

On Hybrid IR and AR Service Provisioning in Elastic Optical Networks

Wei Lu, *Student Member, IEEE*, Zuqing Zhu, *Senior Member, IEEE*, and Biswanath Mukherjee, *Fellow, IEEE*

Abstract—This paper investigates hybrid immediate reservation (IR) and advance reservation (AR) service provisioning in elastic optical networks, with the objective to minimize IR/AR service conflicts. We design algorithms to coordinate service provisioning of IR and AR requests. Specifically, both proactive and reactive IR provisioning schemes are considered to minimize IR service failures. Our AR scheduling algorithm can coordinate AR service provisioning with various IR traffic patterns, balance spectrum utilization in both time and spectral domains, and reduce IR/AR service conflicts. Simulation results verify that our proposed IR + AR schemes can significantly reduce IR service failures as well as routing and spectrum allocation reconfigurations in IR service provisioning. Moreover, the results indicate that the proposed IR + AR schemes can achieve more performance gain from AR flexibility when compared with two existing benchmarks.

Index Terms—Advance reservation, elastic optical network (EON), immediate reservation, service conflict.

I. INTRODUCTION

TODAY, bandwidth-hungry applications have driven Internet traffic to grow exponentially. To accommodate such huge traffic, network operators are planning their networks with 100-Gb/s line speed and beyond. Due to enhanced optical-layer agility brought by the flexible grid, elastic optical networks (EONs) have attracted a lot of research interests and are being considered as a promising physical infrastructure for next-generation backbone networks and future datacenter (DC) networks [1]–[3]. Compared with fixed-grid wavelength division multiplexing (WDM) networks, EONs can achieve higher bandwidth efficiency and more adaptive service provisioning by assigning just-enough numbers of frequency slots (FS^s) to serve requests [4], [5]. Hence, we expect future EONs to accommodate a wide range of applications, which have diverse traffic

characteristics and quality-of-service (QoS) requirements, with flexible network control and management.

It is known that QoS-enabled optical networks can leverage two types of bandwidth reservation schemes, i.e., immediate reservation (IR) and advance reservation (AR), to support heterogeneous applications [6], [7]. Specifically, IR takes care of real-time applications, such as video streaming, and provisions bandwidth immediately upon receiving a request. As these applications usually have unspecified durations, they cannot be planned in advance. On the other hand, by reserving bandwidth in advance for future data transmission, AR can guarantee the availability of network resources for delay-tolerant applications, such as scheduled DC backup. The service provisioning scheme of AR requests is more flexible than that of IR ones. Therefore, if there is no admission control, AR requests may occupy most of network resources, and lead to high blocking of IR requests. Moreover, AR requests can generate two-dimensional (2-D) spectrum fragments (i.e., spectrum fragments existing in both time and spectral domains in a correlated manner [8]) in an EON, and result in reduced spectrum utilization. It is known that request blocking and spectrum utilization are directly related to network operators' revenue. Hence, researchers have tried to address hybrid IR/AR service provisioning in various networks [9]–[16].

Previously, researchers have proposed EON planning and provisioning algorithms for IR requests [17]–[19] and AR requests [8], [20]. Regarding the spectrum-sharing framework for hybrid IR/AR, full sharing has been demonstrated to achieve highest spectrum utilization among full-sharing, strict-partitioning, and flexible-partitioning frameworks in fixed-grid WDM networks [12]. This is appreciated by network operators, especially for improving spectrum utilization in EONs [21],[22]. However, the full-sharing framework also introduces a new problem: IR request preemption, i.e., an IR request may be preempted when it conflicts with a reserved AR request. High request preemption rate can degrade customers' service satisfaction and may reduce network operators' revenue. Therefore, together with the advantage of the full-sharing framework, it is of interest to address how to minimize service conflicts between IR and AR requests while considering the unique spectrum allocation mechanisms in EONs.

Two more ideas can be incorporated with the problem to make it more useful. One is to allow more flexibility in AR, which can result in better spectrum utilization but AR can become more complex because both spectrum and time dimensions are involved in request scheduling. The other is to consider different IR traffic patterns (i.e., evenly and unevenly distributed over time) and design corresponding AR scheduling algorithms to

Manuscript received June 27, 2015; revised August 21, 2015; accepted September 15, 2015. Date of publication September 16, 2015; date of current version October 5, 2015. This work was supported in part by the NSFC Project 61371117, the Fundamental Research Funds for the Central Universities under Grant WK2100060010, the Natural Science Research Project for Universities in Anhui under Grant KJ2014ZD38, the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA06011202, and the UC Davis Networks Lab.

W. Lu is with the School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China, and also with the University of California, Davis, CA 95616 USA (e-mail: luwei11@mail.ustc.edu.cn).

Z. Zhu is with the School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China (e-mail: zqzhu@ieec.org).

B. Mukherjee is with the Department of Computer Science, University of California, Davis, CA 95616 USA (e-mail: bmukherjee@ucdavis.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JLT.2015.2479366

adapt to them. In the latter case, AR flexibility can play an important role for balancing the IR/AR traffics in both time and spectral domains, especially when the peak and off-peak periods of IR traffic are known in advance. This is because, with AR flexibility, it is possible to schedule AR requests in off-peak periods of IR traffic and thus avoid IR/AR spectrum competition in peak periods.

In this paper, by combining the above two ideas, we investigate how to improve service provisioning of hybrid IR and AR requests in an EON, where spectrum resources are fully shared among IR and AR requests. Specifically, we consider a dynamic EON, in which the IR requests' service durations are unknown while the AR ones use the specific start time and specific duration (STSD) reservation model [6]. For the STSD model, we consider both the STSD-fixed and STSD-flexible cases. The STSD-fixed case means that the service start time of an AR request is fixed in the future, while STSD-flexible reservation allows the service start time of an AR request to slide within a pre-defined time window. Regarding the relation between IR and AR requests, we assume that the AR ones possess higher priority, which means that, whenever a service conflict occurs, an in-service IR request has to compromise to the AR request which shares resources with it. Based on this model, we design both proactive and reactive IR provisioning algorithms to minimize IR service failures (i.e., including both service blocking and interruption), and also propose an AR scheduling algorithm that considers different IR traffic patterns to coordinate AR provisioning, to balance resource utilization in both time and spectral domains, and to eventually relieve IR/AR service conflicts in the EON.

The rest of the paper is organized as follows. Section II reviews related work. In Section III, we describe the network model and define the problem of hybrid IR/AR service provisioning in an EON. To minimize IR/AR service conflicts, novel provisioning algorithms are proposed in Section IV. Section V discusses the simulations for performance evaluation. Finally, we summarize the paper in Section VI.

II. RELATED WORK

Based on various networks, previous studies have considered the admission control, spectrum partition, routing and scheduling for IR or AR requests, and tried to limit the service preemption rate of IR ones [9]–[16]. Wischik and Greenberg [9] presented an admission control algorithm for AR requests based on a statistical model that can estimate approximate service preemption rate of IR requests. An admission control scheme was developed for IR requests in [10]. Taking traffic uncertainty into account, Ahmad *et al.* [11] proposed a dynamic look-ahead time based admission control for AR requests. Spectrum full-sharing framework was demonstrated in [12] for fixed-grid WDM networks to achieve much lower blocking probability than the strict- and flexible-partitioning ones.

In [13], preemption-aware routing algorithms for IR requests were designed based on the link cost function that can compute the preemption probability of IR requests. Similarly, based on rigorous statistical analysis of the reservation dynamics, Wu and Dharam [14] proposed a scheduling algorithm for AR

requests such that the number of preempted IR requests could be minimized. Dharam *et al.* [15] investigated the integrated AR scheduling that involved both the routing and preemption phases. Specifically, in the preemption phase, they used a runtime scheme to minimize the number of preempted IR requests. However, none of these studies addressed the resource allocation in the optical layer. Although several routing and wavelength assignment (RWA) algorithms were proposed for fixed-grid WDM networks in [16], they cannot solve the problem of IR/AR service conflicts in EONs as the resource allocation mechanisms in the optical layer are different.

On the other hand, most studies mentioned above assumed that all the AR requests were STSD-fixed ones, which might not reflect the practical situation. Previously, Charbonneau and Vokkarane [23] conducted a comprehensive survey on AR provisioning in WDM networks, in which they discussed the coexistence of STSD-fixed and STSD-flexible requests. Researchers have also considered how to arrange the STSD-flexible AR requests to achieve high service efficiency [24]–[28]. In [24], the problem of scheduling STSD-flexible AR requests was first defined for WDM networks, and the authors solved it exactly with a branch-and-bound method and also proposed a heuristic based on tabu search. In [25] and [26], two separated time scheduling and RWA approaches were proposed to minimize the wavelength utilization of AR requests. Andrei *et al.* [27] designed an integrated service provisioning algorithm based on Lagrangian relaxation, which could optimize the time scheduling and RWA of an AR request jointly. Su *et al.* [28] studied the effect of AR requests' flexibility on resource utilization. Nevertheless, these studies tried to balance AR traffic in time and spectral domains without considering the traffic pattern of IR requests. As we will show later in this work, it would be beneficial to coordinate AR provisioning according to IR traffic pattern so that IR/AR service conflicts can be reduced further.

