resilience case is:

$$\overline{\mathrm{COST}} = 0 - \sum_{\ell\in L} u_\ell^{(3)}\,\Delta_{k_c}\,\left(p_{\ell,c}^{\mathrm{W}} + \delta_k\, p_{\ell,c}^{\mathrm{S}}\right) - \sum_{k\in K} u_k^{(2)}\,\alpha_k - \sum_{\ell\in L}\sum_{\ell'\in L\setminus\{\ell\}} u_{\ell\ell'}^{(4)}\Delta_k p_\ell^{\mathrm{W}} p_{\ell'}^{\mathrm{B}} - \sum_{v\in V_{\mathrm{D}}}\sum_{\ell\in L} u_{v\ell}^{(9)}\Delta_k a_v^{\mathrm{W}} p_\ell^{\mathrm{S}}. \quad (19)$$

Further, the flow constraints need to be modified in order to enforce both working and backup paths to connect to the primary data center. The constraints (12) are replaced by (20):

$$\sum_{\ell\in\omega(v)} p_\ell^{\mathrm{B}} = \begin{cases} 1 & \text{if } v = v_s \\ 2d_v^{\mathrm{B}} - a_v^{\mathrm{W}} & \text{otherwise} \end{cases} \qquad v\in V. \quad (20)$$

E. Improved QoS strategy

As discussed in Section III-C, by enforcing the disjointness between W and S we can reduce the transition time *(i)* for the VNO-resilience case to switch back to the primary data center after clearance of a W failure, and *(ii)* for the PIP-resilience case to switch to the secondary data after two consecutive failures, first on W, then of the primary data center. This can be realized by adding constraints (21) to the pricing problem:

$$p_\ell^{\mathrm{W}} + p_\ell^{\mathrm{S}} \le 1 \qquad \ell\in L. \quad (21)$$

Accordingly, should one want to enforce disjointness between S and B, similar constraints can be added (replacing p_ℓ^{W} with p_ℓ^{B} in (21)).

[2]This ensures protection against single link failures. For a more extensive protection against multiple simultaneous failures, one can model these as shared risk groups (SRGs) and use a similar approach as in [17].

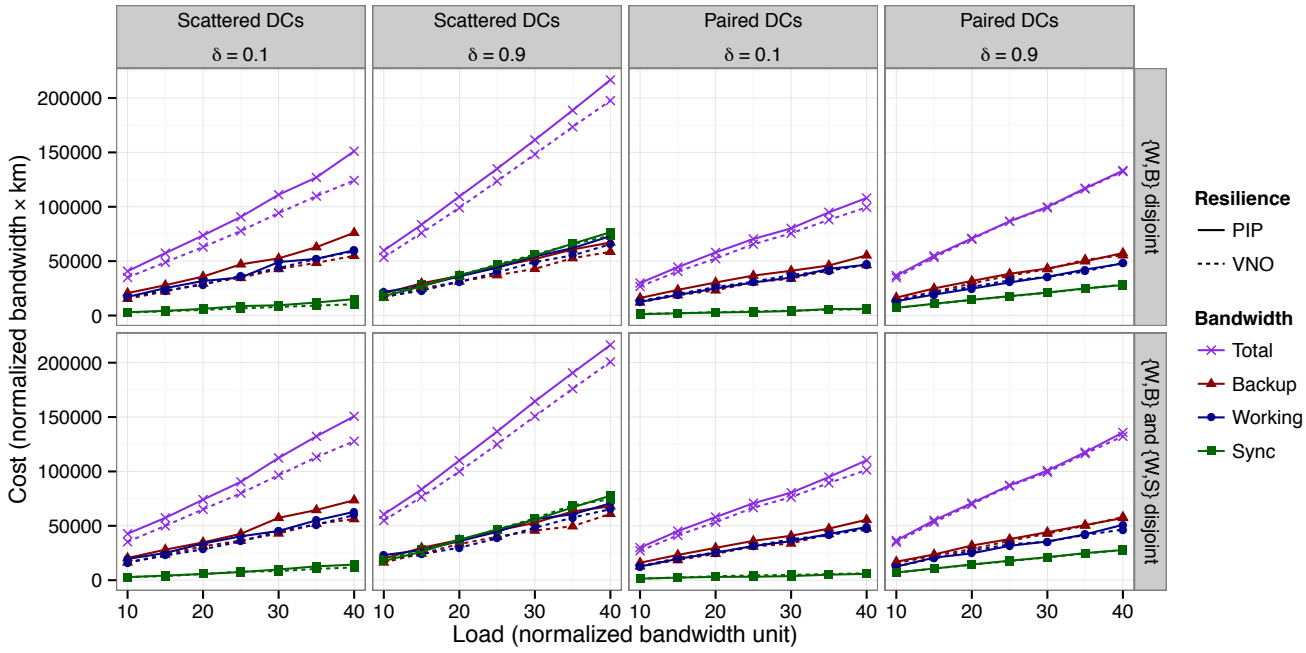

Fig. 3. Experiments on the US topology, for $\{\text{W}, \text{B}\}$ disjointness (top), or both $\{\text{W}, \text{B}\}$ and $\{\text{W}, \text{S}\}$ disjointness (bottom).

