Survivable Manycast, Anycast and Replica Placement in Optical Inter-Datacenter Networks

Marija Furdek[1], Ajmal Muhammad[1], Nina Skorin-Kapov[2]
[1] Royal Institute of Technology (KTH), Stockholm, Sweden, Email: {marifur, ajmalmu}@kth.se
[2] University Centre of Defense, San Javier Air Force Base, Santiago de la Ribera, Spain, Email: nina.skorinkapov@cud.upct.es

ABSTRACT Inter-datacenter networks need to support datacenter communication with the end-users, as well as content replication and synchronization between datacenters in a reliable manner. This paper presents a survivable manycast, anycast and replica placement strategy for optical inter-datacenter networks resulting in reduced overall network resource consumption.

Index Terms— Content placement, inter-datacenter networks, manycast routing, anycast routing.

I. Introduction

The rapidly expanding cloud-based services supported by geographically distributed data centers require a high-capacity, highly reliable and resource-efficient datacenter interconnection network, which can be satisfied only by optical fiber infrastructure. Optical datacenter networks must support the emerging disparate traffic patterns that emerge from connecting the network users to the datacenters (user-DC traffic), as well as due to synchronization among the datacenters (DC-DC traffic). For a given subset of DC nodes that host a desired content, the user-DC traffic typically follows the anycast traffic paradigm, where the user can access the content by connecting to any of the DC nodes in the subset. In other words, a connection is established between a determined source and any one of the set of possible destinations. The DC-DC traffic follows the manycast traffic paradigm, where first a subset of DC nodes is selected to host a specific content, and then a multicast tree is established among the selected nodes. In order to satisfy the immense network traffic growth rates [1], resource-efficient datacenter networking solutions require schemes for assignment of resources at the datacenter and the network level, enabling a consolidated view of both the user-DC and the DC-DC traffic [2].

Besides supporting the ultra-high data rates, datacenter networks must also provide a high degree of resiliency. Optical fiber links are prone to failures, typically caused by construction equipment digging through the fiber ducts. Moreover, entire network nodes which host the datacenters can also fail, e.g., due to a power outage, a natural disaster or a deliberate attack [3]. While replicating content across a set of geographically distributed datacenters inherently improves content accessibility in the presence of failures [4], guaranteeing a certain degree of reliability requires tailored resiliency schemes.

Numerous approaches aimed at providing protection from failures of single, and possibly multiple links and/or DC nodes can be found in the literature. Protection from link failures along working paths of user demands can be provided by reserving resources for backup paths which can be dedicated [5] or shared among multiple working paths that do not fail simultaneously [6]. To ensure protection from failures of primary DC nodes, various survivability approaches assign a backup DC node to each user request, where resources at backup DC nodes can be dedicated [7] or shared among user demands that do not fail simultaneously [8]. The approach in [9] performs replica placement and subsequent resource allocation to user demands in elastic optical datacenter network. The authors in [10] propose an approach to protect multicast transmission among DCs against critical single-node failures.

While the aforementioned approaches address one or more aspects of survivable datacenter network planning, namely, replica placement or survivable routing of user-DC and/or DC-DC traffic, none of them considers the placement of content replicas and resource allocation to both the user-DC and DC-DC traffic type jointly. The approach for joint planning of DC networks in [2] demonstrated significant resource savings obtained through consolidated manycast and anycast routing and replica placement (MARP). In this paper, we extend upon the previous work in [2] by proposing a survivable MARP strategy to protect from single link and single node failures.

II. Survivable Manycast, Anycast and Replica Placement: S-MARP

For a given set of contents, DC nodes and user demands, the survivable manycast, anycast and replica placement (SMARP) problem aims at determining the DC nodes to host the content replicas and establishing a set of connections to support the user-DC and DC-DC traffic in a survivable manner such that the total amount of resources assigned to all connections is minimized.