III. PROBLEM DESCRIPTION

A. Network Model

We model the EON physical topology as a directed graph $G(V, E)$, where V denotes the node set and E is for the set of fiber links, each of which can accommodate B FS'. In the context of this work, we assume that there is no spectrum converter in the EON. Hence, spectrum-continuity constraint needs to be satisfied when setting up a lightpath [17]. Meanwhile, the EON operates in a discrete-time manner, which means that the time is divided into time slots (TS') evenly and network operations happen in those TS'.

When choosing the practical value of TS' duration, two factors should be considered: the speed of lightpath configuration and dynamics of traffic demands in EONs. To ensure effective service provisioning, the time required for lightpath configuration should be at least guaranteed within TS's duration [29],[30]. On the other hand, to adapt to the changing traffic load, network operations should be adjusted according to the dynamics of traffic demands. For instance, some emerging applications, such as data backup and migration in inter-DC networks, may need bandwidth-on-demand on the scale of tens of minutes [29].

We define T as the maximum number of look-ahead TS' which the network operator can observe. To record FS usage on links over time, we define a matrix $[U]_{|E| \times T}$, whose element $u_{e,t}$ is a bitmask that contains B bits to represent the availability of all the FS' on link e in TS t . If the j th FS on link e is available in TS t , $u_{e,t}[j] = 1$, otherwise $u_{e,t}[j] = 0$.

B. Request Models

For the i th incoming IR request, we model it with a tuple $IR(s^i, d^i, n^i, \zeta^i)$, where s^i and d^i are the source and destination nodes, n^i is the bandwidth requirement in terms of FS', and ζ^i is its arrival TS. Here, since we apply the specified start time but unknown duration reservation model to IR requests, it would be challenging for the network operator to provide interruption-free service for them without knowing their service durations, because the IR/AR service conflict can cause competitions on FS'.

AR requests can use either STSD-fixed or STSD-flexible reservation. The service start time of an STSD-fixed request is fixed at a future TS, whereas it can slide within a pre-defined time window for an STSD-flexible request. Hence, we denote an AR request as a tuple $AR(s^a, d^a, n^a, \zeta^a, \delta^a, \tau^a)$, where s^a and d^a are the source and destination nodes, n^a is the bandwidth requirement in term of FS', ζ^a is the earliest service start time, δ^a is the service duration in term of TS', and τ^a is the latest service end time. STSD-fixed means $\tau^a = \zeta^a + \delta^a - 1$, while the STSD-flexible scheme has $\tau^a > \zeta^a + \delta^a - 1$, i.e., the service start time can slide within $[\zeta^a, \tau^a - \delta^a + 1]$. We define the flexibility of an AR request as:

$$\gamma^a = \frac{\tau^a - \zeta^a + 1}{\delta^a} - 1. \quad (1)$$

C. Hybrid IR/AR Service Provisioning

We consider a dynamic network scenario, in which IR and AR requests arrive, hold for a while, and then depart. To serve an $IR(s^i, d^i, n^i, \zeta^i)$, we need to find a route p_{s^i, d^i} and assign an FS block $[f_s^i, f_e^i]$ on it, where f_s^i and f_e^i are the indices of start FS and end FS, satisfying $f_e^i - f_s^i + 1 \geq n^i$. Note that, to avoid interrupting IR service, we need to ensure that available route and FS block, i.e., routing and spectrum assignment (RSA), can be offered for the entire provision period of IR .

On the other hand, when an $AR(s^a, d^a, n^a, \zeta^a, \delta^a, \tau^a)$ arrives, the network operator needs to 1) schedule it with a valid service time window $[t_s^a, t_e^a]$, where t_s^a and t_e^a are the service start TS and end TS, respectively, and 2) find a route p_{s^a, d^a} and reserve an FS block $[f_s^a, f_e^a]$ on it to satisfy its bandwidth requirement, i.e., $f_e^a - f_s^a + 1 \geq n^a$.

Since we consider that AR requests possess higher priority, they can seize the FS' that are occupied by IR requests when IR/AR service conflicts happen. Fig. 1 illustrates an example of IR/AR service conflict where FS block $[1, 3]$ is assigned to IR at t_1 , and later, FS block $[1, 5]$ on the same link is reserved for AR starting from t_2 . If the duration of IR is not over at t_2 , a service conflict will occur, and the IR service will be preempted, if there does not exist an effective lightpath reconfiguration

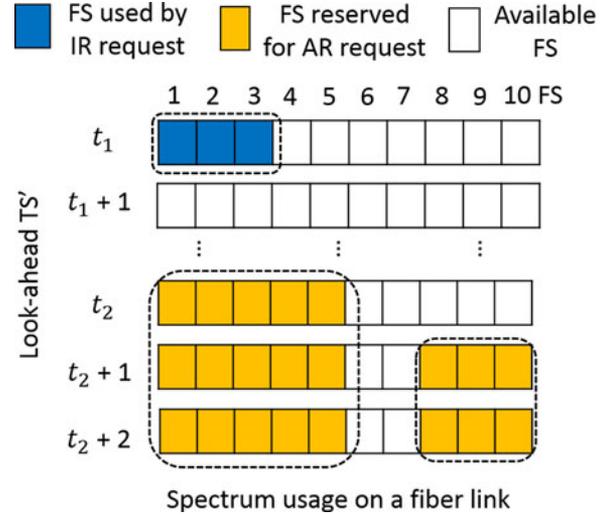


Fig. 1. Example of IR/AR service conflict in an EON.

method. In this work, we try to minimize IR/AR service conflicts by cooperatively managing service provisioning of IR and AR requests.

IV. HYBRID IR/AR SERVICE PROVISIONING ALGORITHMS

A. Proactive Algorithms

Proactive IR/AR service provisioning algorithms try to avoid IR/AR service conflicts, by considering several factors that have impacts on service blocking and interruption jointly.

1) *Proactive IR Provisioning*: To serve $IR(s^i, d^i, n^i, \zeta^i)$, we need to find an available RSA scheme, i.e., route p_{s^i, d^i} and FS block $[f_s^i, f_e^i]$ on it that satisfies $f_e^i - f_s^i + 1 \geq n^i$. Also, to ensure that the service of IR is uninterrupted, we need the RSA scheme(s) to be available for the entire provisioning period of IR . Considering the risk of IR/AR service conflict, we design new IR provisioning algorithms suitable for this case, even though quite a few RSA schemes have been proposed before for IR requests on the basis of different situations. To minimize IR service failures (i.e., including both service blocking and interruption), we consider the following factors in the algorithm design.

- *Spectrum Efficiency (SE)*: Number of FS' to be assigned for an IR request on the route is:

$$N_{SE}^i = n^i \cdot \text{hop}(p_{s^i, d^i}), \quad (2)$$

where $\text{hop}(p_{s^i, d^i})$ returns the hop count of p_{s^i, d^i} . We prefer to use the route with smaller N_{SE}^i to save more spectrum resources and thus reduce request blocking.

- *Spectrum Misalignment (SM)*: Total number of FS' that are misaligned between each link on the route and its adjacent links in the EON within the scope of assigned FS block is calculated as:

$$N_{SM}^i = \sum_{e \in E_{p_{s^i, d^i}}^n} \sum_{j=f_s^i}^{f_e^i} u_{e, t_c}[j], \quad (3)$$

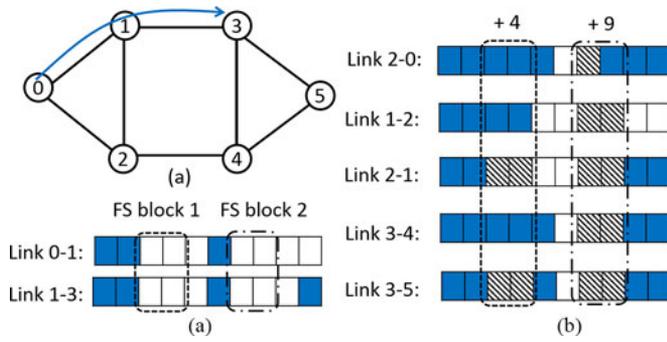


Fig. 2. Example of spectrum misalignment on adjacent links.

where $E_{p_{s^i, d^i}}^n$ is the set of adjacent links for p_{s^i, d^i} , and t_c denotes the current TS. The optimal RSA scheme should misalign the least available FS' on the adjacent links. Hence, we choose the RSA scheme with a smaller N_{SM}^i . Fig. 2 illustrates an example of spectrum misalignment on adjacent links. Suppose we have an IR request with bandwidth requirement of two FS' from *Node 0* to *Node 3* in the topology in Fig. 2(a), where only one route $0 \rightarrow 1 \rightarrow 3$ is considered for simplicity. Fig. 2(b) shows the spectrum usage on links, and there are two available FS blocks for the IR request, i.e., FS blocks [3, 4] and [7, 8]. Note that, when there is a large FS block that has more than one suitable spectrum assignment for the request, we choose the one with the smallest start-FS index to avoid generating spectrum fragments. Fig. 2(c) shows the adjacent links of *Path 0* \rightarrow *1* \rightarrow *3*. For example, *Links 2* \rightarrow *0* and *1* \rightarrow *2* are the adjacent links of *Link 0* \rightarrow *1*. Fig. 2(c) also plots the misaligned FS' on the links within the two FS blocks. We can see that the number of misaligned FS' is 4 within FS block [3, 4], while it increases to 9 within FS block [7, 8]. Therefore, we assign FS block [3, 4] on path $0 \rightarrow 1 \rightarrow 3$ to the request.

