# Disaster-aware service provisioning with manycasting in cloud networks

**S. Sedef Savas · Ferhat Dikbiyik · M. Farhan Habib · Massimo Tornatore · Biswanath Mukherjee**

Received: 17 April 2014 / Accepted: 16 June 2014 / Published online: 7 September 2014

## 1 Introduction

The rapid growth of broadband communications has led to many new web applications such as online interactive maps, social networks, video streaming, cloud computing, and Con-tent Distribution Network (CDN) services, most of which are provided by cloud networks which are composed of several datacenters. A datacenter is a warehouse-scale and massively parallel computing and storage resource. It generally consists of thousands of clustered servers. Cloud-based applications are reshaping the network landscape, by pushing the tradi-tional hierarchical and connectivity-oriented Internet toward a more flat and service-oriented infrastructure as promoted in Web 2.0 [1]. Accordingly, networks provide more direct connections from content/service providers to customers, with services delivered through datacenters. Cloud services require a transmission infrastructure with high capacity, low latency, low cost, and high availability. Optical network-ing technology plays an important role in realizing next-generation cloud solutions in order to meet their stringent requirements [2].

High-capacity optical backbone networks that carry cloud services are exposed to many threats such as malicious attacks, equipment failures, human errors (e.g., misconfig-urations), and large-scale disasters, both human made [e.g., due to weapons of mass destruction (WMD) attacks] and natural. Disasters represent a challenging threat for our net-

S. S. Savas (✉) · M. F. Habib · M. Tornatore · B. Mukherjee
University of California, Davis, CA, USA
e-mail: ssavas@ucdavis.edu

M. F. Habib
e-mail: mfhabib@ucdavis.edu

B. Mukherjee
e-mail: bmukherjee@ucdavis.edu

F. Dikbiyik
Sakarya University, Sakarya, Turkey
e-mail: fdikbiyik@sakarya.edu.tr

M. Tornatore
Politecnico di Milano, Milan, Italy
e-mail: tornator@elet.polimi.it

This work has been supported by the Defense Threat Reduction Agency (DTRA) Grant HDTRA1-10-1-0011. Ferhat Dikbiyik was a Ph.D. student at UC Davis when most of this work was performed.

works as they affect large geographical areas and can trigger correlated and/or cascading failures of multiple network elements, resulting in huge data loss and disruptions in network connectivity. The 2011 Japan Earthquake/Tsunami and 2012 Hurricane Sandy [3] are two recent disasters that have deprived people of essential network services and severely hindered rescue operations for weeks. Moreover, multiple link/node failures caused by disasters may make the network more vulnerable to secondary (post-disaster) failures (e.g., aftershocks or power outages). Post-disasters can be correlated (e.g., an earthquake on a fault line can trigger other earthquakes on the same line) or uncorrelated (sequential WMD attacks), and they need serious attention. Thus, disaster-aware provisioning schemes (e.g., routing around risky disaster areas) have been proposed. Service requirements of bandwidth-hungry cloud services make disaster survivability even more crucial as the data (and revenue) loss caused by large-scale correlated, cascading failures can be very high. To alleviate their impact, new measures should be taken as emerging cloud services make different service paradigms possible since one-to-one connectivity is not required anymore.

Contents or services in cloud systems can be replicated across multiple datacenters located at different nodes from which user demands can be served. As opposed to the traffic that explicitly specifies both source and destination (unicast traffic), cloud networks make new service paradigms feasible which include anycasting, i.e., providing service to cloud users from any of the datacenters that host the service, and manycasting, i.e., providing service from a subset of the datacenters. These models can be applied for services such as file transfer and media streaming [4]. In this work, we propose a novel disaster-aware service-provisioning scheme that multiplexes service over multiple paths to multiple servers/datacenters with manycasting, and the main contribution is to assess the additional resiliency manycasting offers by allowing serving a cloud user from an intelligently selected subset of datacenters that have the requested content/service. Network-survivability studies mostly focus on link failures and neglect datacenter failures. Our solution offers protection against node failures by selecting shared-risk group (SRG whose members may fail simultaneously due to a disaster [5,6]) disjoint datacenters to serve customers. Without changing anything in application layer, manycasting is possible through protocols such as Hypertext Transfer Protocol (HTTP), which makes the proposed solution practical.

Providing 100 % protection against disasters (by routing them via primary and backup paths) would require massive and economically unsustainable bandwidth overprovisioning, as disasters are difficult to predict, statistically rare, and may create large-scale failures. Some works (e.g., [7]) have showed that multipath provisioning outperforms single-path provisioning in terms of flexibility in usage of network resources. Many cloud applications are throughput tolerant and can continue to operate under reduced capacity (e.g., bandwidth), which results in a lower service/perceived quality (e.g., for file delivery, transfer may take longer; for video conferencing, video frames may have lower quality/resolution with reduced capacity, etc.) Thus, service can be maintained with a degraded-service level vs. no service at all [8]. These applications can exploit multipath provisioning, which offers degraded service in case of a failure (partial protection). However, single or multipath provisioning to a single destination (datacenter) does not ensure protection against service failures at datacenters (i.e., node failures). To remedy this problem, Ref. [9] proposes the establishment of primary and backup paths to SRG-disjoint datacenters but it does not benefit from multipath provisioning as only one path is active at a time.

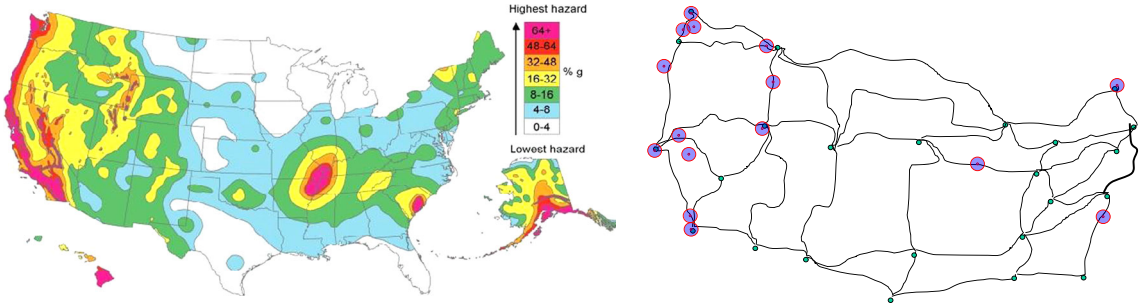
Our solution combines the benefits of multipath and anycast provisioning schemes during the delivery of cloud services. Our proposed manycasting protection offers (1) degraded service in case of large-scale failures due to disasters as in multipath provisioning and (2) resilience to datacenter-node failures as in using a backup-datacenter approach. We consider minimizing the risk of disaster failures, i.e., the loss in case of a disaster by avoiding high-risk paths [10]. Our approach has high chance of successful provisioning, because less bandwidth is requested from the paths and datacenters, and it is inherently survivable to single-link failures (provides degraded service; and, with overprovisioning, it can provide full protection against single failures). Also, instead of only utilizing the most secure paths, it also uses other paths and reduces the possibility of congestion on the secure paths.