A. Problem Definition

Given is a network topology modeled as a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, where a subset of nodes $\mathsf{D} \in \mathcal{V}$ host datacenters. A set of contents C is also given, where k replicas of each content need to be placed at different datacenters. A set A of user demands is also given, where each demand is defined by a source node $s \in \mathcal{V}$ and requested content $c \in \mathsf{C}$. The objective of the S-MARP problem is to determine the locations of all replicas for each content, establish a multicast light-tree among nodes hosting replicas of a particular content, and set up a working and a backup path between the source node of each user request and a DC, such that the total length of established paths is minimized. To protect user demands from different types of failures, we apply dedicated path protection and consider three variants of S-MARP, described as follows. The link-disjoint S-MARP, denoted as S-MARP-L protects user requests from failures along working paths by assigning link-disjoint backup path that connects the source node to the same DC as the working path. The DC-disjoint S-MARP, denoted as S-MARP-DC, provides protection from link- and DC-failures by assigning a backup DC different from the working one, as well as a link-disjoint backup path. Finally, we also consider the optional variant of S-MARP, denoted as S-MARP-O, where the backup path is allowed to use any DC that results in the lowest overall cost of the established solution.

B. Integer Linear Program for the Survivable MARP Problem: S-MARP

Input parameters:

$\mathcal{G}(\mathcal{V},\mathcal{E})$: a directed graph where $\mathcal{V}$ is the set of vertices that represent the network nodes, $\mathcal{E}$ is the set of edges that represent the network links; C: set of contents; $\mathsf{D} \in \mathcal{V}$: set of datacenter locations; k: number of needed replicas for each content; M: set of manycast demands (d,c), where $d \in \mathsf{D}$ is the root/main datacenter for content $c \in \mathsf{C}$; A: set of anycast demands (s,c), where $s \in \mathcal{V}$ is the source node and $c \in \mathsf{C}$ is the requested content; W: set of available wavelengths; $\Upsilon = \{\mathsf{K_1},\mathsf{K_2},\ldots,\mathsf{K_{|D|}}\}$: set of constants with values such that $\sum_{\mathsf{l}=1}^{\hat{\mathsf{l}}-1} \mathsf{K_l} \ll \mathsf{K_{\hat{l}}}$, used for assigning a unique identifier to each node of D;

Variables

- $x_{(i,j)}^{(d,c),w} \in \{0,1\}$- equal to 1 if $w \in \mathsf{W}$ is used on link $(i,j) \in \mathcal{E}$ for manycast demand (d,c) and 0 otherwise;
- $U_i^{(d,c)} \in \mathbb{Z}$- denote the order of vertices included in the route tree for manycast demand (d,c);
- $y_{(i,j)}^{(s,c),w} \in \{0,1\}$- equal to 1 if $w \in \mathsf{W}$ is used on link $(i,j) \in \mathcal{E}$ for working lightpath of anycast demand (s,c) and 0 otherwise;
- $z_{(i,j)}^{(s,c),w} \in \{0,1\}$- equal to 1 if $w \in \mathsf{W}$ is used on link $(i,j) \in \mathcal{E}$ for backup lightpath of anycast demand (s,c) and 0 otherwise;
- $N_d^c \in \{0,1\}$- equal to 1 if $d \in \mathsf{D}$ is selected as one of the location for c and 0 otherwise;
- $\bar{A}_d^{(s,c)} \in \{0,1\}$- equal to 1 if d is selected as the destination node for working lightpath of anycast demand (s,c) and 0 otherwise;
- $\hat{A}_{\hat{d}}^{(s,c)} \in \{0,1\}$- equal to 1 if $\hat{d}$ is selected as the destination node for backup lightpath of anycast demand (s,c) and 0 otherwise;
- $\Omega^{(d,c),w} \in \{0,1\}$- equal to 1 if w is used by (d,c) and 0 otherwise;
- $\bar{\Omega}^{(s,c),w} \in \{0,1\}$- equal to 1 if w is used by (s,c) for working lightpath and 0 otherwise;
- $\hat{\Omega}^{(s,c),w} \in \{0,1\}$- equal to 1 if w is used by (s,c) for backup lightpath and 0 otherwise;

Objective function

$$Minimize: \sum_{(d,c)\in \mathsf{M}} \sum_{(i,j)\in\mathcal{E}} \sum_{w\in \mathsf{W}} x_{(i,j)}^{(d,c),w} + \sum_{(s,c)\in \mathsf{A}} \sum_{(i,j)\in\mathcal{E}} \sum_{w\in \mathsf{W}} y_{(i,j)}^{(s,c),w} + \sum_{(s,c)\in \mathsf{A}} \sum_{(i,j)\in\mathcal{E}} \sum_{w\in \mathsf{W}} z_{(i,j)}^{(s,c),w} \quad (1)$$