- *Congestion Avoidance (CA)*: Number of FS' that are on the route using highly-congested link(s) is calculated as:

$$N_{CA}^i = \beta \cdot n^i \cdot |E_{p_{s^i, d^i}}^c|, \quad (4)$$

where $E_{p_{s^i, d^i}}^c$ represents the set of highly-congested links on p_{s^i, d^i} , operation $|\cdot|$ returns the number of elements in a set, and β is the penalty coefficient. A highly-congested link is defined as the link on which the ratio of used FS' has exceeded a preset threshold (e.g., 60%). If the traffic is unevenly distributed, certain links can become highly congested and the bottleneck to induce service blocking. Using this metric, we try to relieve this situation.

- *Available Time (AT)*: Maximum number of TS' can be obtained as follows by checking the availability of the RSA scheme, i.e., route p_{s^i, d^i} and FS block $[f_s^i, f_e^i]$ on it, along the time axis:

$$N_{AT}^i = |\{t : t \in [t_e, T], \vartheta_\tau = 1, \tau \geq t\}|, \quad (5)$$

where ϑ_t is a binary variable defined as:

$$\vartheta_t = \begin{cases} 1, & \sum_{j=f_s^i}^{f_e^i} \left(\prod_{e \in p_{s^i, d^i}} u_{e,t}[j] \right) = n^i, \\ 0, & \sum_{j=f_s^i}^{f_e^i} \left(\prod_{e \in p_{s^i, d^i}} u_{e,t}[j] \right) < n^i, \end{cases} \quad (6)$$

and \prod executes bitwise dot-product, i.e., \otimes . To minimize service interruption, we use RSA with larger N_{AT}^i .

To consider these factors jointly, we put them into two metrics for evaluating the value of a RSA scheme, i.e., 1) spectrum weight, which is calculated as:

$$W_S^i = N_{SE}^i + N_{SM}^i + N_{CA}^i, \quad (7)$$

and 2) time weight, which is calculated as:

$$W_T^i = N_{AT}^i. \quad (8)$$

Keeping both weights in mind, we define a minimum guaranteed available time ϖ to make our evaluation mechanism work. If there exist RSA scheme(s) that satisfy $W_T^i \geq \varpi$, we only consider such RSA scheme(s) as our candidate(s) and finally select the one with the minimum W_S^i to save spectrum cost as well as have a guaranteed available time; otherwise, we select the RSA scheme that has the maximum W_T^i to minimize service interruption probability. Moreover, if there exist multiple optimal RSA schemes, we select the one with the smallest start-FS index.

Algorithm 1 shows the procedure of our proposed proactive IR provisioning scheme. When an IR request $IR(s^i, d^i, n^i, \zeta^i)$ arrives at t_c , *Line 1* loads the K -shortest path candidates that have been precalculated for each node-pair in the topology. *Lines 2–22* evaluate each path candidate $p_{s^i, d^i, k}$ to determine the optimum FS block $[f_{s,k}^i, f_{e,k}^i]$ using the designed metrics, and record the RSA solution and its W_S^i and W_T^i . If no such FS block exists, *Line 20* records the result. Finally, if feasible RSA solution(s) exist, *Lines 24–30* determine the optimal RSA using the designed evaluation mechanism, set p_{s^i, d^i} and $[f_s^i, f_e^i]$ accordingly, and record *IR* as provisioned. Otherwise, *Line 32* marks *IR* as blocked/interrupted. Note that spectrum assignment is performed from the smallest FS index to the highest FS index. Therefore, using the RSA scheme with the smallest start FS helps to consolidate FS usage in spectral domain, as shown in *Lines 14, 16, 26, and 28*.

2) *Proactive AR Scheduling*: We need to schedule an $AR(s^a, d^a, n^a, \zeta^a, \delta^a, \tau^a)$ with both a valid service time window and an available RSA solution, i.e., $[t_s^a, t_e^a]$, p_{s^a, d^a} , and $[f_s^a, f_e^a]$. Given AR's flexibility, it is possible to relieve IR service blocking and interruption by provisioning AR services in a coordinated manner. Thus, when determining the scheduling scheme for AR, we consider the following factors, part of which are similar to those that have been considered for IR.

- *Spectrum Efficiency (SE)*: Number of FS' to be reserved for an AR request on the route during the service time window is calculated as:

$$N_{SE}^a = n^a \cdot \text{hop}(p_{s^a, d^a}) \cdot \delta^a. \quad (9)$$

Algorithm 1: Proactive IR Provisioning

Input: $IR(s^i, d^i, n^i, \zeta^i), t_c, \{u_{e,t} : \forall e, t \in [t_c, T]\}, \varpi$;
Output: $p_{s^i, d^i}, [f_s^i, f_e^i]$;

- 1 load K -shortest path candidates $\{p_{s^i, d^i, k}\}$ between s^i and d^i ;
- 2 **for each** $p_{s^i, d^i, k}$ **do**
- 3 get FS usage on $p_{s^i, d^i, k}$ as u_{s^i, d^i, k, t_c} ;
- 4 find available FS blocks that are larger than n^i FS';
- 5 **if** such FS block(s) exist **then**
- 6 store the available FS block(s) in set $\{[f_{s,k,n}^i, f_{e,k,n}^i]\}$;
- 7 **for each** FS block $[f_{s,k,n}^i, f_{e,k,n}^i]$ **do**
- 8 tailor a just-enough FS block by reducing $f_{e,k,n}^i$ to $f_{s,k,n}^i + n^i - 1$;
- 9 calculate W_S^i using Eq. (7);
- 10 calculate W_T^i using Eq. (8);
- 11 **end**
- 12 find FS blocks that have $W_T^i \geq \varpi$;
- 13 **if** such FS block(s) exist **then**
- 14 set $[f_{s,k}^i, f_{e,k}^i]$ on $p_{s^i, d^i, k}$ as the one with minimum W_S^i and smallest start-FS index;
- 15 **else**
- 16 set $[f_{s,k}^i, f_{e,k}^i]$ on $p_{s^i, d^i, k}$ as the one with maximum W_T^i and smallest start-FS index;
- 17 **end**
- 18 record the RSA scheme and its W_S^i and W_T^i ;
- 19 **else**
- 20 record no RSA solution found on $p_{s^i, d^i, k}$;
- 21 **end**
- 22 **end**
- 23 **if** feasible RSA solution(s) exist **then**
- 24 find the ones that have $W_T^i \geq \varpi$;
- 25 **if** such RSA solution(s) exist **then**
- 26 set p_{s^i, d^i} and $[f_s^i, f_e^i]$ as the one with minimum W_S^i and smallest start-FS index;
- 27 **else**
- 28 set p_{s^i, d^i} and $[f_s^i, f_e^i]$ as the one with maximum W_T^i and smallest start-FS index;
- 29 **end**
- 30 record IR as provisioned;
- 31 **else**
- 31 mark IR as blocked/interrupted;
- 33 **end**

We choose a scheduling scheme with a smaller N_{SE}^a to leave more spectrum resources for future requests.

- **Spectrum Misalignment (SM):** Number of FS' that are misaligned between the links on the route and their adjacent links within the scope of assigned FS block during the service time window is calculated as:

$$N_{SM}^a = \sum_{e \in E_{p_{s^i, d^i, k}^a}} \sum_{t=t_s^a}^{t_e^a} \sum_{j=f_s^a}^{f_e^a} u_{e,t}[j]. \quad (10)$$

We choose an AR scheduling scheme with smaller N_{SM}^a .

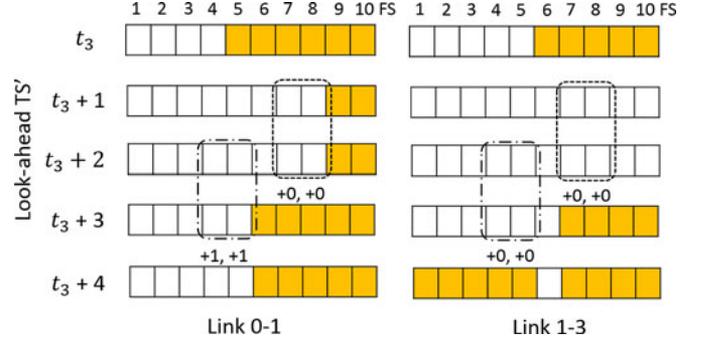


Fig. 3. Example on time fragments.

