Resilient backbone networks for multi-site data centers: Exploiting anycast (re)routing for multi-period traffic

Ting Wang, Brigitte Jaumard

CSE, Concordia University, Montréal (QC) H3G 1M8 Canada

bjaumard@cse.concordia.ca

Chris Develder

INTEC – IBCN, Ghent University – iMinds, Technologiepark Zwijnaarde 15, BE-9052 Gent, Belgium

chris.develder@intec.ugent.be

Abstract: Using column generation, we dimension networks interconnecting multi-site data centers (DCs), for multi-period traffic with anycast routing and resilience against single link/DC failures. We quantify (backup) resource savings from rerouting demands lasting multiple periods.

OCIS codes: 060.0060, 060.4256, 060.4257.

1. Introduction

Thanks to the performance of optical networking, applications serving both business and end user needs are increasingly offered from remote data centers (DCs), adopting the cloud computing paradigm [1]. Given the high bandwidth and low latency of (optical) network connections towards those DCs, the exact location of the data center serving a particular customer's request does not really matter (as long as resources are used efficiently). This leads to the adoption of anycast routing: providers can choose what particular DC to use from their multi-site DC constellation. Anycast can also be exploited for resiliency, by rerouting demands to alternate backup DCs if a failure affects either the primary DC or the (optical) network connection towards it. We previously explored the potential benefits of adopting such anycast principle to minimize the resource capacity requirements, by solving network dimensioning problems with static traffic (e.g., [2]). In the current paper, we rather study time-varying traffic and assume that routing and DC assignment can be revised at discrete points in time. We build on our latest work on this topic [3], and contribute with (i) a new column generation model that uses path-based (rather than link-based) formulations to limit rerouting, (ii) an exploration of the effect of parallel execution of multiple so-called pricing problems (PPs), and (iii) a wider range of experiments, with different topology set-ups in terms of DC locations and traffic variations over time.

2. Dimensioning resilient backbone networks for multi-site DCs

2.1. Problem Statement

We will dimension a backbone network interconnecting nodes where demands originate as well as the various data centers spread across multiple sites (see Fig. 2 for an example). We are *given*
• The network topology, in terms of network nodes (e.g., OXCs) and links interconnecting them, and
• The service requests to be routed in each of the consecutive time periods, where requests are specified in terms of a given source node, the number of unit demands (where a unit represents an amount of bidirectional bandwidth to provide from the source to a DC to be chosen), and its duration (i.e., a holding time of 1 or more periods).
We then need to *find* for each request, and for each time period it is active, the routes from its source to a selected DC, *such that* the total amount of required network resources (i.e., the bandwidth crossing each link, summed over all links) is minimized and each request remains operational under given failure scenarios. For the latter, we will require requests to be protected against single failures of either a single link or a single DC.

We adopt VNO-resilience (as presented, e.g., in [3]). The working path (π^{W}) routes the services from their source node (v_S) towards the primary DC (d^{W}), the protection path (π^{B}) towards the backup DC (d^{B}), while π^{W} and π^{B} are disjoint in their physical layer mapping. In addition, a synchronization path (π^{S}) is established in order to handle
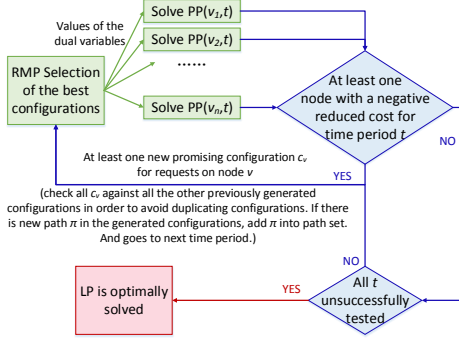
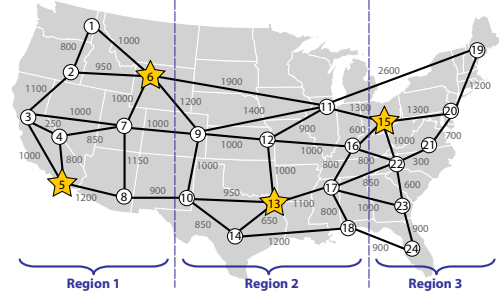Fig. 1: Our parallel PP solving scheme.

Fig. 2: A 3-region topology with 4 DC locations (indicated with stars).

migration and failure routing requirements when a DC failure occurs: services then need to be rerouted from the chosen primary DC destination d^{W} to the backup d^{B}.

2.2. Modeling

We here describe only the general ideas of our column generation (CG) model, and refer to [4] (which is available upon request to the authors) for the full mathematical model. The general idea of column generation is to split the overall optimization problem into a (restricted) master problem (RMP), and a pricing problem (PP). The RMP decides on the optimal combination out of a set of candidate "configurations", C. The PP finds/creates new configuration(s) that, when added to the RMP configuration set C leads to an RMP solution with a better objective value. More concretely, a "configuration" in our model is associated with a given source node v_{S}, and comprises a set of the three paths as required by the VNO-resilience scheme (primary path π^{W} from source to a chosen DC, backup path π^{B} to a potentially different DC, and synchronization path π^{S} between the primary and the backup DCs).

The primary objective is to minimize the amount of required network resources to serve all requests. We compare three rerouting strategies: (i) in the baseline *Scenario I* each request sticks to the same routing configuration for all time periods of its holding time, (ii) *Scenario II* allows a request to change its backup and/or synchronization paths from one period to the next, and (iii) *Scenario III* permits complete rerouting of all paths, including the working path.

To solve the resulting CG model, multiple strategies can be devised to decide what PP(s) to solve before adding the new configuration(s) to the RMP (and solve it). We propose to solve multiple PPs in parallel, as illustrated in Fig. 1.