The rest of this study is organized as follows. Section 2 presents a brief background on risk calculation for disasters, protection schemes, and related works. We describe the proposed scheme at a high level and state the problem in Sect. 3, and formulate the optimal ILP in Sect. 4. Since ILP is intractable for large problem instances, we propose heuristics in Sect. 5. Section 6 presents numerical results, and we conclude the study in Sect. 7.

## 2 Background

Disaster repercussions, e.g., disconnections, data loss, and service disruptions, can be minimized using survivable provisioning schemes (pre-configured, before disasters) and restoration schemes (reactive, after disasters).

Some studies propose survivable provisioning to proactively alleviate the impact of disasters. They determine possible disaster zones in the network, e.g., risk (hazard) maps to highlight its vulnerable regions using interdisciplinary con-




(a) (b)

**Fig. 1** Exploitation of hazard maps to determine disaster zones: **a** US seismic-hazard map, and **b** US-wide fiber cable topology with high-risk disaster zones (by matching fiber map with seismic-hazard map). The hazard levels below 32 % are neglected. *Circles* represent disaster zones and their center is at the epicenter of the predicted earthquakes, and their radius is 100 miles as damage of an earthquake may span this distance [12]

tributions from climatology, geology, etc. (e.g., Fig. 1a shows risky zones of US with a heat map against earthquake disaster) [11]. By utilizing risk maps, we can estimate the probability of occurrence of a disaster and probability of a network device getting damaged by this disaster. These two parameters give us the risk levels of disaster zones. Figure 1a shows seismic-hazard map of US with its risk levels. Locations of high-risk earthquake zones (Fig. 1b), which have different probabilities of failures, can be determined by matching a network topology with the seismic-hazard map. Using these maps, network planners can develop solutions that select less risky regions for routing connections, and hence, the expected loss will be minimized and network becomes better prepared to handle a disaster.

The set of links or nodes that are vulnerable to a common failure (e.g., a disaster) can be represented as a SRG [6]. The most prevalent protection strategy against disasters is to route connections over disaster-zone-disjoint (i.e., SRG disjoint) primary and backup paths (or using multiple primary paths, e.g., multipath provisioning) as depicted in Fig. 2a, where 1B is the requested bandwidth. However, fully protecting primary paths with backups requires extensive resource usage, especially for multiple failures (as in disasters).

Some services may accept a reduced level of bandwidth during failures, depending on their characteristics. For services that can tolerate reduced bandwidth, network operators may offer partial protection, possibly at lower cost. The partial-protection guarantee is determined by the connection's degraded-service tolerance, e.g., backup path in Fig. 2a can be downgraded to the connection's minimum service requirement, so it becomes partially protected.

Multipath routing, i.e., multiplexing a connection over multiple paths, is another scheme for providing partial protection [13]. For a multi-path-routed service, even if some paths are down or overloaded, the other paths may provide the required degraded service. Thus, some SLAs for partial protection can be satisfied without any redundant resource allocation. Figure 2b shows an example degraded-service-aware multipath provisioning [14], which guarantees a minimum tolerable bandwidth of 60 % of the required bandwidth under normal operation even in case of failure of one of the paths. During normal operation, customer receives 1 B, which is multiplexed over three paths (one with 0.4 B, two with 0.3 B) and any predicted disaster in the network, affects only one path, hence the guaranteed bandwidth is at least 0.6 B.

The shifting paradigm toward cloud computing is creating new opportunities for optimizing disaster-aware network design. Contents in cloud systems are replicated at multiple servers/datacenters and a user's service need not be confined to one particular datacenter. New service models are enabled such as anycasting, i.e., providing service from any of the datacenters that hosts the requested service, and manycasting, i.e., providing service from a subset of the datacenters. These models can be exploited, e.g., for file transfer and media streaming, to enhance the resilience of cloud services. Resilience against destination-node failures is very crucial due to cloud services and datacenters hosting content. Backup-path protection using anycasting is shown in Fig. 2c. Authors in [15] propose a model for placing datacenters and routing services under the anycast model, such that fast cloud service protection is considered in case of single-link and single-node failure events. Risk concept has not been considered in these works, and they perform routing by shortest-path calculation.

Our proposed disaster-aware provisioning scheme exploits the manycasting service paradigm enabled by cloud services which is depicted in Fig. 2d. We use inverse multiplexing over multiple paths (the least risky ones) to provision bandwidth for services distributed over multiple servers/datacenters with manycasting. Also, it ensures degraded service (vs. no service at all) after a failure without using extra resources since it uses multipath routing. Both anycasting scheme in Fig. 2c and our proposed manycasting scheme in Fig. 2d are resilient against destination-node failures (since the paths