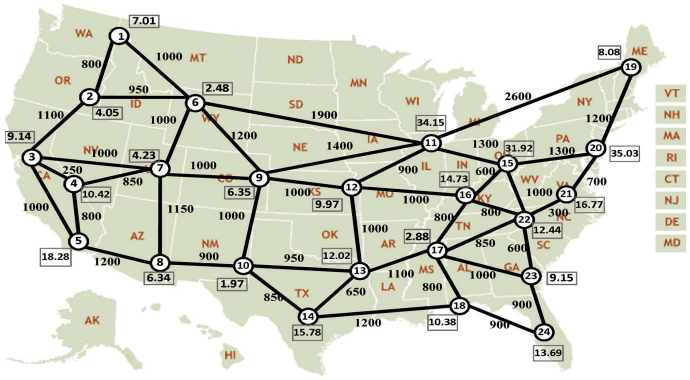

Fig. 4. The US topology, as reproduced from [19].

V. Numerical Results

We run experiments on the 24-node US nationwide backbone network shown in Fig. 4 with 4 data centers. The network has 43 non-directional links, labeled with their lengths in km. The bandwidth requirement for each service request is generated randomly with uniform distribution between 0 and 1 normalized bandwidth units. We generate uniform traffic, i.e., the source node of a request is chosen randomly, and vary the total requested bandwidth (i.e., the total load) from 10 to 40 units (the number of generated requests varied from 22 to 83). As per the CG model, each request is individually provisioned: requests originating from the same source node are not forced to follow the same paths towards the same data centers.

To study the effect of DCs location, we consider two sets of DC locations. In the first set, DCs are fairly uniformly *scattered* over the geographical region: {WY(6), AZ(8), IL(11), AL(18)}. In the second set, DCs are selected in *paired* locations: {CA(3), UT(7), KY(16), NC(22)}. (A motivation for using *paired* locations could be to aim to have similar path lengths, and hence latencies, to both the primary and backup data centers[3].) For each DC constellation, we carry out the experiment for two synchronization parameter settings: $\delta = 0.1$ and $\delta = 0.9$.

We expect VNO-resilience to outperform PIP-resilience in all settings, since under VNO-resilience we have more flexibility to choose the backup paths than for PIP-resilience (indeed, the physical routing as obtained in the latter case is always also allowed in VNO-resilience). This is confirmed by our results shown in Fig. 3, which we now discuss in detail.

A. Effect of DC locations and synchronization bandwidth (δ)

First of all, going from *scattered* to *paired* DC locations, we find that the total bandwidth cost is reduced by roughly 30% (for the same δ and resilience scheme). This can be explained by the fact that paired DCs enable more sharing, since the backup paths go to 2 regions (east and west) instead of 4, and the synchronization paths are shorter.

Intuitively, we expect the *paired* DC configuration to have lower cost differences between VNO- and PIP-resilience. Indeed, VNO-resilience's potential advantage mainly stems from shorter backup route options avoiding the inter-DC path, yet this path is quite short in the paired DC case and thus does

[3]We verified that for the chosen *paired* DC locations, the majority of the source nodes indeed has one of the pairs as two closest, path-disjoint, DCs among the four given in total.

not amount to a heavy penalty. Our results confirm this, and the cost advantage VNO-resilience even is neglible in the $\delta = 0.9$ case: for high δ the synchronization bandwidth becomes more dominant (thus limiting VNO's gain in terms of lower backup bandwidth). Obviously, overall cost for both VNO- and PIP-resilience and both DC settings does increase for higher δ.

B. Effect of disjointness of W and S

In our experiments, the penalty for adding the disjointness between W and S is very small at less than 5%. It is likely that in most cases, W and S are already link-disjoint which is also intuitively understandable. This suggests that we can improve the quality of the resilience (in terms of recovery times, see Section III-C) by enforcing the disjointness between W and S, and only pay an almost negligible extra bandwidth cost.

VI. Conclusion

We have carefully outlined the various options in providing resilient virtual networks for cloud services, thus under an anycast traffic scenario: we only assumed the traffic sources to be given, while destinations can be chosen among a set of given data center (DC) locations. We considered a virtualized network environment, where virtual network operators (VNOs, that will provision the cloud service requests) make use of underlying physical infrastructure offered by physical infrastructure providers (PIP). We explained the different mappings in a VNO- vs a PIP-resilience scenario, comprising not just working and backup paths, but also explicitly accounting for the synchronization path (and associated bandwidth cost) between primary and secondary data centers. We indeed provide resilience against both network and DC failures. Our thorough discussion of the various failure scenarios revealed disjointness requirements for that synchronization path that can improve the quality of resilience in terms of recovery times.

