A Disaster-Resilient Multi-Content Optical Datacenter Network Architecture

M. Farhan Habib[1], Massimo Tornatore[21], Marc De Leenheer[31], Ferhat Dikbiyik[4] and Biswanath Mukherjee[1]

[1]*Department of Computer Science, University of California Davis*
[2]*Department of Electronics and Information, Politecnico di Milano*
[3]*Department of Information Technology, Ghent University- IBBT*
[4]*Department of Electrical and Computer Engineering, University of California Davis*
{mfhabib, mleenheer, fdikbiyik, bmukherjee}@ucdavis.edu, tornator@elet.polimi.it

ABSTRACT

Cloud services based on datacenter networks are becoming very important. Optical networks are well suited to meet the demands set by the high volume of traffic between datacenters, given their high bandwidth and low-latency characteristics. In such networks, path protection against network failures is generally ensured by providing a backup path to the same destination, which is link-disjoint to the primary path. This protection fails to protect against disasters covering an area which disrupts both primary and backup resources. Also, content/service protection is a fundamental problem in datacenter networks, as the failure of a single datacenter should not cause the disappearance of a specific content/service from the network. Content placement, routing and protection of paths and content are closely related to one another, so the interaction among these should be studied together. In this work, we propose an integrated ILP formulation to design an optical datacenter network, which solves all the above-mentioned problems simultaneously. We show that our disaster protection scheme exploiting anycasting provides more protection, but uses less capacity, than dedicated single-link protection. We also show that a reasonable number of datacenters and selective content replicas with intelligent network design can provide survivability to disasters while supporting user demands.

Keywords: Datacenter Network, Optical Network, Content Protection, Network Survivability, Network Optimization.

1. INTRODUCTION

Even though protection in optical mesh networks has been well studied, network survivability has mainly focused on single-link failure events. A few multi-failure studies have been conducted, focusing on dual-link failures. Protection against cascading and correlated multiple link/node failures is a topic that needs serious attention. Multiple failures generally occur due to natural disasters such as earthquake, hurricane, tsunami, tornado, etc. or human-made disasters such as weapons of mass destruction (WMD) and electromagnetic pulse (EMP) attacks [1]. Such disasters affect specific geographic areas; as a result, a set of collocated nodes and links go down simultaneously e.g., the 2011 earthquake and tsunami in Japan and the 2008 China Sichuan earthquake caused massive damage to telecom networks in large geographic areas. These events indicate that it is crucial to study disaster protection mechanisms for communication networks. In this work, we focus on the problem of assigning network resources to connection requests and placing heterogeneous contents in geo-diverse locations in datacenter networks for disaster protection.

Typically, datacenters host computing and storage resources, and the resources are served to customers through a network of datacenters, popularly referred to as the cloud. In such a network, heterogeneous (i.e., different) contents and/or services are replicated over multiple datacenters, so that a user request can be served by any datacenter that supports the specified content and/or service, to optimize availability and performance. Most popular web applications are served by such datacenter networks, while other forms of cloud computing include distributed grid computing [2] and some mission-critical applications with high bandwidth requirements. To meet the demands set by the high volume of traffic between datacenters, optical networks are ideally suited, given their high-bandwidth and low-latency characteristics. In this study, we consider a datacenter network with a fiber-optic backbone that provides circuit-switched paths for high-bandwidth connection requests.

Traditionally, network protection against single-link failures is ensured by providing a backup path to the same destination (i.e., datacenter), which is link-disjoint to the primary path. This scheme has been refined by the introduction of a backup datacenter [2], thereby adding protection against failures of a single datacenter (destination node). However, this protection scheme fails to protect against disasters located in an area that contains both primary and backup resources (either network links or datacenters). Also, content/service protection is another fundamental problem in datacenter networks, as the failure of a single datacenter should not cause the disappearance of a specific content/service from the whole network. Thus, three problems, namely content/service placement, path protection, and content/service protection, should be addressed simultaneously. A Shared Risk Group (SRG) is defined as a set of nodes and links which might be affected simultaneously by a

single disaster event. To provide disaster protection, primary and backup paths as well as multiple locations of content/services should be SRG-disjoint.