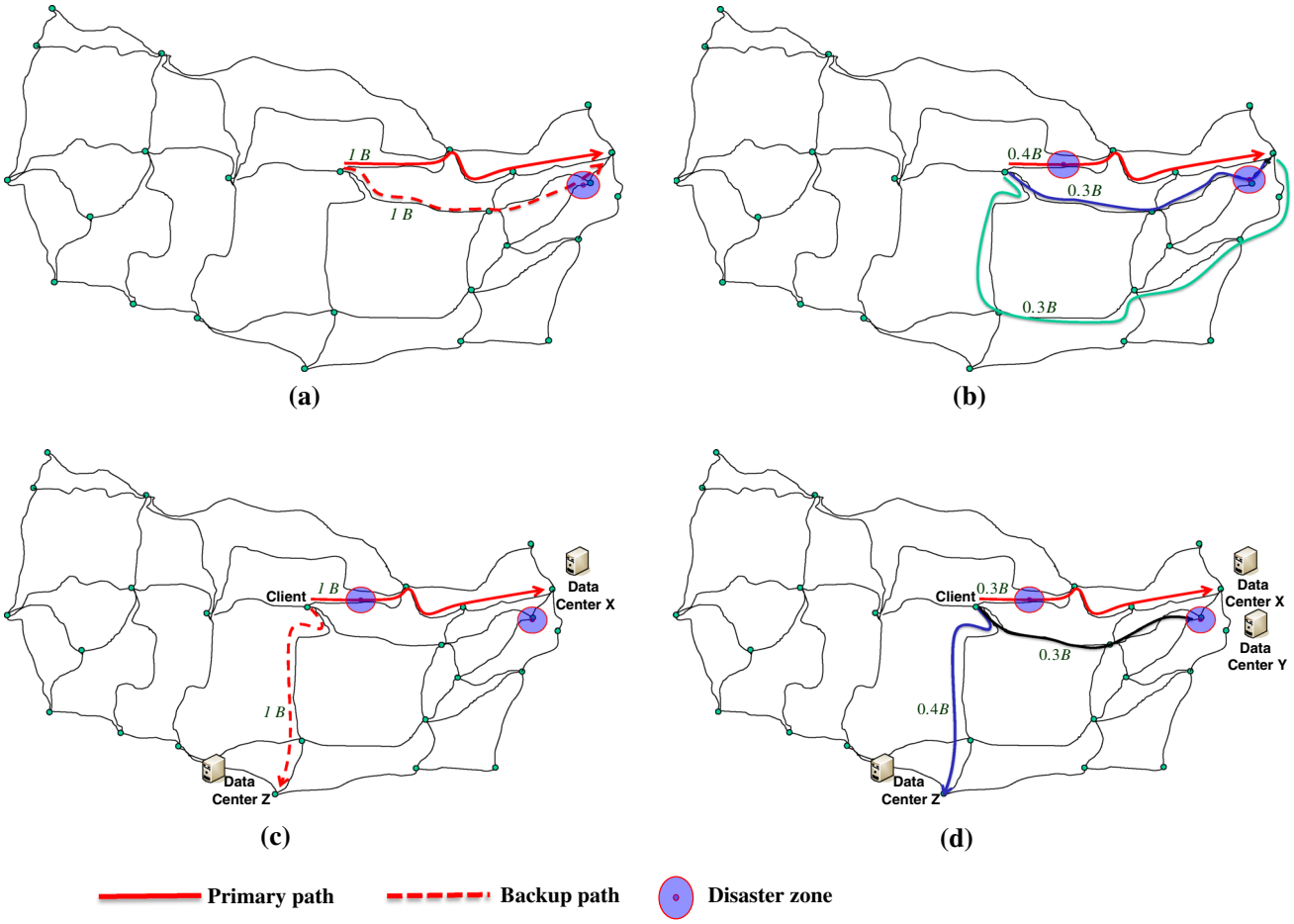



**Fig. 2** Disaster-aware provisioning schemes: **a** full protection via backup path (Partial protection can be realized by lowering bandwidth of backup path), **b** partial protection via multipath routing, **c** full protection via anycasting (partial protection also possible), and **d** proposed manycasting scheme with partial protection

connect to disaster-zone-disjoint destination nodes, so the service will not be disrupted if such a node fails due to a disaster).

Manycasting is not a new concept, and it has been discussed in several works in the context of group communication that enables communication with an arbitrary (client-specified) number of group members. Anycast and multicast communication are special cases of manycast in which the target number of group members is one and infinity, respectively [16]. However, in this study, we explore manycast provisioning for cloud computing in the context of disaster survivability which, to the best of our knowledge, is the first study to focus on this issue. We present disaster-aware manycast provisioning, with specific attention to high-capacity backbone networks. Compared with the classic multipath routing problem, our proposed solution is much more complex since this approach introduces new constraints such as SLAs to satisfy delay requirement, and only the datacenters that do not violate that amount can be used to get the service. Also, risk, resource usage, and service delay parameters are highly dependent on the network and datacenter placement. For a connection request, optimizing selected paths both considering their risk levels and distances between source and destination nodes of that path (a.k.a. resource usage and latency) may be challenging since a low-risk path can be established to a distant datacenter while there are closer datacenters to the source.

## 3 Overview of the proposed approach and problem statement

### 3.1 Overview of the proposed approach

We propose to provision multiple paths, which are the least risky ones, to multiple datacenters, which we call Multipath to Multiple Destinations (MMD). We define risk using a probabilistic model (i.e., we consider disaster occur-

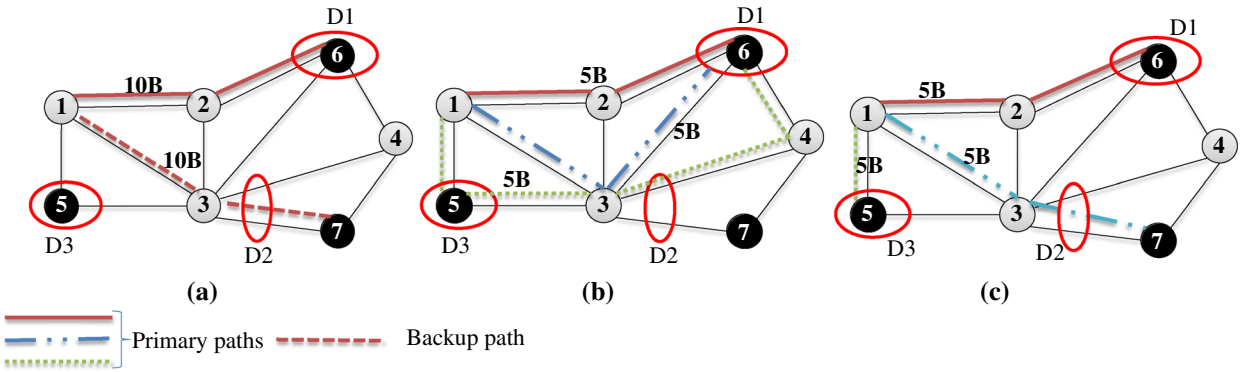



**Fig. 3** Different risk-aware provisioning schemes. **a** backup path to backup dest. (BBD), **b** multipath to single destination (MSD), **c** proposed multipath to multiple destination (MMD)

rence probabilities as well as the probability of the network resources being affected by a disaster, as in [10,17]).

We analyze the network with a probabilistic disaster model, where a network equipment in a disaster zone fails with some probability, which depends on its distance from the disaster's epicenter, type of the disaster, etc. [16]. A network equipment fails with a probability which depends on many factors such as the equipment's dimension (e.g., length of a link), specification (i.e., shock resistance), distance from the attack center, and functional dependencies on other components, e.g., power station.

The main benefits and properties of the proposed scheme are as follows. First, it provides protection against secondary failures (e.g., aftershocks) along with single-link and disaster failures by provisioning multiple SRG-disjoint paths. When a connection loses one path after a disaster, it will still have multiple SRG-disjoint paths that protect it against post-disaster failures. Second, it is resilient against destination-node failures (since we have multiple destination nodes, the service will not be disrupted if one of them fails due to a disaster). Third, this scheme can ensure degraded service (vs. no service at all) after a failure without using extra resources. The connection can also be partially protected according to customer needs, and this partial protection amount will be defined in the service request. The topology of the network and the bandwidth requirement of a connection might limit the amount of partial protection; yet this model can offer the requested degraded service level, if possible. Another important aspect of this scheme is that it provides a minimum degraded-service ratio which is defined in a connection request by the customers. It is the minimum level of partial protection ensured even if just a single path survives after disasters. This ratio assures a certain level of acceptable protection when post-disaster failures occur since each path provisions at least that amount. To reduce latency among different paths, the paths in our scheme only use datacenters within a certain distance away from the customers, which is generally referred to as differential delay constraint (DDC) [18].