Constraints

The objective (1) is to minimize the number of wavelength-links required to establish the requested manycast light-trees and survivable anycast lightpaths. Constraint (2) selects k locations for content replicas. Constraints (3-8) are used to build each manycast light-tree. Constraint (3) specifies that there must be at least one wavelength associated with content c outgoing from the main DC node. Constraint (4) prohibits any incoming wavelengths to the main DC location. Constraint (5) ensures that the destination nodes of each tree host the associated content replicas. Constraint (6) specifies that each node other than the main DC can have outgoing wavelengths only if it has incoming wavelengths. Constraint (7) ensures that nodes which are not in the set of possible DC locations and have an incoming wavelength must have at least one outgoing wavelength. Note that the last two constraints (i.e., constraints (6,7)) differ from the flow conservation constraints, since the number of outgoing wavelength links for intermediate nodes in a light-tree can be greater than the number of input wavelength links due to branching of the tree. Constraint (8) prevents the formation of loops in established paths. Constraints (9,10) enforce the wavelength continuity constraint for manycast light-trees.

Constraints (11-13) are used to establish the anycast working lightpaths. Constraint (11) enforces flow conservation for each working lightpath by making sure that the number of outgoing and incoming wavelength links is equal for every node along its path, except for the source node of the demand and the selected destination DC node. Constraint (12) ensures that only one DC is assigned per anycast working lightpath. Constraint (13) guarantees that a DC can be selected for a working lightpath