- **Time Segmenting (TS):** Number of additional time fragments that are generated due to the reserved FS' on a route during the selected service time window is:

$$N_{TS}^a = \sum_{e \in p_{s^i, d^i}^a} \sum_{j=f_s^a}^{f_e^a} \phi_{e,j}, \quad (11)$$

where $\phi_{e,j}$ is a binary variable defined as:

$$\phi_{e,j} = \begin{cases} 0, & u_{e, t_s^a-1}[j] \oplus u_{e, t_e^a+1}[j] = 1, \\ 1, & \text{otherwise.} \end{cases} \quad (12)$$

The more time fragments we introduce, the more severe 2-D spectrum fragmentation there is. Considering the fact that 2-D spectrum fragmentation increases IR service interruptions and reduces spectrum utilization [31], we try to minimize N_{TS}^a when choosing the scheduling scheme for AR. We show an example on time fragments in Fig. 3. Suppose that the AR request $AR(0, 3, 2, t_3 + 1, 2, t_3 + 3)$ arrives in the network in Fig. 2(a), and we only consider the path candidate $0 \rightarrow 1 \rightarrow 3$ for simplicity. The figure also plots spectrum usage on the links on the route along the time axis. According to the time parameters of AR, two service time windows are feasible, i.e., $[t_3 + 1, t_3 + 2]$ and $[t_3 + 2, t_3 + 3]$. During each one, a suitable FS block has been marked with a rectangle, i.e., FS block [7, 8] during $[t_3 + 1, t_3 + 2]$ and FS block [4, 5] during $[t_3 + 2, t_3 + 3]$. We observe that no time fragment will be introduced by FS block [7, 8] during $[t_3 + 1, t_3 + 2]$, since FS block [7, 8] at TS' t_3 and $t_3 + 3$ have already been reserved on both links. However, FS block [4, 5] during $[t_3 + 2, t_3 + 3]$ introduces two time fragments for FS' 4 and 5 on Link $0 \rightarrow 1$. Thus, we prefer to reserve FS block [7, 8] on Path $0 \rightarrow 1 \rightarrow 3$ during $[t_3 + 1, t_3 + 2]$ for AR.

- **Traffic Intensity (TI):** This factor is considered for non-uniform IR traffic. It is known that traffic in a backbone network usually follows an uneven but predictable pattern when being observed over a relatively long period of time (e.g., a day or a month) due to the habitual usage of network users. Fig. 4 illustrates an uneven IR traffic profile. There are nine time periods during each of which the IR traffic is marked with an intensity mark, namely, high (H), medium (M), or low (L). With the knowledge of IR TI, we can reduce IR service failures by trying to schedule AR

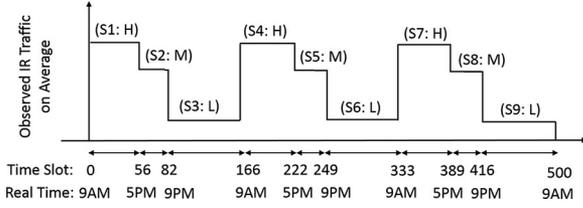


Fig. 4. Example of an uneven IR traffic profile.

requests within the off-peak periods. However, serving too many AR requests in off-peak periods may create undesirable peak periods as well, and result in high IR service failures. Hence, regarding TI awareness, we consider the TI of both IR and AR traffics. We denote the time period set as $\{S_n : n \in [1, \Delta]\}$, where Δ is the number of time periods. The IR traffic profile is $\{\xi_t^i : t \in [1, T]\}$, where ξ_t^i indicates the IR TI at t , which is the average value within $[1, T]$. We define the following variables:

- $\{\psi_t^a : t \in [1, T]\}$: scheduled AR traffic profile where ψ_t^a denotes the AR TI scheduled at t . Whenever an AR request $AR(s^a, d^a, n^a, \zeta^a, \delta^a, \tau^a)$ has been successfully scheduled within the service time window $[t_s^a, t_e^a]$, we need to update the traffic profile as:

$$\psi_t^a = \psi_t^a + n^a, \forall t \in [t_s^a, t_e^a]. \quad (13)$$

- $\{\xi_t^a : t \in [1, T]\}$: average AR TI, where ξ_t^a is:

$$\xi_t^a = \frac{\sum_{\tau \in S_p} \psi_\tau^a}{|S_p|}, \quad (14)$$

and S_p represents the time period that t belongs to. Note that, when $\{\psi_t^a : t \in [1, T]\}$ has been updated, $\{\xi_t^a : t \in [1, T]\}$ needs to be updated accordingly.

- ω_{t_s, t_e} : total average TI of IR and AR traffics within time window $[t_s, t_e]$, which is calculated as:

$$\omega_{t_s, t_e} = \frac{\sum_{t=t_s}^{t_e} \xi_t^i + \xi_t^a}{t_e - t_s + 1}. \quad (15)$$

Before scheduling an AR request, we first get all the feasible service time windows, denoted as set $\{[t_{s,n}^a, t_{e,n}^a]\}$. Then, we use *Algorithm 2* to shrink the size of $\{[t_{s,n}^a, t_{e,n}^a]\}$ by only keeping those that have relatively low IR and AR TI, where a factor ρ is introduced to control the ratio of remaining service time windows. *Lines 1–3* calculate $\omega_{t_{s,n}^a, t_{e,n}^a}^a$ for each $[t_{s,n}^a, t_{e,n}^a]$. Then, with $\{\omega_{t_{s,n}^a, t_{e,n}^a}^a\}$, we get the lower and upper bounds of $\omega_{t_{s,n}^a, t_{e,n}^a}^a$ in *Line 4*. Finally, *Line 5* removes certain AR service time windows accordingly. Note that we only apply *Algorithm 2* when an AR request can be scheduled within multiple time periods.

To consider the first three factors (i.e., SE, SM, and TS)¹ jointly, we put them into two metrics for evaluating the value of an AR scheduling scheme, i.e., 1) spectrum weight as:

$$W_S^a = N_{SE}^a + N_{SM}^a, \quad (16)$$

¹TI is not considered here, as it will be addressed when we shrink the size of AR service time window set before AR scheduling.

Algorithm 2: Cutting Down AR Service Time Windows

Input: $\{[t_{s,n}^a, t_{e,n}^a]\}$, $\{\xi_t^i : t \in [1, T]\}$, $\{\xi_t^a : t \in [1, T]\}$, ρ ;

Output: $\{[t_{s,n}^a, t_{e,n}^a]\}$;

- 1 **for each** $[t_{s,n}^a, t_{e,n}^a]$ **do**
- 2 calculate $\omega_{t_{s,n}^a, t_{e,n}^a}^a$ according to Eq. (15);
- 3 **end**
- 4 get the lower bound as $\min(\{\omega_{t_{s,n}^a, t_{e,n}^a}^a\})$ and the upper bound as $\min(\{\omega_{t_{s,n}^a, t_{e,n}^a}^a\}) + \rho \cdot (\max(\{\omega_{t_{s,n}^a, t_{e,n}^a}^a\}) - \min(\{\omega_{t_{s,n}^a, t_{e,n}^a}^a\}))$;
- 5 remove the time windows whose $\omega_{t_{s,n}^a, t_{e,n}^a}^a$ is larger than the upper bound;

and 2) time weight that is calculated as:

$$W_T^a = N_{TS}^a. \quad (17)$$

For these two metrics, W_S^a is primary while W_T^a is secondary since we want to emphasize on spectrum savings. Specifically, we try to find AR scheduling scheme(s) with minimum W_S^a , and if there exist multiple such AR scheduling schemes, we select the one with the minimum W_T^a .

Algorithm 3 shows the overall procedure of the proposed proactive AR scheduling algorithm. When $AR(s^a, d^a, n^a, \zeta^a, \delta^a, \tau^a)$ arrives, we first leverage the TI awareness to cut down its service time windows $\{[t_{s,n}^a, t_{e,n}^a]\}$, and input the new set to *Algorithm 3*. Then, we load the K precalculated path candidates, as explained in *Line 1*. *Lines 2–19* consider each service time window $[t_{s,n}^a, t_{e,n}^a]$ and path candidate $p_{s^a, d^a, k}$ to find the optimal FS block using the designed metrics, and to record the AR scheduling scheme with its W_S^a and W_T^a . If no such FS block exists, *Line 16* records the result accordingly. Finally, if feasible AR scheduling scheme(s) exist, *Lines 21 and 22* determine the optimal scheme using the designed evaluation mechanism, set $[t_s^a, t_e^a]$, p_{s^a, d^a} and $[f_s^a, f_e^a]$, and record AR as scheduled; otherwise, *Line 24* marks AR as blocked. Note that we perform spectrum allocation for the AR requests in the opposite direction of that for the IR ones, i.e., from the largest FS index to the smallest FS index, to reduce IR/AR service conflicts. This is why we prefer to select the one with the highest start FS when there are multiple optimal AR scheduling schemes, as shown in *Lines 13 and 21*.