We compare our scheme with two benchmark schemes, namely multipath to single destination (MSD) and Backup path to Backup Destination (BBD). In MSD, multiple paths that use a single datacenter are provisioned for a request, i.e., multipath provisioning. In BBD, one primary and one backup path are provisioned for a connection in which both the paths and their destination nodes are SRG disjoint.

Figure 3 illustrates three different risk-aware provisioning schemes on a 7-node network with datacenters located at nodes 5, 6 and 7; and three disaster zones (DZs) D1, D2, and D3. We consider a content request from node 1 with 10B requested bandwidth and full protection in case of a disaster, and all datacenters have the requested content. In Fig. 3a, a backup path is provided to datacenter at node 7, which is SRG disjoint to the primary datacenter. When any disaster occurs, this scheme still provides 10B bandwidth but consumes 20B bandwidth. Backup-path scheme becomes vulnerable to single-link failures after either a disaster in D1 or D2 occurs (not survivable against post-disaster failures). Figure 3b illustrates multipath provisioning to a single datacenter. The connection will be lost if a disaster in D1 occurs while it will be robust to other disasters. Due to this instability, we cannot guarantee a certain amount of service. This scheme also provides degraded service. Figure 3c shows the proposed solution which provisions multipaths to multiple datacenters. After any one of the disasters occurs, the service remains survivable against post-disasters only in this scheme. Also, the proposed scheme is robust against multiple disasters unless all the SRG-disjoint paths get disconnected.

### 3.2 Problem statement

We consider a WDM optical backbone network where datacenters can be placed at a selected subset of network nodes. Contents are not fully replicated (where every content exists in all datacenters) since it may create more background traffic between datacenters due to synchronization.

Disaster-aware manycast routing can be stated as follows: *Given the network topology with nodes and fiber links interconnecting them, including source sites (where requests for cloud services originate), candidate destinations (i.e., the cloud servers), and disaster risk map of the network, find multipaths to multiple datacenters for each of the requests such that the risk (i.e., expected bandwidth loss) of the network during a disaster event is minimized.*

## 4 Mathematical formulation

The mathematical formulation of multipath provisioning to multiple datacenters is formally described below, and it turns out to be an Integer Linear Program (ILP). Datacenter locations and content placement are given as input. At most one path will be established to a datacenter for a connection. We consider k-shortest paths from each node to each datacenter as an input. The primary objective is to find a set of paths for all connection requests that minimizes the expected bandwidth loss in the event of disasters and the secondary objective is to minimize network resource usage.

*Input Parameters*:

- $G(V,\ E)$: Network topology where $V$ is the set of nodes and E is the set of directed links where all links are unidirectional.
- $P = \{p | p =< s_p,\ d_p,\ L_p,\ E_p >\}$: Set of k-shortest paths where $s_p$, $d_p$, and $L_p$ are source, destination, and length of the path in km, respectively. $E_p$ is the set of links on path p.
- $D = \{d | d =< v_d,\ C_d >\}$: Set of datacenters where $v_d$ is the node where datacenter $d$ is located and $C_d$ is the set contents datacenter *d hosts*.
- $T = \{t | t =< s_t,\ c_t,\ B_t,\ \mu_t,\ n_t,\ P_t >\}$: Set of connection requests where $s_t$ is the source, $B_t$ is the bandwidth request, $\mu_t$ is the partial-protection ratio (e.g., 0.5 for 50 % protection and 1.0 for full protection), $n_t$ is the minimum degraded-service ratio (for 30 % minimum protection, this value is 0.30), $c_t$ is the requested content of connection $t$, and $P_t \subset P$ is the set of possible paths to use for connection t where $s_t = s_p$.
- $Y = \{y | y =< E_y,\ \rho_y >\}$: Set of DZs where $E_y$ is the set of links that are members of Disaster $y$ and $\rho_y$ is the probability that Disaster y causes a failure.
- $K$: Maximum number of paths allowed for a connection.
- $U_p^y \in \{0,\ 1\}$: Equal to 1 if path $p$ goes through SRG $y$.
- $A_p^e \in \{0,\ 1\}$: Equal to 1 if link e $\in E_p$.
- $H_{c_t}^{d_p} \in \{0,\ 1\}$: Equal to 1 if datacenter $d$ has content $c$.
- $\pi_e$: Capacity of link $e$ which is the product of the total number of wavelengths on link $e$ and single wavelength capacity (can be different for each link).
- $\epsilon$ is an arbitrarily small number.
- *DDC*: Differential delay constraint which limits the usage of distant datacenters.

*Integer Variables*:

- $\alpha_p^t$: Bandwidth used on path $p$ for connection $t$.
- $X_t$: Number of paths for connection $t$.
- $Z_p^t \in \{0,\ 1\}$: 1 if path p is used for connection t.

*Objective*:

*Minimize* $\sum_{y\in Y}\left(\sum_{t\in T}\left(\sum_{p\in P_t}\alpha_p^t.U_p^y\right)\right).\rho_y + \epsilon.\sum_{t\in T}\sum_{p\in P_t}\alpha_p^t + Z_p^t.L_p$

The first term minimizes the expected bandwidth loss (risk) in case of disasters. The second term ensures that shortest paths are selected and minimum possible bandwidth is provisioned.

*Constraints*:

$$\sum_{p\in P_t}\alpha_p^t \geq B_t \quad \forall t \in T \tag{1}$$

$$\sum_{p\in P_t}\sum_{t\in T}\alpha_p^t.A_p^e \leq \pi_e \quad \forall e \in E \tag{2}$$

$$\sum_{p\in P_t} Z_p^t = X_t \quad \forall t \in T \tag{3}$$

$$X_t \leq K,\ X_t \geq 2 \tag{4}$$

$$Z_p^t \leq \alpha_p^t \forall t \in T \quad \forall t \in T, \forall p \in P_t \tag{5}$$

$$Z_p^t \geq \alpha_p^t / M \tag{6}$$

$$Z_p^t \leq \frac{H_{c_t}^{d_p} + 1}{2} \quad \forall t \in T\ \forall p \in P_t \tag{7}$$

$$\sum_{p\in P_d} Z_p^t \leq 1 \quad \forall t \in T\ \forall d \in D \tag{8}$$

$$\sum_{p\in P_t} Z_p^t.U_p^y \leq 1 \quad \forall t \in T, \forall y \in Y \tag{9}$$

$$\sum_{p\in P_t} Z_p^t.A_p^e \leq 1 \quad \forall t \in T, \forall e \in E \tag{10}$$

$$|(L_p - L_q).Z_p^t.Z_q^t| \leq DDC \quad \forall p, q \in P_t (p \neq q), \forall t \in T \tag{11}$$

$$\sum_{p\in P_t}\alpha_p^t - \sum_{q\in P_t}\alpha_q^t.U_q^y \geq \mu_t.B_t \quad \forall t \in T, \forall q \in P_t, \forall y \in Y \tag{12}$$

$$\alpha_p^t \geq Z_p^t.nt.Bt \quad \forall t \in T, \forall p \in P_t \tag{13}$$