3. Numerical results

We consider a 24-node US network illustrated in Fig. 2. Given the anycast assumption, demands have a fixed source, whereas the destination DC is to be chosen by the routing algorithm. The amount of traffic generated varies in terms of time-of-day (i.e., period), as well as per region (which are assumed to have distinct, artificial time zones). For the illustrated 3-region case, 33.3% originates from Region 1, 37.5% from Region 2, and the remaining 29.2% from Region 3. (We additionally also experimented with a 4-region case, where 29.2% originates from Region 1, 16.6% from Region 2, 25% from Region 3, and 29.2% from Region 4.) For a given Region, the traffic varies during the day, with 48% of the Region's traffic in 8 am–4 pm, 38% in 4 pm–12 am, and the remaining 14% from 12 am–8 am. The volume of traffic originating in a given Region at a specific time periods, is further partitioned in terms of different holding times. We consider mixes of single-period and two-period traffic and experiment with 20%, 50% or 80% two-period traffic. Figure 3 plots resulting bandwidth requirements, relative to the baseline Scenario I without rerouting.

4. Conclusion

We have defined a new column generation model (detailed in [4]) to solve a multi-period traffic dimensioning problem for resilient backbone networks for multi-site data centers. We applied it in an experiment on a 24-node network with 3 cyclic time periods, time-shifted across 3 regions with their own distinct time zones. That experiment's results aimed to quantify the potential bandwidth savings achievable by rerouting demands that span multiple time periods. Our main observations are:
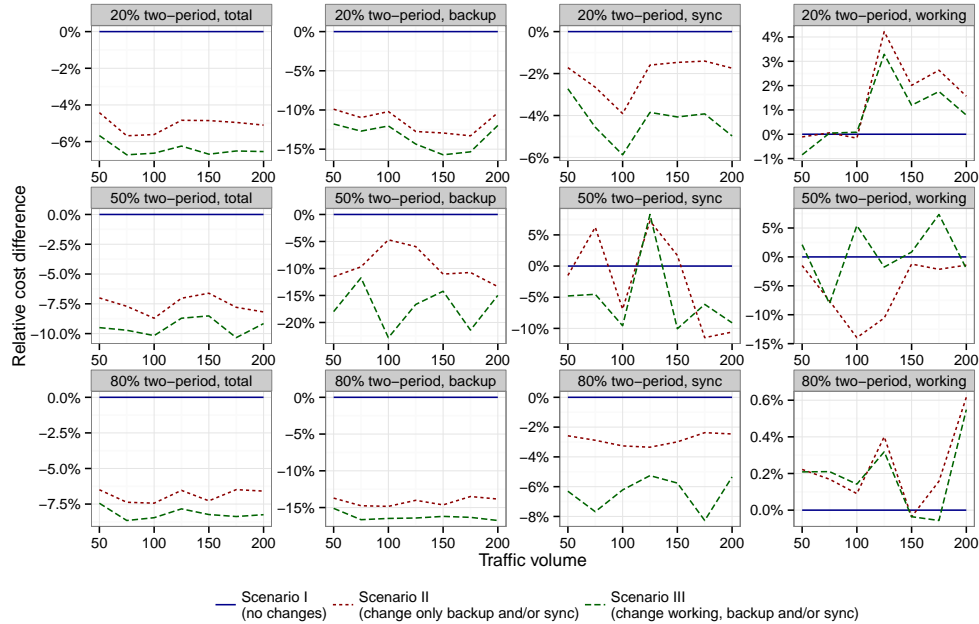
Fig. 3: The bandwidth requirements for time-varying traffic, for the 3-Region topology.

- The bandwidth savings mainly stem from *backup* paths (because of increased sharing with the requests starting in the 2nd period of two-period requests).
- Part of backup capacity savings are negated by longer *working* paths, chosen to avoid overlap among concurrent demands, and thus allow more sharing of backup capacity.
- When working paths may be rerouted (Scenario III), the reduction in bandwidth requirements is slightly higher than when only rerouting backup/synchronization paths (Scenario II), but the difference is not significant.
- The overall bandwidth savings from rerouting for the 3-Region topology seem not to exceed 10%.

Furthermore, we also showed that by adopting a parallel pricing problem (PP) solving strategy, we can realize a substantial amount of (wall clock) time for solving the dimensioning problem. That overall saving is achieved by solving multiple PPs in parallel (one for each source node), and only re-solving the RMP with the newly found configurations (at most one per source node) after they all completed (see Fig. 1). Compared to a naive serial approach that only solves one PP at a time, and re-solves the RMP each time a new configuration is found (and thus added to the RMP), we clearly save about $1/N$ time per round (where a "round" comprises solving one PP for each source node), with N the number of source nodes. Even though in the parallel case we have to iterate over slightly more such rounds — indeed, adding 1 configuration for a given source node v_s impacts what other configuration for a different source node v'_s might share bandwidth with the new v_s configuration — the net saving in total time still is substantial (e.g., a factor of about $14\times$ in our case study) compared to a naive serial execution plan.

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," Proc. IEEE **100**, 1149–1167 (2012).
2. C. Develder, J. Buysse, B. Dhoedt, and B. Jaumard, "Joint dimensioning of server and network infrastructure for resilient optical grids/clouds," IEEE/ACM Trans. Netw. **22**, 1591–1606 (2014).
3. T. Wang, B. Jaumard, and C. Develder, "A scalable model for multi-period virtual network mapping for resilient multi-site data centers," in "Proc. 17th Int. Conf. Transparent Optical Netw. (ICTON 2015)," (Budapest, Hungary, 2015), pp. 1–8.
4. T. Wang, "Time-varying resilient virtual networking mapping for multi-location cloud data centers," Master's thesis, Concordia University, Montreal, Canada (2016).