B. Reactive Algorithms

In addition to the proactive ones, we also design reactive algorithms to deal with IR/AR service conflicts. When an IR service is about to be interrupted, we have two options to avoid it: 1) reconfigure IR to use a new lightpath, or 2) delay the scheduled AR requests that are still flexible to release more spectrum resources for IR . The first one is relatively straightforward, but lightpath reconfigurations can cause high operational cost. For the second one, we need to design a sophisticated AR rescheduling algorithm, while the advantage is that no operational cost would be induced as the scheduled AR requests have not been physically set up yet.

Here, we design *Algorithm 4* to apply the first scheme, and evaluate the operational cost from lightpath reconfigurations.

Algorithm 3: Proactive AR Scheduling

Input: $AR(s^a, d^a, n^a, \zeta^a, \delta^a, \tau^a), \{[t_{s,n}^a, t_{e,n}^a]\}, \{u_{e,t} : \forall e, t \in [\zeta^a, \tau^a]\};$
Output: $[t_s^a, t_e^a], p_{s^a, d^a}, [f_s^a, f_e^a];$

- 1 load K -shortest path candidates $\{p_{s^a, d^a, k}\}$ between s^a and d^a ;
- 2 **for each** $[t_{s,n}^a, t_{e,n}^a]$ **do**
- 3 **for each** $p_{s^a, d^a, k}$ **do**
- 4 get FS usage on $p_{s^a, d^a, k}$ within $[t_{s,n}^a, t_{e,n}^a]$;
- 5 find available FS blocks larger than n^a FS’;
- 6 **if** such FS block(s) exist **then**
- 7 denote the available FS block(s) as set $\{[f_{s,k,n,m}^a, f_{e,k,n,m}^a]\};$
- 8 **for each** FS block $[f_{s,k,n,m}^a, f_{e,k,n,m}^a]$ **do**
- 9 tailor a just-enough FS block by reducing $f_{s,k,n,m}^a$ to $f_{e,k,n,m}^a - n^a + 1$;
- 10 calculate W_S^a with Eq. (16);
- 11 calculate W_T^a with Eq. (17);
- 12 **end**
- 13 set $[f_{s,k,n}^a, f_{e,k,n}^a]$ on $p_{s^a, d^a, k}$ during $[t_{s,n}^a, t_{e,n}^a]$ as the one with minimum W_S^a and W_T^a but highest start-FS index;
- 14 record the AR scheduling scheme and its W_S^a and W_T^a ;
- 15 **else**
- 16 record no AR scheduling scheme found on $p_{s^a, d^a, k}$ during $[t_{s,n}^a, t_{e,n}^a]$;
- 17 **end**
- 18 **end**
- 19 **end**
- 20 **if** feasible AR scheduling scheme(s) exist **then**
- 21 set $[t_s^a, t_e^a], p_{s^a, d^a}$ and $[f_s^a, f_e^a]$ as the one with minimum W_S^a and W_T^a but highest start-FS index;
- 22 record AR as scheduled;
- 23 **else**
- 24 mark AR as rejected;
- 25 **end**

We will consider the second scheme in our future work. In *Algorithm 4*, the maximum lightpath reconfiguration times are set as M for each IR , and variable m^i is initialized as 0 for each IR and then used to record the number of lightpath reconfigurations that have been performed. *Line 1* checks whether the lightpath reconfigurations for IR have been run out. If not, *Algorithm 1* is triggered to get the new RSA solution, as in *Lines 2* and *3*; otherwise, *Line 5* reports that IR is interrupted. Therefore, if $M = 0$, no reactive IR reprovisioning would be allowed and we only have proactive IR provisioning; otherwise, both proactive IR provisioning and reactive IR reprovisioning would be performed with $M > 0$.

C. Complexity Analysis

The time complexity of *Algorithm 1* is $O(K \cdot B \cdot |E| \cdot T \cdot n^i)$. The time complexity of *Algorithm 2* is $O(\gamma^a \cdot \delta^a + 1)$, where $(\gamma^a \cdot \delta^a + 1)$ is the number of feasible service time

Algorithm 4: Reactive IR Reprovisioning

Input: $IR(s^i, d^i, n^i, \zeta^i), m^i, t_c, \{u_{e,t} : \forall e, t \in [t_c, T]\}, \varpi, M;$
Output: $p_{s^i, d^i}, [f_s^i, f_e^i], m^i;$

- 1 **if** $m^i < M$ **then**
- 2 $m^i = m^i + 1$;
- 3 perform *Algorithm 1*;
- 4 **else**
- 5 mark IR as interrupted;
- 6 **end**

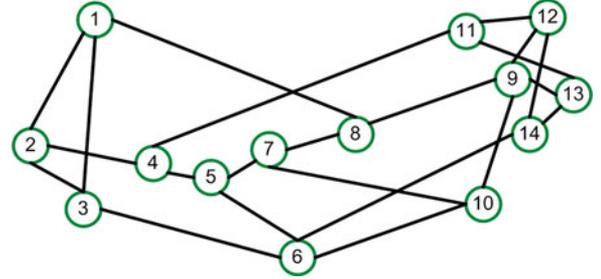


Fig. 5. NSFNET topology.

windows. For *Algorithm 3*, its time complexity is $O((\gamma^a \cdot \delta^a + 1) \cdot K \cdot B \cdot |E| \cdot n^a \cdot \delta^a)$, where $O(K \cdot B \cdot |E| \cdot n^a \cdot \delta^a)$ is the time complexity of the for-loop that covers *Lines 3–18*. With *Algorithm 2*, the time complexity of *Algorithm 3* can be further reduced due to the reduced service time windows. In the worst case, an IR request can have M lightpath reconfigurations, and the complexity of *Algorithm 4* to calculate them is $O(M \cdot K \cdot B \cdot |E| \cdot T \cdot n^i)$.

V. PERFORMANCE EVALUATION

In this section, we conduct numerical simulations to evaluate the proposed algorithms.

A. Simulation Parameters

Simulations use the NSFNET topology in Fig. 5 [32] as the physical topology. Each fiber link is assumed to accommodate 358 FS’, each of which can provide a capacity of 12.5 Gb/s². The maximum number of look-ahead TS’ that can be observed by the network operator is set as $T = 500$ TS’. When generating the IR requests, we consider both uniform and non-uniform traffic patterns and simulate around 6×10^4 IR requests in each simulation to ensure that the result on IR service failure ratio has sufficient statistical accuracy. For each $IR(s^i, d^i, n^i, \zeta^i)$, its s^i - d^i pairs are randomly selected, the bandwidth requirement n^i is uniformly distributed within $[1, 10]$ FS’, and the duration follows a negative exponential distribution with an average of 5 TS’, which is unknown to the network operator. Fig. 6 shows the examples of aggregated IR traffic patterns. To obtain the average IR traffic profile $\{\xi_t^i : t \in [1, T]\}$, we randomly generate several

²Supposing C-Band is deployed in the network, each fiber link has ~ 4.475 THz bandwidth to allocate.

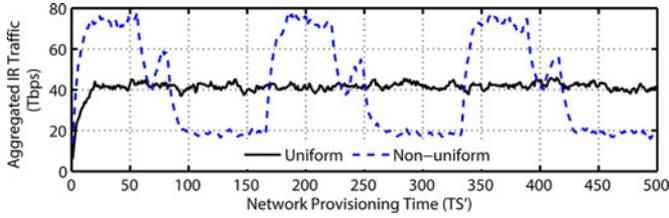


Fig. 6. Aggregated IR traffic under different traffic patterns.

sets of dynamic IR requests and calculate average TI during each time period S_n , $n \in [1, \Delta]$. Meanwhile, we generate AR requests with a Poisson traffic model, i.e., the requests arrive with an average rate of λ^a per TS and their service durations follow a negative exponential distribution with an average of $\frac{1}{\mu^a}$ TS'. Then, the AR traffic can be quantified with $\frac{\lambda^a}{\mu^a}$ in Erlangs.

To investigate the effect of AR traffic on IR traffic, we test several AR traffic loads, which are 25, 50, 75, 100, and 125 Erlangs. For each $AR(s^a, d^a, n^a, \zeta^a, \delta^a, \tau^a)$, its s^a - d^a pair is randomly selected, bandwidth requirement n^a is uniformly distributed within $[1, 16]$ FS', earliest service start time ζ^a is set according to AR request's book-ahead time³, which ranges within $[20, 50]$ TS', duration δ^a follows a negative exponential distribution with an average of 10 TS', and latest service end time τ^a is $\tau^a = (\gamma^a + 1) \cdot \delta^a + \zeta^a - 1$ according to Eq. (1), where γ^a is the flexibility of the request and is uniformly distributed within $[0, 2]$. Here, we consider both STSD-fixed and STSD-flexible AR requests. If $\gamma^a = 0$, we have an STSD-fixed AR request; otherwise, it is STSD-flexible.