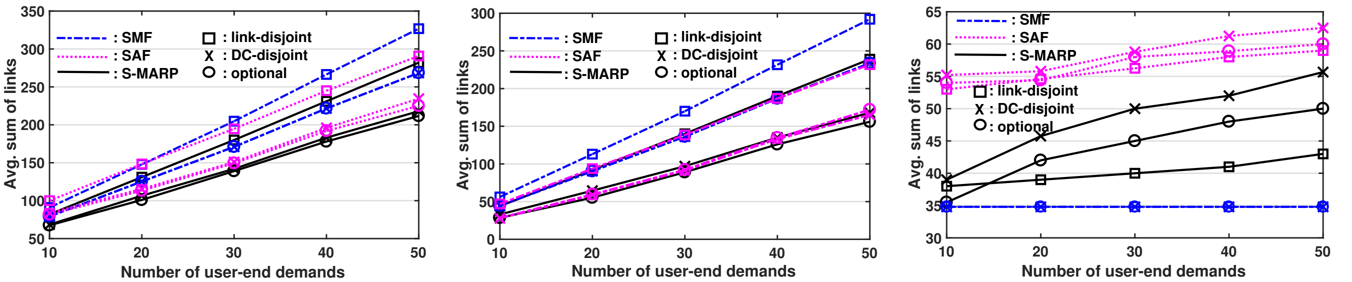
$\sum_{d\in \mathsf{D}} N_d^c = \mathsf{k}, \quad \forall c \in \mathsf{C}$	(2)	$\bar{A}_d^{(s,c)} \le N_d^c,\ \forall (s,c) \in \mathsf{A}, \forall c \in \mathsf{C}, \forall d \in \mathsf{D}$	(13)
$\sum_{w\in \mathsf{W}} x_{(d,j)}^{(d,c),w} \ge 1,\ \forall (d,c) \in \mathsf{M}$	(3)	$\sum_{w\in \mathsf{W}} \bar{\Omega}^{(s,c),w} = 1,\ \forall (s,c) \in \mathsf{A}$	(14)
$\sum_{w\in \mathsf{W}} x_{(i,d)}^{(d,c),w} = 0,\ \forall (d,c) \in \mathsf{M}$	(4)	$y_{(i,j)}^{(s,c),w} + y_{(j,i)}^{(s,c),w} \le \bar{\Omega}^{(s,c),w},$ $\forall (s,c) \in \mathsf{A}, \forall w \in \mathsf{W}, \forall (i,j) : \lvert j\rvert > \lvert i\rvert$	(15)
$\sum_{d\in \mathsf{D}} \big((\sum_{i:(i,d)\in \mathscr{E}} \sum_{w\in \mathsf{W}} \mathsf{K}_d \cdot x_{(i,d)}^{(d,c),w}) - \mathsf{K}_d \cdot N_d^c \big) = 0,$ $\forall (d,c) \in \mathsf{M}$	(5)	$\sum_{i:(i,m)\in \mathscr{E}} \sum_{w\in \mathsf{W}} z_{(i,m)}^{(s,c),w} - \sum_{j:(m,j)\in \mathscr{E}} \sum_{w\in \mathsf{W}} z_{(m,j)}^{(s,c),w} = \begin{cases} -1, m = s \\ \hat{A}_{\hat{d}}^{(s,c)}, \quad \forall \hat{d} \in \mathsf{D}, \forall (s,c) \in \mathsf{A} \\ 0, \text{otherwise} \end{cases}$	(16)
$\sum_{j:(i,j)\in \mathscr{E}} x_{(i,j)}^{(d,c),w} - \lvert \mathscr{V} \rvert \cdot \sum_{j:(j,i)\in \mathscr{E}} x_{(j,i)}^{(d,c),w} \le 0,$ $\forall (d,c) \in \mathsf{M}, \forall w \in \mathsf{W}, \forall i \in \mathscr{V} \setminus \mathsf{d}$	(6)	$\sum_{\hat{d}\in \mathsf{D}} \hat{A}_{\hat{d}}^{(s,c)} = 1,\ \forall (s,c) \in \mathsf{A}$	(17)
$\sum_{j:(j,i)\in \mathscr{E}} x_{(j,i)}^{(d,c),w} - \sum_{j:(i,j)\in \mathscr{E}} x_{(i,j)}^{(d,c),w} \le 0,$ $\forall (d,c) \in \mathsf{M}, \forall w \in \mathsf{W}, \forall i \in \mathscr{V} \setminus \mathsf{D}$	(7)	$\hat{A}_{\hat{d}}^{(s,c)} \le N_{\hat{d}}^c,\ \forall (s,c) \in \mathsf{A}, \forall c \in \mathsf{C}, \forall \hat{d} \in \mathsf{D}$	(18)
$U_i^{(d,c)} - U_j^{(d,c)} + \lvert \mathscr{V} \rvert \cdot x_{(i,j)}^{(d,c),w} \le \lvert \mathscr{V} \rvert - 1,$ $\forall (d,c) \in \mathsf{M}, \forall w \in \mathsf{W}, \forall (i,j) \in \mathscr{E}$	(8)	$\sum_{w\in \mathsf{W}} \hat{\Omega}^{(s,c),w} = 1,\ \forall (s,c) \in \mathsf{A}$	(19)
$\sum_{w\in \mathsf{W}} \Omega^{(d,c),w} = 1,\ \forall (d,c) \in \mathsf{M}$	(9)	$z_{(i,j)}^{(s,c),w} + z_{(j,i)}^{(s,c),w} \le \hat{\Omega}^{(s,c),w},$ $\forall (s,c) \in \mathsf{A}, \forall w \in \mathsf{W}, \forall (i,j) : \lvert j\rvert > \lvert i\rvert$	(20)
$x_{(i,j)}^{(d,c),w} + x_{(j,i)}^{(d,c),w} \le \Omega^{(d,c),w},$ $\forall (d,c) \in \mathsf{M}, \forall w \in \mathsf{W}, \forall (i,j) : \lvert j\rvert > \lvert i\rvert$	(10)	$\sum_{w\in \mathsf{W}} y_{(i,j)}^{(s,c),w} + \sum_{w\in \mathsf{W}} z_{(i,j)}^{(s,c),w} \le 1,$ $\forall (i,j) \in \mathscr{E}, \forall (s,c) \in \mathsf{A}$	(21)
$\sum_{i:(i,m)\in \mathscr{E}} \sum_{w\in \mathsf{W}} y_{(i,m)}^{(s,c),w} - \sum_{j:(m,j)\in \mathscr{E}} \sum_{w\in \mathsf{W}} y_{(m,j)}^{(s,c),w} = \begin{cases} -1, m = s \\ \bar{A}_d^{(s,c)}, \quad \forall d \in \mathsf{D}, \forall (s,c) \in \mathsf{A} \\ 0, \text{otherwise} \end{cases}$	(11)	for $\mathrm{S-MARP-L} \quad (N_d^c = N_{\hat{d}}^c)$ for $\mathrm{S-MARP-DC} \quad (N_d^c \ne N_{\hat{d}}^c)$ for $\mathrm{S-MARP-O} \quad (N_d^c = N_{\hat{d}}^c) \vee (N_d^c \ne N_{\hat{d}}^c)$	(22)
$\sum_{d\in \mathsf{D}} \bar{A}_d^{(s,c)} = 1,\ \forall (s,c) \in \mathsf{A}$	(12)	$\sum_{(d,c)\in \mathsf{M}} x_{(i,j)}^{(d,c),w} + \sum_{(s,c)\in \mathsf{A}} y_{(i,j)}^{(s,c),w} + \sum_{(s,c)\in \mathsf{A}} z_{(i,j)}^{(s,c),w} \le 1,$ $\forall (i,j) \in \mathscr{E}, \forall w \in \mathsf{W}$	(23)