In this study, we propose a method to design datacenter networks while providing disaster survivability. Our formulation solves the following problems simultaneously: content/service placement (i.e., replication), as well as routing and disaster protection for both paths and content. Although previous work has investigated resiliency in optical datacenter networks [2][3], none of them consider content/service placement or disaster survivability.

The rest of the study is organized as follows. Section 2 describes the proposed model for resource assignment and content placement providing content and path protection from disaster in a datacenter network. Section 3 presents some illustrative examples and discusses the results. Section 4 concludes the study.

2. A MODEL FOR RESOURCE ASSIGNMENT AND CONTENT PLACEMENT IN A DATACENTER NETWORK

We consider a circuit-switched optical datacenter network. Optical cross-connects are assumed to be opaque as in practice today, so there is no wavelength-continuity constraint. We formally state the problem of assigning paths to high-bandwidth connections, content replica placement, and providing disaster protection for both paths and contents, and formulate the problem using an integer linear program (ILP) as shown below.

Given:

- $G=(V,E)$: Physical topology, where V is the set of nodes and E is the set of directed links.
- $V' \subset V$: Set of datacenter locations.
- D: Set of shared risk groups (SRGs).
- C': Set of contents.
- T: Set of requests (s, c), where s is the source node and c is the content; $s \in V, c \in C'$.
- C: Capacity of link (i,j). M, K: Constant.

Variables:

- $P_{(i,j)}^{(s,c)} \in \{0,1\}$: link ($i$, j) is used in the primary path for request (s, c).
- $B_{(i,j)}^{(s,c)} \in \{0,1\}$: link ($i$, j) is used in the backup path for request (s,c).
- $A_d^{(s,c)} \in \{0,1\}$: $d \in V'$ is used as the primary datacenter for request (s,c).
- $\overline{A}_d^{(s,c)} \in \{0,1\}$: $d \in V'$ is used as the backup datacenter for request (s,c).
- $R^{(c,d)} \in \{0,1\}$: content $c \in C'$ is replicated at datacenter $d \in V'$.
- $\pi_{(i,j)} \in N$: total number of wavelengths in link (i,j) used for backup paths.
- $\alpha_x^{(s,c)} \in \{0,1\}$: primary path for request (s,c) goes through SRG $x \in D$.
- $\beta_x^{(s,c)} \in \{0,1\}$: backup path for request (s,c) goes through SRG $x \in D$.
- $\beta_{(i,j),x}^{(s,c)} \in \{0,1\}$: link ($i$,$j$) is used in the backup path for request (s,c) if the primary path is down due to a disaster occurring at SRG $x \in D$.

Objective:

$$\min\left(\sum_{(i,j)}\pi_{(i,j)} + \sum_{(i,j)}\sum_{(s,c)}P_{(i,j)}^{(s,c)}\right)$$

Constraints:

$$\sum_{j:(i,j)\in E}P_{(i,j)}^{(s,c)} - \sum_{j:(j,i)\in E}P_{(j,i)}^{(s,c)} = \begin{cases} 1, \ if\ i == s \\ -A_i^{(s,c)}, \ if\ i \in V' \\ 0, \ otherwise \end{cases} \quad \forall_{(s,c)\in T}, \forall_{i\in V} \tag{1}$$

$$\sum_{j:(i,j)\in E}B_{(i,j)}^{(s,c)} - \sum_{j:(j,i)\in E}B_{(j,i)}^{(s,c)} = \begin{cases} 1, \ if\ i == s \\ -\overline{A}_i^{(s,c)}, \ if\ i \in V' \\ 0, \ otherwise \end{cases} \quad \forall_{(s,c)\in T}, \forall_{i\in V} \tag{2}$$