The simulations evaluate the IR + AR service provisioning algorithms with four metrics, i.e., IR service failure ratio, average RSA reconfiguration times, IR service interruption ratio, and IR service blocking ratio. IR service failure ratio represents the ratio of the IR service that is blocked or interrupted, average RSA reconfiguration times count the average number of reconfigurations that have been done for each IR service in a simulation, IR service interruption ratio is the ratio of the IR service that is disrupted to release spectrum resources for the AR service, and IR service blocking ratio is the ratio of the IR service that is blocked due to lack of spectrum resources. Actually, IR service failure ratio is the sum of IR service interruption ratio and IR service blocking ratio. Table I summarizes the key simulation parameters.

B. Effect of AR Traffic Load

To evaluate the performance of the proposed IR + AR service provisioning algorithms, we use two traditional IR + AR algorithms as baselines, namely, "KSP-FF + LSTR" and "KSP-FF + SFSSI." Both of them use K -shortest path routing and first-fit spectrum allocation scheme (KSP-FF) in [17] for IR service provisioning and reprovisioning. However, for AR service provisioning, "KSP-FF + LSTR" applies the least-spectrum-to-reserve (LSTR) AR scheduling scheme in [8] while "KSP-FF + SFSSI" incorporates the smallest-FS-starting-index (SFSSI)

TABLE I
SIMULATION PARAMETERS

Network topology	NSFNET
T , Number of look-ahead TS' for network operator	500
B , Number of FS' on each fiber link	358
Number of IR requests in each simulation	$\sim 6 \times 10^4$
Bandwidth requirements of IR requests	$[1, 10]$ FS'
Bandwidth requirements of AR requests	$[1, 16]$ FS'
Book-ahead time of AR requests	$[20, 50]$ TS'
γ^a , Flexibility of AR requests	$[0, 2]$
K , Number of alternate-path candidates	5
ϖ , Guaranteed available time of IR requests	10 TS'
M , Maximum RSA reconfigurations of IR requests	5

AR scheduling scheme in [20]. For comparison, we denote our proposed IR + AR scheme as "Pro. IR + AR." We set $\rho = 1$ for the simulations that use uniform IR traffic pattern, and do not apply *Algorithm 2* to reduce the AR service time windows, i.e., the algorithm is not TI-aware. This is because, when the IR traffic pattern is uniform, there is no need to do so. On the other hand, when the IR traffic pattern is non-uniform, the simulations compare $\rho = 1$ with $\rho = 0.5$ to evaluate the effectiveness of TI awareness.

Fig. 7(a)–(c) shows the results with uniform IR traffic. We use them to indicate the effect of AR traffic load. Fig. 7(a) compares the results on IR service failure ratio. As expected, IR service failure ratio increases with AR traffic load. Due to the effort on spectrum saving, "KSP-FF + LSTR" achieves slight lower service failure ratio than "KSP-FF + SFSSI." More importantly, our proposed algorithm, i.e., "Pro. IR + AR, $\rho = 1$," always achieves much lower IR service failure ratio than the two baselines, which verifies its effectiveness on improving IR service provisioning quality.

Fig. 7(b) shows the results on average RSA reconfiguration times for each IR in case of $M = 5$. With increase of AR traffic load, the average RSA reconfiguration times grow quickly, due to the rise of IR/AR service conflicts. However, "Pro. IR + AR, $\rho = 1$ " can reduce the average RSA reconfigurations by half or more, when compared to the benchmarks. This means that our proposed algorithm can achieve significant reduction on the operational cost. Fig. 7(c) compares the results on IR service interruption ratio. Again, "Pro. IR + AR, $\rho = 1$ " achieves the lowest IR service interruption ratio, followed by "KSP-FF + SFSSI" and "KSP-FF + LSTR." In all, the results discussed above indicate that both spectrum efficiency and IR/AR load balancing are important for handling hybrid IR/AR service provisioning well, which is also the reason why our proposed IR + AR scheme can always perform the best.

Fig. 8(a)–(c) shows the results with non-uniform IR traffic. Now, to evaluate the effectiveness of TI awareness, we test both $\rho = 1$ and $\rho = 0.5$. When comparing the results from "KSP-FF + LSTR," "KSP-FF + SFSSI" and "Pro. IR + AR, $\rho = 1$," we get the same conclusions as those from Fig. 7(a)–(c). Then, we compare "Pro. IR + AR, $\rho = 1$ " and "Pro. IR + AR, $\rho = 0.5$." In Fig. 8(a), we observe that "Pro. IR + AR, $\rho = 0.5$ " achieves lower IR service failure ratio for all the AR traffic loads, indicating the effectiveness of TI awareness. We also notice that "Pro.

³Time between AR request's arrival and its earliest service start time.

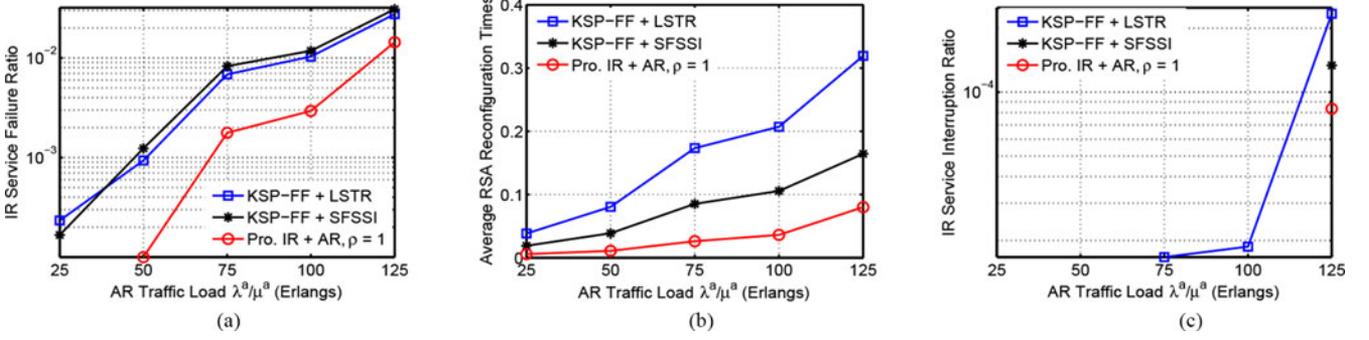


Fig. 7. Effect of AR traffic load with uniform IR traffic. (a) IR service failure ratio. (b) Average RSA reconfigurations. (c) IR service interruption ratio.

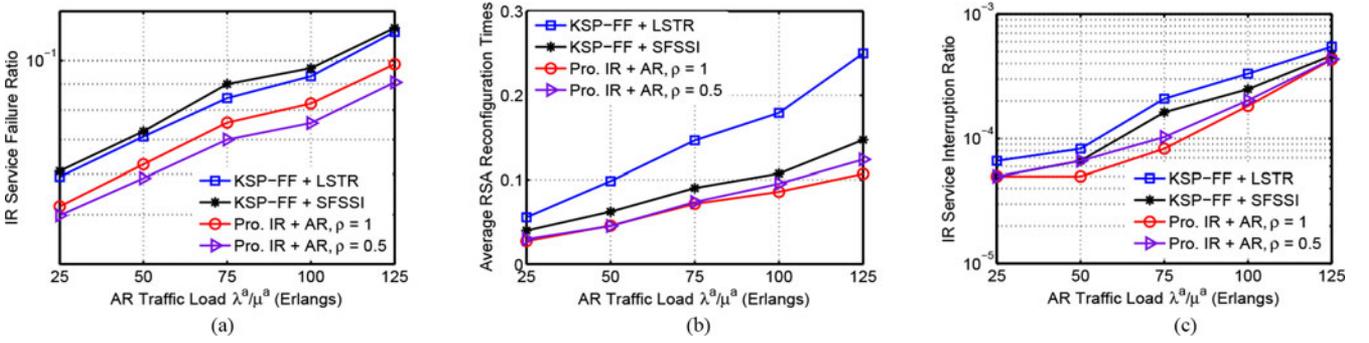


Fig. 8. Effect of AR traffic load with non-uniform IR traffic. (a) IR service failure ratio. (b) Average RSA reconfigurations. (c) IR service interruption ratio.

IR + AR, $\rho = 0.5$ ” requires more RSA reconfigurations than “Pro. IR + AR, $\rho = 1$,” as shown in Fig. 8(b). This is because, while scheduling AR requests with TI awareness of IR traffic, “Pro. IR + AR, $\rho = 0.5$ ” can create unbalanced AR traffic during different time periods, which results in more IR/AR service conflicts. Therefore, more RSA reconfigurations are needed, and then “Pro. IR + AR, $\rho = 0.5$ ” can achieve almost the same IR service interruption ratio as that from “Pro. IR + AR, $\rho = 1$,” as shown in Fig. 8(c).

It is also interesting to notice that, at the same AR traffic load, “Pro. IR + AR, $\rho = 1$ ” performs more RSA reconfigurations in the non-uniform IR traffic scenario than in the uniform one. This is because, in the non-uniform IR traffic scenario, the competitions between IR and AR requests become more intense during the traffic peaks, which makes “Pro. IR + AR, $\rho = 1$ ” interrupt more IR requests instead of serving them completely. Note that, in Figs. 7(c) and 8(c), the IR service interruption ratios are around 10^{-4} , which makes IR service blocking ratio and IR service failure ratio almost equal. This is why the results on IR service blocking ratio are not given.