if and only if the requested content is replicated at that DC. Constraints (14,15) enforce the wavelength continuity constraint for the anycast working lightpaths. Similarly, constraints (16-18) ensure the establishment of anycast backup lightpaths while constraints (19-20) inflict the wavelength continuity constraint for the backup lightpaths. Constraint (21) forces the path of the anycast backup lightpath to be link-disjoint with respect to the path of the anycast working lightpath. Constraint (22) selects the appropriate destination DC node for anycast working and backup lightpaths for each type of S-MARP scheme. Finally, constraint (23) enforces the wavelength clash constraint by making sure that each wavelength on a certain network link is allocated to at most one lightpath or light-tree.

III. Numerical Results

The performance of the proposed approaches is evaluated and compared to alternative strategies for manycast and anycast RWA for replica placement that minimize the network resource usage for each type of traffic independently. The first strategy, denoted as Survivable Anycast-First (SAF), first finds an optimal RWA along with replica placement for the survivable anycast demands and then uses this replica placement as input for finding the RWA for the inter-DC traffic. Note, this second step implies solving the multicast RWA problem (not manycast) since the destination nodes are known. The second strategy, denoted as Survivable Manycast-First (SMF), first solves the manycast RWA problem optimally for inter-DC traffic and then uses this replica placement to solve the anycast RWA problem for the survivable user-driven demands. Similar to SMARP, three different cases, i.e., link-disjoint, DC-disjoint, and optional are considered for SAF and SMF. All ILP formulations are solved by running the commercially available CPLEX on the 16 node European topology with 21 bidirectional fiber links, each one supporting 128 wavelengths [11]. 10 test cases were run for each strategy with the values of C, D, and k set to 10, 8, and 3, respectively.

Fig.1(a) shows the overall resource usage obtained by all three S-MARP variants as well as the corresponding versions of SAF and SMF. In all test scenarios, S-MARP-O obtains the lowest resource usage. Compared to the SMF-O and SAF-O, S-MARP-O reduces the resource usage by 16% and 7%, respectively. When backup paths are limited to connect to the same DC as the corresponding working paths, compared to the SMF-L and SAF-L, the S-MARP-L approach reduces the resource usage by 12% and 9% on average, respectively. The savings are comparable for the DC-disjoint approaches, where the average resource usage of S-MARP-DC is by 19% and 12% lower than the one obtained by SMF-DC and SAF-DC, respectively. When different versions of S-MARP are compared, the S-MARP-O performs the best due to the highest degree of flexibility. Interestingly, the S-MARP-DC variant which forces the working and backup paths of a same demand to connect to distinct DCs obtains 22% lower resource usage than S-MARP-L which connects both paths to the same DC. A similar trend is also present for the DC and L variants of the SMF and SAF approaches. This indicates the ability of the algorithm to optimally place content at different DCs so as to provide survivability from DC failures at a lower overall network resource usage.