$$\sum_{d\in V'}A_d^{(s,c)} = 1, \forall_{(s,c)\in T} \tag{3}$$

$$\sum_{d\in V'}\overline{A}_d^{(s,c)} = 1, \forall_{(s,c)\in T} \tag{4}$$

$$A_d^{(s,c)} + \overline{A}_d^{(s,c)} \le 1, \forall_{(s,c)\in T}\forall_{d\in V'} \tag{5}$$

$$R^{c,d} \ge A_d^{(s,c)} + \overline{A}_d^{(s,c)}, \forall_{d\in V'}\forall_{(s,c)\in T} \tag{6}$$

$$\sum_{(s,c)}P_{(i,j)}^{(s,c)} + \pi_{(i,j)} \le C, \forall_{(i,j)\in E} \tag{7}$$

$$\sum_{d\in V'}R^{c,d} \le K, \forall_{c\in C'} \tag{8}$$

$$\frac{\sum_{(i,j)\in x} P_{(i,j)}^{(s,c)}}{M} \le \alpha_x^{(s,c)} \le \sum_{(i,j)\in x} P_{(i,j)}^{(s,c)}, \forall_{(s,c)\in T} \forall_{x\in D} \quad (9)$$

$$\frac{\sum_{(i,j)\in x} B_{(i,j)}^{(s,c)}}{M} \le \beta_x^{(s,c)} \le \sum_{(i,j)\in x} B_{(i,j)}^{(s,c)}, \forall_{(s,c)\in T} \forall_{x\in D} \quad (10)$$

$$\alpha_x^{(s,c)} + \beta_x^{(s,c)} \le 1, \forall_{(s,c)\in T} \forall_{x\in D} \quad (11)$$

$$\pi_{(i,j)} \ge \sum_{(s,c)} B_{(i,j),x}^{(s,c)}, \forall_{(i,j)\in E} \forall_{x\in D} \quad (12)$$

$$\beta_{(i,j),x}^{(s,c)} \le \alpha_x^{(s,c)}, \forall_{(s,c)\in T} \forall_{x\in D} \forall_{(i,j)\in E} \quad (13)$$

$$\beta_{(i,j),x}^{(s,c)} \le B_{(i,j)}^{(s,c)}, \forall_{(s,c)\in T} \forall_{x\in D} \forall_{(i,j)\in E} \quad (14)$$

$$\beta_{(i,j),x}^{(s,c)} \ge \alpha_x^{(s,c)} + B_{(i,j)}^{(s,c)} - 1, \forall_{(s,c)\in T} \forall_{x\in D} \forall_{(i,j)\in E} \quad (15)$$

The first term of the objective function minimizes the shared backup resources, and the second term minimizes the primary resources. We have 15 sets of constraints shown in Eqn. (1)-(15). Equations (1) and (2) enforce flow conservation for primary and backup paths, respectively. Equations (3) and (4) make the assignment of a datacenter for both of the primary and backup paths, respectively. Equation (5) ensures that the same datacenter is not used for both primary and backup paths of a request. Equation (6) does content replica placement. Equation (7) is the capacity constraint. Equation (8) bounds the number of replicas, where *k* is the maximum number of replicas per content. Equations (9) and (10) set the value of $\alpha_x^{(s,c)}$ and $\beta_x^{(s,c)}$, respectively (here M is a large integer). $\alpha_x^{(s,c)}$ $\left(\beta_x^{(s,c)}\right)$ is 1 if the primary (backup) path for request (*s*, *c*) goes through SRG x. Equation (11) ensures the SRG-disjoint property of primary and backup paths. Equations (13)-(15) set values for $\beta_{(i,j),x}^{(s,c)}$, which is 1 if the primary path for request (*s, c*) goes through SRG *x* and the backup path uses link (*i, j*). Equation (12) bounds the number of wavelengths used in a link for shared protection. $\pi_{(i,j)}$ denotes the total number of wavelengths in link (*i, j*) that is shared by multiple backup paths if a disaster happens. The right hand side denotes the number of backup paths that goes through link (*i, j*) if the corresponding primary paths are down due to a disaster in SRG *x*. As primary and backup paths for a connection are SRG-disjoint and connect to two different datacenters, it ensures that content is replicated to disaster disjoint datacenters.