Equation (1) enforces that the total reserved bandwidth is at least equal to the requested bandwidth. Equation (2) ensures that the total bandwidth usage in a link does not exceed the link capacity. Equations (3) and (4) constrain the number of multipaths per connection. Equations (5) and (6) define $Z_p^t$ where M is a large number. Equation (7) ensures that only the paths destined to datacenters with the requested content are

selected for a connection. Equation (8) ensures that at most one path is provisioned to a specific datacenter for a connection. Equations (9) and (10) ensure that every path used for a connection is SRG disjoint and link disjoint, respectively. As the number of multipaths increases, the survivability against post-disaster failures also increases. Possible multipath number is dependent on the availability of the network resources, content replication amount as well as the DDC. Equation (11) ensures that the differential path distances of a connection are selected to fulfill DDC (whose practical value is 8 ms [18]). Equation (12) ensures that, in case of single-link or disaster failures, requested level of degraded service is provided. When the requested degraded-service ratio is 1, full protection with overprovisioning is provided. Equation (13) gives the minimum level of degraded service required in case only one path survives (e.g., if a connection's multipath number is 4, it will take minimum degraded service after any three simultaneous failures).

## 5 Heuristic

Solving the ILP in Sect. 4 is not an easy task for large-scale problem instances. Although the optimization techniques such as column generation [19], Langrange relaxation [20], etc. can be applied, they are still ILP-based or exhaustive search algorithms and thus not fully scalable. To make the design more scalable for large problem instances, in this section, we propose a heuristic. We use this heuristic for static traffic, where all requests are known beforehand, though it can also be applied for dynamic traffic, where requests arrive and are processed one-by-one. We provide the pseudo code of the proposed heuristic in Fig. 4 to provision a set of connection requests $\boldsymbol{T}$.

The heuristic starts with initialization steps (lines 1–16). First, we construct a bipartite graph (see Fig. 5) in which the source nodes are on the left set and datacenters on the right. Second, we pre-calculate the k-shortest paths from each possible source to each datacenter in order to make future calculations more efficiently. All procedures starting with "*sort*", are regular sort operations (e.g., `sortPathsbyRisk` sorts connections based on their risk level, `sortCostofSolutionSet` sorts provisioning solutions for the connections request based on the cost of the paths with assigned capacity.) Once the initialization is complete, we switch to the running state where, for each request $t$, we invoke the provision procedure. The procedure starts with determining the candidate datacenters which can serve the requested content (lines 18–23) and form the set $D$'. The next step is to determine possible datacenter sets to provision a connection. Since we limit the system with maximum of four paths per connection, all possible two-to-four elements combination of $D$' is determined. For this purpose, we first

Input: $G(V, E)$, $T = \{ t \mid t = <s_t, b_t, \mu_t, c_t> \}$

```
1   procedure initialize( void ):
2   //Build a bipartite graph
3   set N, D to empty set
4   foreach (v in V)
5         if (v is a datacenter)
6               D.add(v)
7         else N.add(v) endif
8   endforeach
9   foreach (n in N)
10        foreach (d in D)
11        K_nd = kshortest(n, d, k)
12        sortPathsbyRisk(K_nd, +C^k_n,d ∀k)
13        Z_nd = select the least risky m
14        paths between nodes {n,d}
15        endforeach
16  endforeach
17  foreach (t in T)
18        provision(t) endforeach
19
20  procedure provision(t):
21  Set D' to an empty set.
22  foreach (d in D)
23        if (C_d contains c_t)
24        D'.add (d) endif
25  endforeach
26  set solution set S to empty set.
27  set numMaxCon = min(size(D'), 4)
28  for i = 2 to numMaxCon
29      set D'' to Combination(D',i)
30  //D'' is vector of datacenter set.
31      set sol to findSolution(t,D'')
32      S.add (sol)
33  endfor
34  sortCostofSolutionSet(S)
35  return S[0]
36
37  procedure findSolution (t, D'')
38  set S' to empty set. //sol. set
39  foreach (D_s in D'')
40      foreach possible combinations of
41        m least risky paths
42           Let P_t be the selected
43              paths to datacenters in
44              D_s among all Z_nd where n
45              is s_t and d is in D_s.
46           if(assignCapacity(t,P_t)*)
47             set Solution S to P_t with
48             capacities assigned.
49           set costOfS by Eq.(14)
50           S'.add(S)
51      end_foreach
52  end_foreach
53  sortCostofSolutionSet(S')
54  return S'[0]
```

**Fig. 4** Pseudo code for proposed heuristic. $^{+}C^{k}_{(n,d)}$ is the risk of the kth path between n and d shown in Eq. (15). *We use a modified version of capacity assignment algorithm in[14]. It assigns bandwidths to the paths by satisfying degraded service constraint

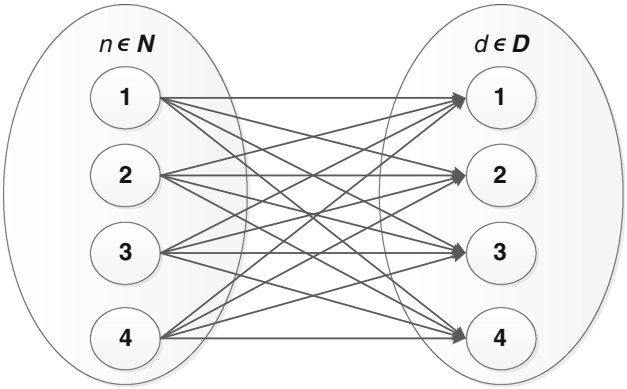



**Fig. 5** Bipartite graph: each edge $\langle n, d \rangle$ represent k-shortest paths between nodes $\{n, d\}$ and each path $P^k_{n,d}$ is weighted by its risk of failure found in Eq. (15)

construct the set D" which is the vector of all possible set of candidate datacenters using the combination operation. For instance, if i (number of connections) is set to two and if the content is available in three servers (A, B, and C), then, D" is the set of ({A, B}, {A, C}, {B, C}, {A, B, C}). The basic idea is to forward set D" to the findSolution procedure along with the incoming connection request, record the results (lines 28–33), and then sort all results based on the cost function to choose the one with the minimum cost (lines 28–34). (For sake of simplicity, we omit the case where no possible result exists.) In findSolution procedure (lines 37–54), we use a similar iterative method as above and break D" into subproblems, calculate the cost for each possible solution, and then choose the one with the lowest cost again. At this point, the problem boils down to finding the best (lowest cost) solution for a given set of candidate datacenters (e.g., A, B) which correspond to determining the paths from the source to each datacenter, assigning capacity to each path and calculating the cost. We iterate over all possible path combinations for a datacenter set to which we are trying to provision paths for a connection.