C. Effect of Lightpath Reconfigurations

Frequent lightpath reconfigurations induce undesirable operational cost for network operator. Hence, we investigate the effect of maximum RSA reconfiguration times M by changing its value from 0 to 5. Fig. 9(a)–(d) shows the results on the four metrics with uniform IR traffic when $\frac{\lambda^a}{\mu^a} = 75$ Erlangs. Fig. 9(a) shows that the IR service failure ratios of all the IR + AR schemes decrease with M . But, when $M \geq 2$, the

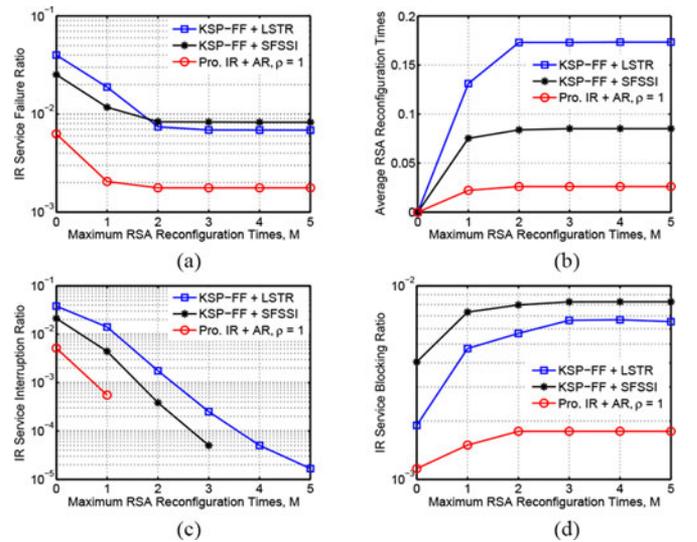


Fig. 9. Effect of maximum RSA reconfiguration times with uniform IR traffic when $\frac{\lambda^a}{\mu^a} = 75$ Erlangs. (a) IR service failure ratio. (b) Average RSA reconfigurations. (c) IR service interruption ratio. (d) IR service blocking ratio.

decreasing trend becomes less obvious, especially for “Pro. IR + AR, $\rho = 1$.” As expected, in Fig. 9(b), the average RSA reconfiguration times from all the schemes increase with M . Also, when $M \geq 2$, the average RSA reconfigurations keep almost unchanged. These observations can be explained by using the results in Fig. 9(c). Fig. 9(c) shows that, when $M \geq 2$, the IR service interruption ratios of all the schemes become very small. Hence, no more RSA reconfigurations are needed. Conversely,

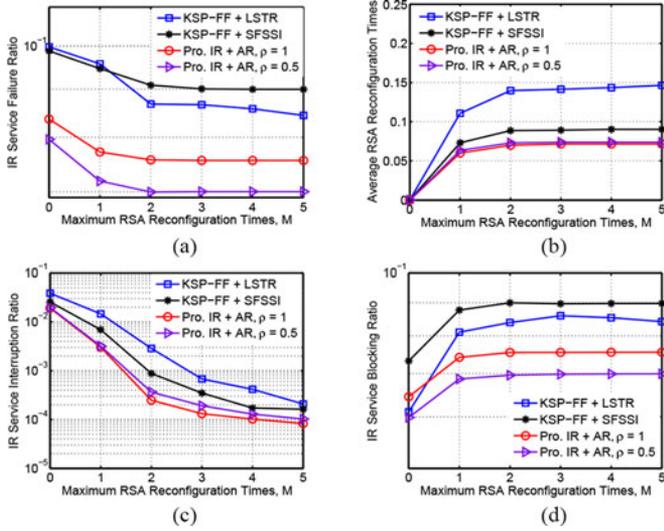


Fig. 10. Effect of maximum RSA reconfiguration times with non-uniform IR traffic when $\frac{\lambda^a}{\mu^a} = 75$ Erlangs. (a) IR service failure ratio. (b) Average RSA reconfigurations. (c) IR service interruption ratio. (d) IR service blocking ratio.

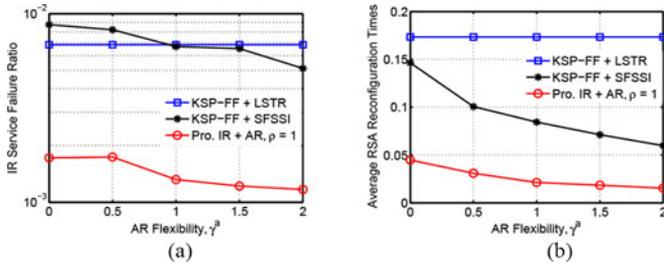


Fig. 11. Effect of AR flexibility with uniform IR traffic pattern when $M = 5$ and $\frac{\lambda^a}{\mu^a} = 75$ Erlangs. (a) IR service failure ratio. (b) Average RSA reconfigurations.

in Fig. 9(d), we note that the IR service blocking ratios of all the schemes increase significantly with M , but when $M \geq 2$, they keep almost unchanged or even turn to decrease. This observation suggests that there is a tradeoff between IR service interruption ratio and IR service blocking ratio, especially when the former is relatively high.

Fig. 10(a)–(d) shows the results with non-uniform IR traffic. In Fig. 10(a), “Pro. IR + AR, $\rho = 0.5$ ” achieves lower IR service failure ratio than “Pro. IR + AR, $\rho = 1$ ” in all cases, which verifies the effectiveness of TI awareness again. Fig. 10(b) shows that “Pro. IR + AR, $\rho = 0.5$ ” still needs more RSA reconfigurations than “Pro. IR + AR, $\rho = 1$ ” with the same M , due to the same reason explained above. With the additional RSA reconfigurations, “Pro. IR + AR, $\rho = 0.5$ ” achieves almost the same IR service interruption ratio but lower IR service blocking ratio, as shown in Fig. 10(c) and (d).

D. Effect of AR Flexibility

The flexibility of AR requests (i.e., γ^a) affects the network operator’s ability to coordinate the IR/AR traffic, i.e., a larger flexibility provides more freedom on the resource management for AR. To investigate the effect of AR flexibility, Fig. 11(a)

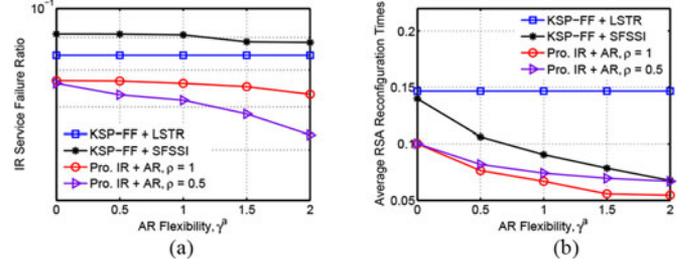


Fig. 12. Effect of AR flexibility with non-uniform IR traffic when $M = 5$ and $\frac{\lambda^a}{\mu^a} = 75$ Erlangs. (a) IR service failure ratio. (b) Average RSA reconfigurations.

and (b) shows the results on IR service failure ratio and average RSA reconfiguration times with uniform IR traffic, for different γ^a . In Fig. 11(a), we observe that “KSP-FF + LSTR” achieves no gain from AR flexibility, but both “KSP-FF + SFSSI” and “Pro. IR + AR, $\rho = 1$ ” obtain obvious gains. This is because “KSP-FF + LSTR” focuses too much on spectrum efficiency and become inflexible in AR scheduling, while the other two can utilize AR flexibility proactively to balance IR/AR traffics and benefit from this.

In Fig. 11(b), we can see the same trend on average RSA reconfiguration times. Average RSA reconfigurations required by “KSP-FF + LSTR” remain unchanged with the increase of AR flexibility, but both “KSP-FF + SFSSI” and “Pro. IR + AR, $\rho = 1$ ” can reduce average RSA reconfigurations effectively when AR flexibility is larger. Fig. 12(a) and (b) shows the results with non-uniform IR traffic. Here, “Pro. IR + AR, $\rho = 1$ ” and “Pro. IR + AR, $\rho = 0.5$ ” have similar performance under the effect of AR flexibility. However, we still notice that the decreasing trend of “Pro. IR + AR, $\rho = 0.5$ ” in Fig. 12(a) is more obvious than that of “Pro. IR + AR, $\rho = 1$,” which indicates that “Pro. IR + AR, $\rho = 0.5$ ” can achieve more gain on IR service failure ratio from the AR flexibility increase. This again verifies the effectiveness of TI awareness.