3. ILLUSTRATIVE NUMERICAL EXAMPLES

We present illustrative results by solving our ILP on NSFNet and COST239 networks as shown in Fig. 1. A study shows that a disaster zone (DZ) can span upto 160 km [4]. Following this, we specify 14 DZs for NSFNet and 7 DZs for COST239. Figure 2 compares the wavelength usage for shared DZ protection with dedicated and shared single-link (SL) protection. Note that, also for SL protection, the primary and backup paths connect to different datacenters to avoid destination node failure, though it cannot protect against disaster failure. NSFNet has datacenters at nodes 2, 6, and 9; and COST239 has datacenters at nodes 4, 5, and 9. We see that DZ protection uses more wavelengths than shared SL protection but fewer wavelengths than dedicated SL protection. Dedicated SL protection has more probability of being survivable in case of multiple random link failures than shared SL protection. But, in reality, failures of multiple non-correlated links are very unlikely. Rather, it is more likely that a set of correlated links/nodes are down simultaneously due to a disaster. Table 1 shows that, without DZ protection, significant numbers of connections are vulnerable to disaster failures, even though primary and backup datacenters are different in both dedicated and shared SL protections in this example. These results indicate that DZ protection, though it uses fewer wavelengths, provides more protection against disasters than dedicated SL protection. Table 1 also shows that numbers of paths vulnerable to disasters in SL protection are higher in COST239 than in NSFNet because COST239 is denser with shorter links than NSFNet. The more dense a network is, the more vulnerable it is to a disaster failure.

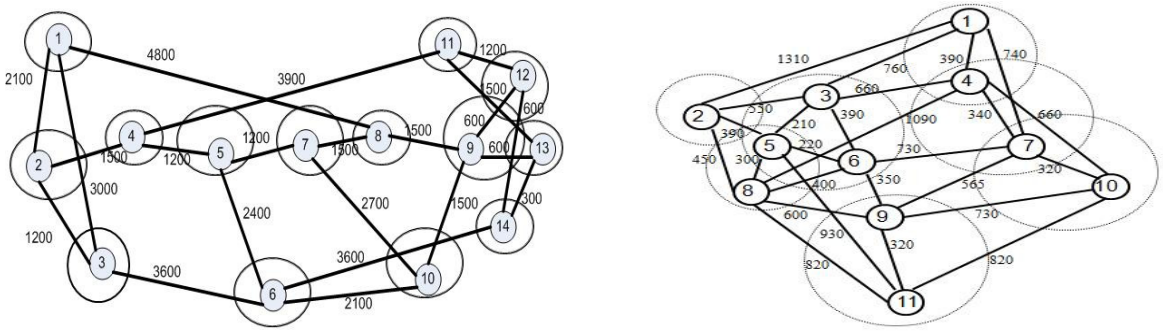

Figure 1: NSFNet (left) and COST239 (right) with link lengths in km. SRGs are shown in circles.

Figure 3 shows the effect of number of content replicas on wavelength usage for shared and dedicated protection in NSFNet. Datacenter locations are 2, 5, 9, and 11. Note that, with a small increase in wavelength usage, number of replicas can be decreased significantly. Based on user demands, replicas are distributed through the network which allows flexibility to choose primary and backup datacenters. More replicas do not always provide more flexibility to choose a shorter path, but more replicas mean more usage of storage resources.

Table 2 shows the effect of number of datacenters on wavelength usage in shared protection with unconstrained number of replicas in NSFNet. The number of wavelengths reduces significantly as the number of datacenters increases, but after a certain value, increasing the number of datacenters does not help much to reduce wavelength utilization. Thus a reasonable number of datacenters with intelligent network design can provide survivability to disasters while supporting user demands. Our model can help network designers to decide on the number of datacenters and replicas that would achieve optimal performance. This result also shows that survivability to disasters does not require having many datacenters.