A set of possible paths between a source to a datacenter set is called a solution. We assign capacities to every solution using capacity assignment in [14] while satisfying the connection request's bandwidth and degraded-service requirement.

$$C_s = \sum_{y \in Y} \left( \sum_{p \in P_s} \alpha^s_p \cdot U^y_p \right) \cdot \rho_y + \epsilon \cdot \sum_{p \in P_s} \alpha^s_p \tag{14}$$

$$C^k_{n,d} = \sum_{y \in Y} U^y_p \cdot \rho_y \tag{15}$$

At the last step, we calculate the cost of the solution using the Eq. (14) which is the expected bandwidth loss in case of all predefined disasters and the total bandwidth provisioned for the connection. Even if the number of loops seems to be high, given that k may be bounded to a low value and a requested content is available in a limited number of datacenters, the computations are done efficiently, and in much faster way compared to the ILP.

## 6 Illustrative numerical examples

We present illustrative results by solving the ILP formulations and heuristics on the network in Fig. 6. Since the ILP approach is not scalable, we can optimally solve the ILP only in small problem instances. This provides a benchmark to gauge our heuristic performance. For large scenarios, we focus on checking the feasibility of the heuristic solution.

We compare three disaster-aware provisioning schemes: multipath to single destination (MSD), our proposed mul-

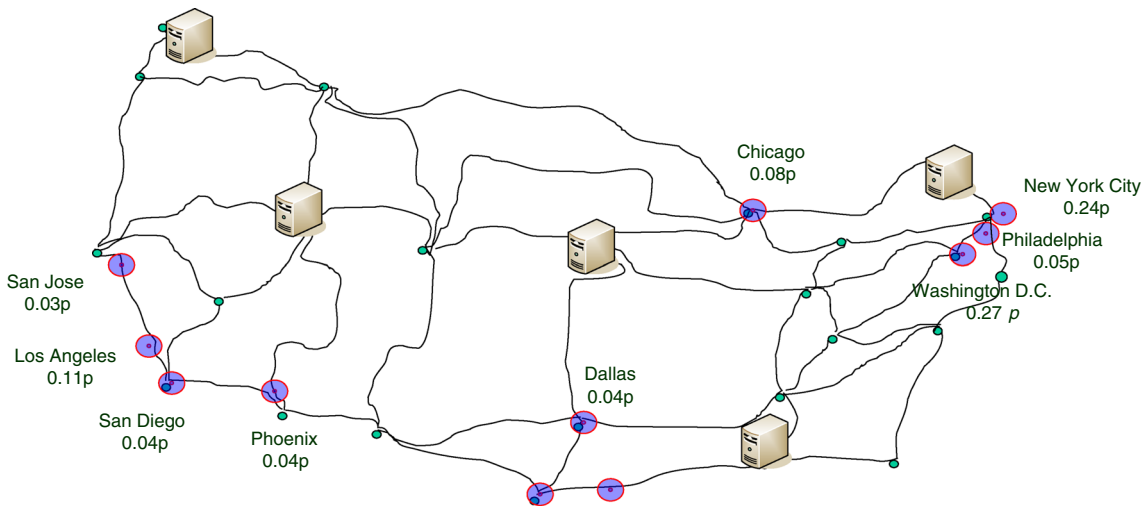



**Fig. 6** US-wide network with WMD disaster zones shown in circles with attack probabilities, where $p$ is the probability of a WMD attack targeting US

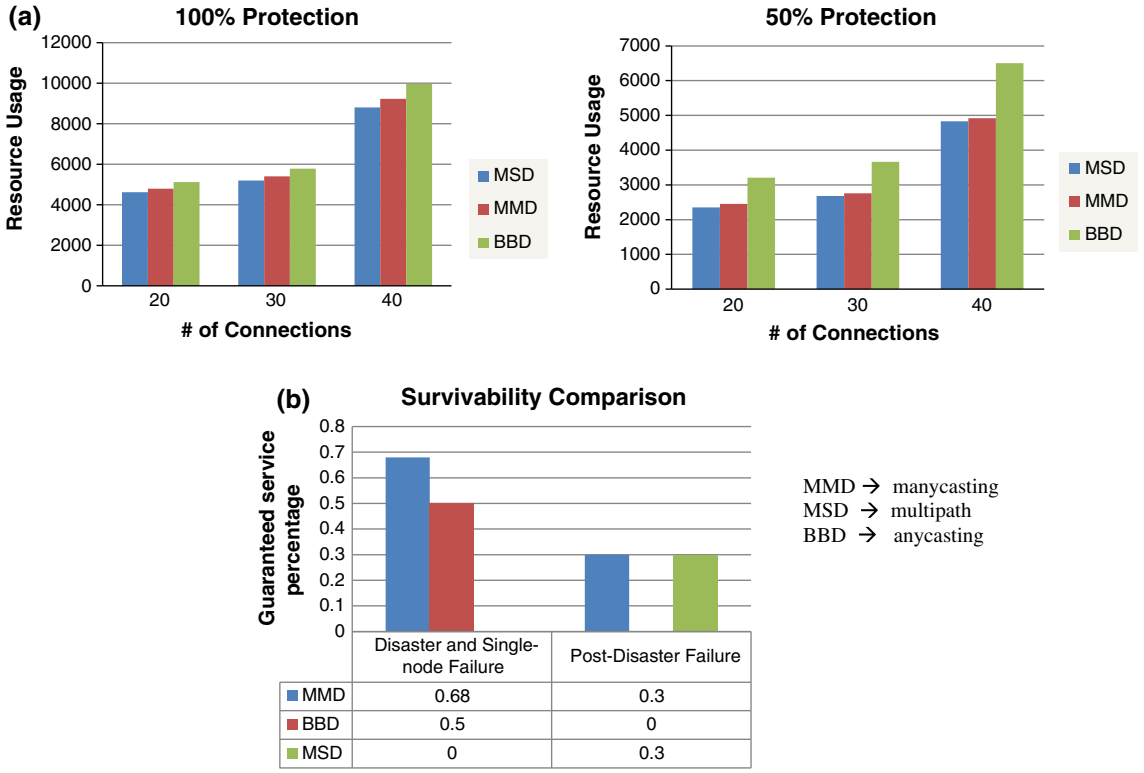



**Fig. 7** Comparison of the proposed scheme with benchmarks using ILP: **a** resource usage comparison of different risk-aware provisioning schemes with same protection (100 and 50 %) against disaster failures, **b** guaranteed service amount (worst case scenario)

tipath to multiple destination (MMD), and backup path to backup destination (BBD) (i.e., SRG-disjoint primary and backup paths to SRG-disjoint datacenters), in terms of resource usage and survivability. All three schemes perform routing in a risk-aware manner (i.e., they select their paths to reduce the expected loss in case of a disaster). In all schemes, all paths are link disjoint and SRG disjoint. The ILP formulations for MSD and BBD are derived from our MMD model with minor changes.