We subsequently detailed scalable models to find routings and DC allocations for cloud requests, with minimal cost, for the proposed resilience strategies (VNO vs PIP) and options for the synchronization path (one or two disjoint ones). Our results showed that the intuitively expected advantage of VNO-resilience actually can be quite limited, when DCs occur in paired configurations (which may be desirable to obtain similar latencies towards both primary and backup DC). Moreover, if the synchronization bandwidth becomes a substantial fraction of the actual traffic bandwidth, this relative cost advantage becomes very limited.

Acknowledgment

B. Jaumard has been supported by a Concordia University Research Chair (Tier I) and by an NSERC (Natural Sciences and Engineering Research Council of Canada) grant.

References

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

[2] N. Chowdhury, K. Mosharaf, and R. Boutaba, "A survey of network virtualization," *Computer Networks*, vol. 54, no. 5, pp. 862–876, April 2010.

[3] I. Barla, D. Schupke, and G. Carle, "Resilient virtual network design for end-to-end cloud services," in *Proceedings of the 11th international IFIP TC 6 conference on Networking - Volume Part I*. Heidelberg: Springer-Verlag Berlin, 2012, pp. 161–174.

[4] ——, "Delay performance of resilient cloud services over networks," in *IEEE 10th International Symposium on Parallel and Distributed Processing with Applications (ISPA)*, 2012, pp. 512 – 517.

[5] I. Barla, D. Schupke, M. Hoffmann, and G. Carle, "Optimal design of virtual networks for resilient cloud services," in *Proceedings of IEEE/VDE Workshop on Design of Reliable Communication Networks - DRCN*, Budapest, Hungary, 4–7 March 2013.

[6] M. Bui, B. Jaumard, and C. Develder, "Anycast end-to-end resilience for cloud services over virtual optical networks (invited)," in *15th International Conference on Transparent Optical Networks (ICTON)*, Cartagena, Spain, June 2013.

[7] K. Lee and E. Modiano, "Cross-layer survivability in WDM-based networks," in *Annual Joint Conference of the IEEE Computer and Communications Societies - INFOCOM*, Rio de Janeiro, Brazil, April 2009, pp. 1017–1025.

[8] 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 *IEEE Global Telecommunications Conference - GLOBECOM*, Miami, FL, USA, December 2010, pp. 1–6.

[9] M. Bui, B. Jaumard, C. Cavdar, and B. Mukherjee, "Design of a survivable VPN topology over a service provider network," in *Proceedings of IEEE/VDE Workshop on Design of Reliable Communication Networks - DRCN*, Budapest, Hungary, 4–7 March 2013, pp. 1–8.

[10] B. Jaumard, M. Bui, B. Mukherjee, and C. Vadrevu, "{IP} restoration vs. optical protection: Which one has the least bandwidth requirements?" *Optical Switching and Networking*, vol. 10, no. 3, pp. 261 – 273, 2013.

[11] J. Jiang, T. Lan, S. Ha, M. Chen, and M. Chiang, "Joint VM placement and routing for data center traffic engineering," in *IEEE Annual Joint Conference of the IEEE Computer and Communications Societies - INFOCOM*, Orlando, FL, USA, March 2012, pp. 2876–2880.

[12] M. Alicherry and T. Lakshman, "Network aware resource allocation in distributed clouds," in *IEEE Annual Joint Conference of the IEEE Computer and Communications Societies - INFOCOM*, Orlando, FL, USA, March 2012, pp. 963–971.

[13] W.-L. Yeow, C. Westphal, and U. Kozat, "Designing and embedding reliable virtual infrastructures," in *Proc. 2nd ACM SIGCOMM Workshop Virtualized Infrastructure Systems and Architectures (VISA 2010)*, New Delhi, India, 3 Sep. 2010, pp. 33–40.

[14] C. Develder, B. Mukherjee, B. Dhoedt, and P. Demeester, "On dimensioning optical grids and the impact of scheduling," *Photonic Network Communications*, vol. 17, no. 3, pp. 255–265, June 2009.

[15] 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.

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

[17] C. Develder, J. Buysse, B. Dhoedt, and B. Jaumard, "Joint dimensioning of server and network infrastructure for resilient optical grids/clouds," *IEEE/ACM Transactions on Networking*, 2014, to appear.

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

[19] M. Batayneh, D. Schupke, M. Hoffmann, A. Kirstadter, and B. Mukherjee, "On routing and transmission-range determination of multi-bit-rate signals over mixed-line-rate WDM optical networks for carrier ethernet," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1304–1316, October 2011.