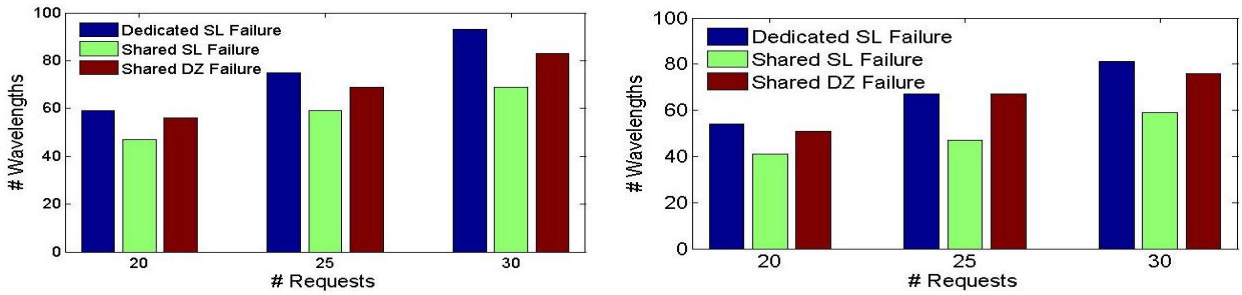

Figure 2: Total wavelength usage for three protection schemes in (a) NSFNet and (b) COST239.

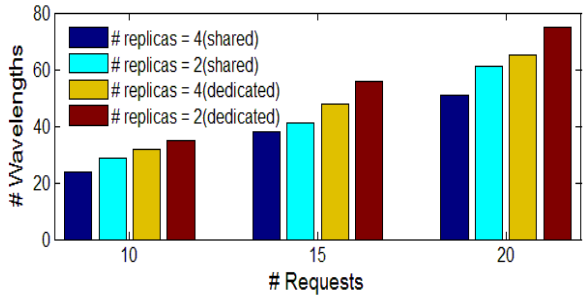

Figure 3: Wavelength usage for different number of replicas in NSFNet (number of datacenters = 4).

Table 1: Number of unprotected paths from a disaster

Protection Scheme	Number of connections					
	NSFNet			***COST239***		
	20	25	30	20	25	30
Dedicated SL Failure	4	4	5	7	8	9
Shared SL Failure	4	2	6	6	7	13
Shared DZ Failure	0	0	0	0	0	0

Table 2: Total wavelength usage for varying number of datacenters in shared protection in NSFNet

Number of Datacenters	2	3	4	5
Number of Wavelengths	62	45	37	33

4. CONCLUSION

We presented a model to design optical datacenter networks. Our formulation supports contents with different characteristics and provides survivability to both paths and content from disasters. We found that disaster protection scheme provides more protection, but uses fewer wavelengths than dedicated single-link protection. We also showed that a reasonable number of datacenters and content replicas with intelligent network design can provide survivability to disasters while supporting user demands.

ACKNOWLEDGEMENT

This work has been supported by the Defense Threat Reduction Agency (DTRA) Program "Network Adaptability from WMD Disruption and Cascading Failures".

REFERENCES

[1] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano: Assessing the Vulnerability of the Fiber Infrastructure to Disasters, *in Proc. IEEE INFOCOM*, Brazil, April 2009.

[2] J. Buysse, M. De Leenheer, C. Develder, and B. Dhoedt: Exploiting Relocation to Reduce Network Dimensions of Resilient Optical Grids, *in Proc. 7th International Workshop on Design of Reliable Communication Networks,* Washington DC, October 2009.

[3] C. Develder, B. Dhoedt, B. Mukherjee, and P. Demeester: On Dimensioning Optical Grids and the Impact of Scheduling, *in Photonic Network Communications,* vol. 17, no. 3, pp. 255-265, 2009.

[4] T. L. Weems: How Far is Far Enough, *in Disaster Recovery Journal*, vol. 16, no. 2, Spring 2003.