We study a 24-node US-wide network (Fig. 6) with 32 channels and 10 Gbps per channel per link. As disaster scenarios, we consider WMD attacks and 10 most-populated cities and Washington DC as possible targets, as in [10]. WMD zones are shown in circles. We assume that, if a disaster occurs, all nodes/links in the affected zone are down. We consider 5 datacenters as shown in Fig. 6 placed at safe nodes (not in a WMD zone), which minimizes the total distance of each node to its nearest datacenter. For datacenter placement, we use the greedy algorithm for the Center Selection problem in [21]. In this study, as ILP is computationally extensive, we consider a small example with 15 contents and 3 geographically distributed replicas per content. The bandwidth distribution of the connection requests is OC-3 (∼ 150 M), OC-12 (∼ 600 M), OC-48 (∼ 2.5 G), and OC-192 (∼ 10 G) which follows the ratios 40:30:20:10 (which is a realistic bandwidth distribution in a practical network [14]).

Content replica number limits the maximum number of multipath to 3. For nodes with degree <3, it is infeasible to find 3 link-disjoint paths; so, fewer paths will be provisioned for them. To compare our schemes, we only considered connections that can have 3 or more SRG-disjoint paths.

Figure 7a compares the total resource usage of our scheme with MSD and BBD schemes where all schemes provide same level of protection. Our scheme (MMD) consumes approximately 25 % less resources than BBD in case of 50 % protection and 10 % less resources in case of full protection while providing same amount of survivability against single-link and single-node (with a WMD attack) failures. However, while BBD may become vulnerable to even single-link failures after a disaster, our scheme offers survivability against post-disaster failures as well. Resource consumption of MMD and MSD are very close. In this particular setting, only around 5 % more resource is consumed by MMD. In return, it provides protection against node failures, while in MSD, connectivity may be lost. The resource usage directly depends on the number of datacenters and their geo-distribution as our approach tries to provide service from as many datacenters as possible satisfying DDC. In Fig. 7b, we compare the schemes according to their guaranteed minimum service amount in different failure scenarios (disaster, post-disaster, and node failure) when the resource usage is almost

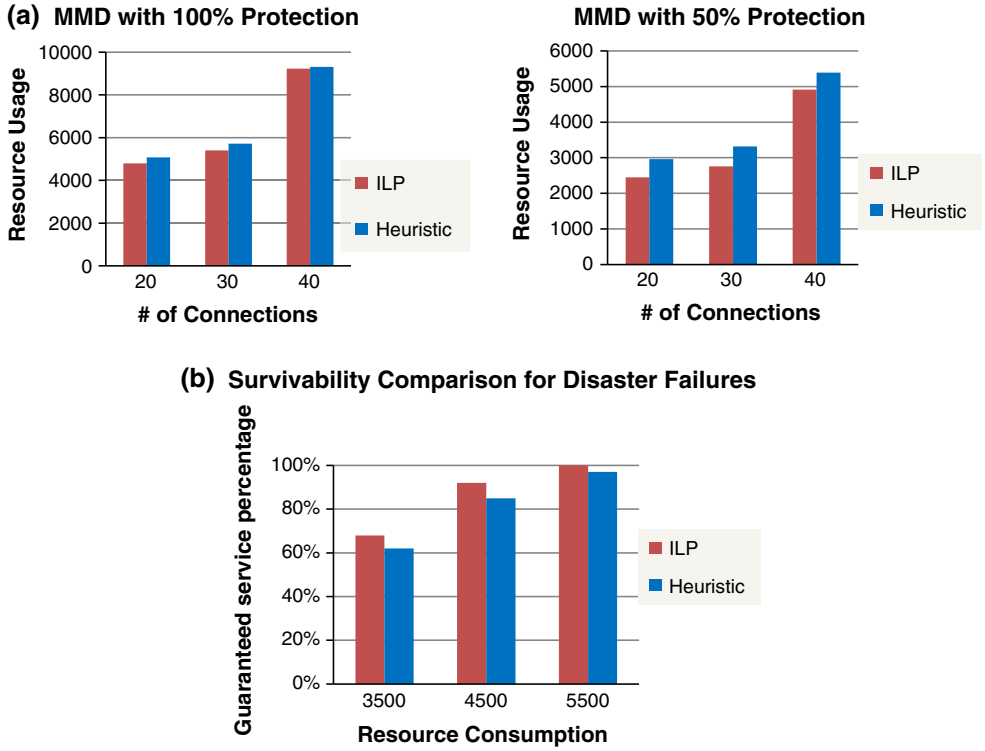



**Fig. 8** ILP versus heuristic comparison: **a** resource usage comparison of ILP and heuristic approaches for MMD scheme, **b** guaranteed service amount comparison between ILP and heuristic when they consume same amount of bandwidth

same (but the *guaranteed* service levels they offer differ). Tests are done with 20, 30, and 40 connections; since results are similar, we provide only the results with 30 connections. Resource usage is defined as the total bandwidth consumed on all links in the network. For example, if a connection path with 3 links has 5 B capacity, then it uses 15B in total. We set all schemes' resource usage to around three values in terms of optical carrier (OC): 3,500, 4,500, and 5,500, and resource usages of all schemes are within 2 % confidence interval. We observe that our scheme (MMD) provides 10–20 % more protection than BBD in all cases. MSD and MMD results are similar except in node failures. If datacenters are not located in risk-free nodes, then MSD cannot guarantee any degraded service in case of node failures, because all connection bandwidth can fail if the datacenter it connects is down. To sum up, our scheme offers comprehensive protection for different failure scenarios.

To demonstrate the effectiveness of our heuristic, we compare the heuristic and ILP results. We compare ILP's resource usage with that of the heuristic on the 24-node network for 20, 30, and 40 connections, 5 datacenters shown in Figs. 6, and 3 replicas per content. Figure 8a shows that our heuristic's resource usage has a similar trend with ILP. The small gap-to-optimality of around 20 % confirms the superior performance of the heuristic. Figure 8b shows that, when the same amount of resource is provided for both ILP and heuristic, their survivability characteristics are comparable.

### 6.1 Effect of the number of replicas per content on the performance of manycasting

Datacenter placement is a fundamental issue in supporting cloud services in optical networks. It concerns not only the cost of providing cloud services, but also the service availability against failures via proper service replicas. Under the manycast model, datacenter placement is important in balancing between network cost, service latency, availability, as well as providing survivability in our case. On the one hand, cloud service providers wish to direct user demands to nearby datacenters with the smallest latencies and the minimum transmission costs by replicating the content to many datacenters so that they can all be served. To achieve this, the number of datacenters is also an important metric, and a promising approach can be achieved by densely distributing the datacenters in the area of interest. On the other hand, deploying a datacenter is costly, so this requires the number of datacenters to be minimized. The trade-off between network cost, service latency, availability, and survivability makes 'how many datacenters', 'where to place them', and 'how many replicas per content' questions a fundamental optimization problem.