(a) (b) (c)

Figure 1. (a): Total resource consumption vs. the number of end-user demands; (b): Resource consumption for lightpaths vs. the number of end-user demands; (c): Resource consumption for light-trees vs. the number of end-user demands

Fig.1(b) and Fig.1(c) break down the total resource usage into the resources used for lightpaths serving the user-DC traffic and light-trees serving the DC-DC traffic. SMF obtains the highest lightpath resource usage in all three variants, due to the fact it establishes multicast communication between DCs first, before proceeding to serving the user demands. Consequently, it yields the lowest light-tree resource usage (Fig.1(c)) for all three variants. Conversely, the SAF approach obtains the highest light-tree resource usage (Fig.1(c)) because it begins by minimizing the network resources needed to support user-DC communication. Consequently, it obtains the lowest lightpath resource usage (Fig.1(b)). The proposed S-MARP approach obtains a favorable trade-off between the resources used for user-DC and DC-DC communication compared to the benchmarking approaches that prioritize either type. For instance, S-MARP-O uses the same amount of resources for user-DC lightpaths as SAF-O, while reducing the DC-DC light-tree resource usage by 23%. Compared to the SMF approach, it can also be noticed how S-MARP-O trades off the minimization of the overall resource usage (Fig.1(a)) for a slight increase of the total light-tree length (Fig.1(c)).

IV. Conclusion

This paper studies the problem of manycast and anycast routing and replica placement in datacenter networks to guarantee user demand survivability in the presence of single link and DC failures. By considering three variations of the S-MARP problem, we investigate the tradeoffs between the usage of network resources to support user-DC lightpaths and DC-DC light-trees necessary to protect from link and/or DC failures. Compared to approaches that solve the routing of user-DC and DC-DC traffic subsequently, along with finding the corresponding replica placement, the S-MARP approach is capable of obtaining the lowest overall resource usage.

V. Acknowledgment

This work was jointly supported by the Swedish Research Council (VR) framework grant No. 2014-6230, Celtic-Plus sub-project SENDATE-EXTEND funded by Vinnova, and H2020-ICT-2014 project 5GEx (Grant Agreement no. 671636), and Spanish Grant TEC2014-53071-C3-1-P (ONOFRE).

References

[1] Global Cloud Index: Forecast and Methodology, 2014-2019, *Cisco white paper*.

[2] A. Muhammad, N. Skorin-Kapov, and L. Wosinska, "Manycast and anycast routing for replica placement in datacenter networks," in *Proc. of ECOC*, Sep. 2015.

[3] N. Skorin-Kapov, M. Furdek, S. Zsigmond, and L. Wosinska, "Physical-layer security in evolving optical networks," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 110–117, Aug. 2016.

[4] C. Natalino, A. Yayimli, L. Wosinska, and M. Furdek, "Content accessibility in optical cloud networks under targeted link cuts," in *Proc. of ONDM*, 2017.

[5] R. Goscien, K. Walkowiak, and M. Klinkowski, "Joint anycast and unicast routing and spectrum allocation with dedicated path protection in elastic optical networks," in *Proc. of RNDM*, Mar. 2014.

[6] X. Li, S. Huang, S. Yin, B. Guo, Y. Zhao, J. Zhang, M. Zhang, and W. Gu, "Shared end-to-content backup path protection in k-node (edge) content connected elastic optical datacenter networks," *Optics Express*, vol. 24, no. 9, pp. 9446–9464, Apr. 2014.

[7] M. Habib, M. Tornatore, M. De Leenheer, F. Dikbiyik, and B. Mukherjee, "Design of disaster-resilient optical datacenter networks," *Journal of Lightwave Technology*, vol. 30, no. 16, pp. 2563–2573, Aug. 2012.

[8] C. Natalino, P. Monti, L. Franca, M. Furdek, L. Wosinska, C. R. Frances, and J. W. Costa, "Dimensioning optical clouds with shared-path shared-computing (spsc) protection," in *Proc. of HPSR*, Jul. 2015.

[9] R. Goscien and K. Walkowiak, "Modeling and optimization of data center location and routing and spectrum allocation in survivable elastic optical networks," *Optical Switching and Networking*, vol. 23, pp. 129–143, 2017.

[10] D. A. P. Davis, J. M. Plante, and V. M. Vokkarane, "Critical resource multicast protection in data center networks," in *Proc. of ICC*, May 2015.

[11] T. O. A. Drakos and A. Stavdas., "Performance benchmarking of core optical networking paradigms," *Optics Express*, vol. 20, pp. 17 421–17 439, 2012.