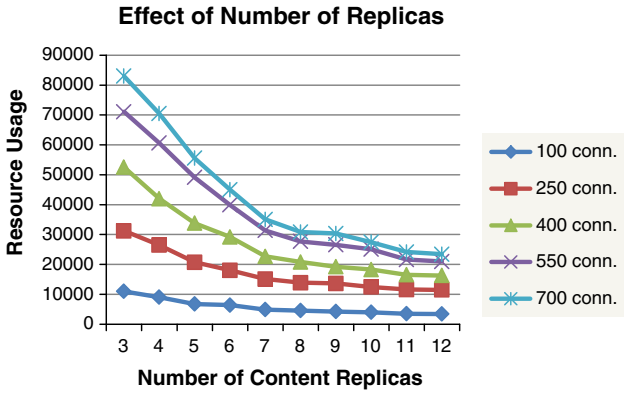



**Fig. 9** Effect of content number on resource usage

We analyzed our proposed scheme's sensitivity against number of replicas per content, and replica distribution. We noticed that the number of datacenters has limited effect on the performance of manycasting when the number of replicas is fixed. Figure 9 shows the effect of number of replicas on resource usage in the 24-node network. To experiment with more replicas, we use 12 datacenters where we repeat the simulations for 30 different combinations of datacenter locations. We also run the simulations for different replica distributions on these 12 datacenters and take the average. The trends of almost all replica distributions are very similar, up to a point; as the number of replicas increase, the resource usage dramatically decreases and then the decrease saturates. More replicas do not always provide more flexibility to choose a shorter path; rather, more replicas mean more usage of storage resources and more usage of bandwidth to perform replication and synchronization (i.e., consistency among replicas of a content). For our example, after the 7th replica per content, there is no significant gain in terms of resource usage. To conclude, resource usage reduces significantly as the number of datacenters increases (to have a higher number of replicas, we also need more datacenters), but after a certain value, increasing the number of datacenters does not help much to reduce resource utilization.

## 7 Conclusion

Recent disasters have shown that current survivability schemes in our networks are lacking and enhanced schemes are needed. The advent of cloud services delivered by datacenter networks gives novel opportunities to provide protection against disasters in a cost-effective way. By exploiting manycasting, which is enabled by the destination-agnostic cloud services, we can improve the network's adaptability against disasters. We proposed a disaster-aware manycast provisioning scheme that multiplexes service over multiple relatively less risky paths destined to multiple servers/datacenters with manycasting. In illustrative examples, our proposed methods were applied to a US-wide network topology, and we showed its advantageous properties compared to benchmark disaster-aware provisioning techniques.

## References


[1] Sultan, N.: Knowledge management in the age of cloud computing and Web 2.0: experiencing the power of disruptive innovations. Int. J. Inf. Manag. **33**(1), 160–165 (2013)

[2] Develder, C., De Leenheer, M., Dhoedt, B., Pickavet, M., Colle, D., De Turck, F., Demeester, P.: Optical networks for grid and cloud computing applications. Proc. IEEE **100**(5), 1149–1167 (2012)

[3] Habib, M.F., Tornatore, M., Dikbiyik, F., Mukherjee, B.: Disaster survivability in optical communication networks. Comput. Commun. **36**(6), 630–644 (2013)

[4] Charbonneau, N., Vokkarane, V.: Routing and wavelength assignment of static manycast demands over all-optical wavelength-routed WDM networks. IEEE/OSA J. Opt. Commun. Netw. **2**(7), 442–455 (2010)

[5] Das, G., et al.: SRLG identification from time series analysis of link state data. In: Proceedings of COMSNETS, Bangalore, India, January 2011

[6] Lu, S., Yang, X., Ramamurthy, B.: Shared risk link group (SRLG)-diverse path provisioning under hybrid service level agreements in wavelength-routed optical mesh networks. IEEE/ACM Trans. Netw. **13**(4), 918–931 (2005)

[7] Kuperman, G., Modiano, E., Narula-Tam, A.: Analysis and algorithms for partial protection in mesh networks. In: IEEE INFOCOM, April 2011

[8] Chang, H.: A multipath routing algorithm for degraded-bandwidth services under availability constraint in WDM networks. In: Advanced Information Networking and Applications Workshops (WAINA) (2012)

[9] Habib, M.F., Tornatore, M., Leenheer, M.D., Dikbiyik, F., Mukherjee, B.: Design of disaster-resilient optical datacenter networks. IEEE/OSA J. Lightw. Technol. **30**(16), 2563–2573 (2012)

[10] Dikbiyik, F., Tornatore, M., Mukherjee, B.: Minimizing the Risk From Disaster Failures in Optical Backbone Networks. J. Lightw. Technol. **32**(18), 3175–3183 (2014)

[11] Neumayer, S., Zussman, G., Cohen, R., Modiano, E.: Assessing the vulnerability of the fiber infrastructure to disasters. In: IEEE INFOCOM, April 2009

[12] Weems, T.L.: "How far is far enough", Disaster Recovery Journal, vol. 16, no. 2, Spring 2003

[13] Zhang, W., Tang, J., Wang, C., Soysa, S.: Reliable adaptive multipath provisioning with bandwidth and differential delay constraints. In: IEEE INFOCOM, March 2010

[14] Huang, S., Xia, M., Martel, C., Mukherjee, B.: A multistate multipath provisioning scheme for differentiated failures in telecom mesh networks. IEEE/OSA J. Lightw. Technol. **28**(11), 1585–1596 (2010)

[15] Xiao, J., Wen, H., Wu, B., Jiang, X., Ho, P., Zhang, L.: Joint design on DCN placement and survivable cloud service provision over all-optical mesh networks. IEEE Trans. Commun. **62**(1), 235–245 (2014)

[16] Carter, C., Yi, S., Ratanchandani, P., Kravets, R.: Manycast: exploring the space between anycast and multicast in ad hoc networks. In: IEEE Mobile Comput. Netw. (MobiCom), April 2003

[17] Agarwal, P., et al.: The resilience of WDM networks to probabilistic geographical failures. In: Proceedings of IEEE INFOCOM, Shanghai, China, April 2011
[18] Huang, S., Martel, C., Mukherjee, B.: Survivable multipath provisioning with differential delay constraint in telecom mesh networks. IEEE/ACM Trans. Netw. **19**(6), 657–669 (2011)
[19] Barnhart, C., Johnson, E.L., Nemhauser, G.L., Savelsbergh, M.W., Vance, P.H.: Branch-and-price: column generation for solving huge integer programs. Oper. Res. **46**(3), 316–329 (1998)
[20] Fisher, L.: The Lagrangian relaxation method for solving integer programming problems. Manag. Sci. **27**(1), 1–18 (1981)
[21] Kleinberg, J., Tardos, E.: Algorithm design. Addison Wesley, Reading (2006)