E. IR/AR Load Balancing

By considering several factors that can impact service blocking and interruption jointly, the proposed IR + AR schemes can coordinate IR and AR traffics wisely in both time and spectral domains to reduce IR/AR service conflicts. To make this conclusion more visible, Figs. 13 and 14 illustrate the aggregated traffic from IR and AR requests over time with uniform and non-uniform IR traffic patterns. Fig. 13 indicates that the IR and AR traffics can compete for network resources, i.e., an increase in AR traffic leads to a decrease in IR traffic. Moreover, we notice that the IR traffic accommodated by “Pro. IR + AR, $\rho = 1$ ” is higher than those by “KSP-FF + LSTR” and “KSP-FF + SFSSI” at most TS’. In Fig. 14(a), we can see that “Pro. IR + AR, $\rho = 0.5$ ” can accommodate even more IR traffic than “Pro. IR + AR, $\rho = 1$.” This can be explained with the results on aggregate AR traffic in Fig. 14(b). Basically, the AR traffic distribution from “Pro. IR + AR, $\rho = 0.5$ ” is also non-uniform and follows the opposite trend of that of the IR traffic in Fig. 14(a), but we cannot see such trends on the results from “Pro. IR + AR, $\rho = 1$.” This is because “Pro. IR + AR, $\rho = 0.5$ ” can purposely

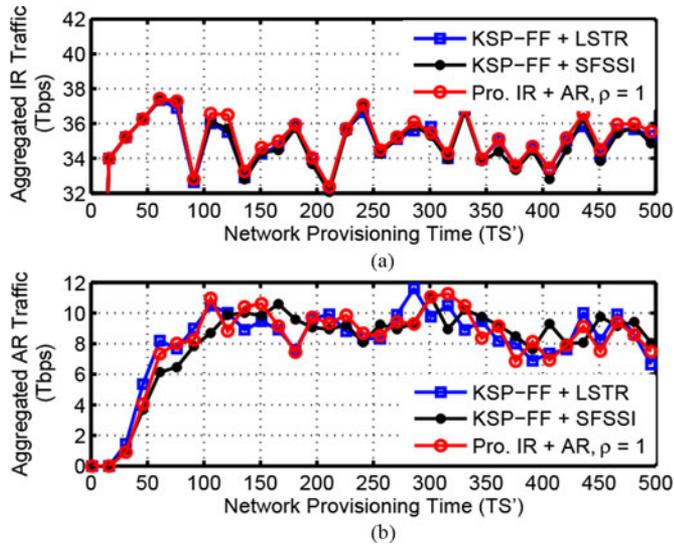


Fig. 13. Aggregated IR/AR traffics with uniform IR traffic when $\gamma^a \in [0, 2]$ and $\frac{\lambda^a}{\mu^a} = 75$ Erlangs.

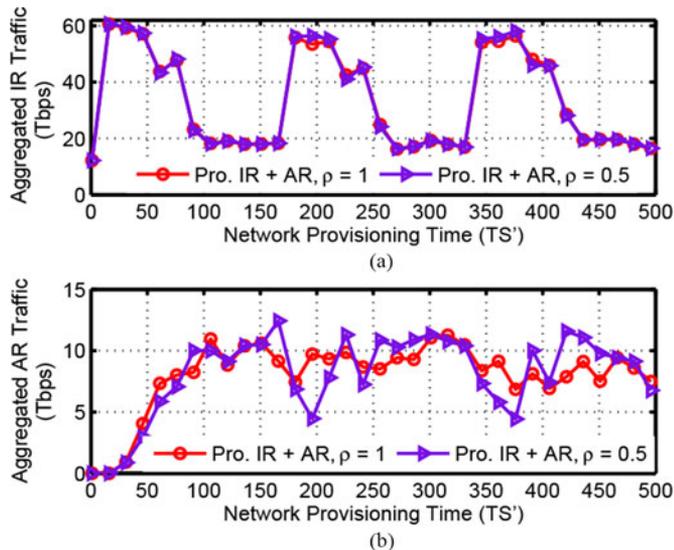


Fig. 14. Aggregated IR/AR traffics with non-uniform IR traffic when $\gamma^a \in [0, 2]$ and $\frac{\lambda^a}{\mu^a} = 75$ Erlangs.

drop service time windows that fall in peak periods of IR traffic, and thus is more likely to schedule the AR request within an off-peak service time window, while “Pro. IR + AR, $\rho = 1$ ” treats all the service time windows equally without TI awareness of IR traffic.

VI. CONCLUSION

In this paper, we investigated hybrid IR and AR service provisioning in EONS where spectrum resources are fully shared among IR and AR requests, aiming to minimize IR/AR service conflicts. We designed algorithms to provision IR and AR services in a coordinated manner. We also proposed an AR scheduling algorithm, which can coordinate AR service provisioning with various IR traffic patterns, balance spectrum utilization in both time and spectral domains, and relieve IR/AR

service conflicts. Simulation results verified that the proposed IR + AR schemes can significantly reduce IR service failures as well as RSA reconfigurations in IR service provisioning. Moreover, the results indicated that the proposed IR + AR schemes can achieve more performance gain from AR flexibility when compared with two benchmark algorithms. Note that this work has not considered spectrum defragmentation, but it is known that AR service provisioning can generate 2-D spectrum fragments in EONS. Hence, we expect that considering spectrum defragmentation together with the proposed AR rescheduling approaches would be an interesting research direction to explore in the future.

REFERENCES

- [1] O. Gerstel, M. Jinno, A. Lord, and S. Yoo, “Elastic optical networking: A new dawn for the optical layer?” *IEEE Commun. Mag.*, vol. 50, no. 2, pp. s12–s20, Feb. 2012.
- [2] L. Gong and Z. Zhu, “Virtual optical network embedding (VONE) over elastic optical networks,” *J. Lightw. Technol.*, vol. 32, no. 3, pp. 450–460, Feb. 2014.
- [3] P. Lu, L. Zhang, X. Liu, J. Yao, and Z. Zhu, “Highly-efficient data migration and backup for big data applications in elastic optical inter-data-center networks,” *IEEE Netw.*, 2015, to be published.
- [4] M. Zhang *et al.*, “Bandwidth defragmentation in dynamic elastic optical networks with minimum traffic disruptions,” in *Proc. IEEE Int. Commun. Conf.*, Jun. 2013, pp. 3894–3898.
- [5] L. Gong, W. Zhao, Y. Wen, and Z. Zhu, “Dynamic transparent virtual network embedding over elastic optical infrastructures,” in *Proc. IEEE Int. Commun. Conf.*, Jun. 2013, pp. 3466–3470.
- [6] J. Zheng and H. Moustafa, “Routing and wavelength assignment for advance reservation in wavelength-routed WDM optical networks,” in *Proc. IEEE Int. Commun. Conf.*, Aug. 2002, pp. 2722–2726.
- [7] W. Lu, S. Ma, C. Chen, X. Chen, and Z. Chen, “Implementation and demonstration of revenue-driven provisioning for advance reservation requests in OpenFlow-controlled SD-EONS,” *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1727–1730, Oct. 2014.
- [8] W. Lu and Z. Zhu, “Dynamic service provisioning of advance reservation requests in elastic optical networks,” *J. Lightw. Technol.*, vol. 31, no. 10, pp. 1621–1627, May 2013.
- [9] D. Wischik and A. Greenberg, “Admission control for booking ahead shared resources,” in *Proc. IEEE INFOCOM*, Mar. 1998, pp. 873–882.
- [10] A. Greenberg, R. Srikant, and W. Whitt, “Resource sharing for book-ahead and instantaneous-request calls,” *IEEE/ACM Trans. Netw.*, vol. 7, no. 1, pp. 10–12, Aug. 1999.
- [11] I. Ahmad, J. Kamruzzaman, and S. Aswathanarayanan, “A dynamic approach to reduce preemption in book-ahead reservation in QoS-enable networks,” *Comput. Commun.*, vol. 29, pp. 1443–1457, May 2006.
- [12] D. Rousseau, J. Triay, and V. Vokkarane, “Improving service differentiation of immediate and advance reservation in resource-partitioned optical WDM networks,” in *Proc. Int. Conf. Comput. Netw. Commun.*, Jan. 2013, pp. 173–179.
- [13] I. Ahmad and J. Kamruzzaman, “Preemption-aware instantaneous request call routing for networks with book-ahead reservation,” *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1456–1465, Nov. 2007.
- [14] Q. Wu and P. Dharam, “Advance bandwidth scheduling with minimal impact on immediate reservations in high-performance networks,” in *Proc. IEEE Netw. Oper. Manage. Symp.*, Apr. 2012, pp. 679–682.
- [15] P. Dharam, Q. Wu, and Y. Wang, “Advance bandwidth reservation with deadline constraint in high-performance networks,” in *Proc. Int. Conf. Comput., Netw. Commun.*, Feb. 2014, pp. 1041–1045.
- [16] E. Escalona, S. Spadaro, J. Comellas, and G. Junyent, “Advance reservation algorithm for service oriented optical networks,” presented at the Eur. Conf. Optical Communication, Berlin, Germany, Sep. 2007.
- [17] K. Christodoulopoulos, I. Tomkos, and E. Varvarigos, “Elastic bandwidth allocation in flexible OFDM-based optical networks,” *J. Lightw. Technol.*, vol. 29, no. 9, pp. 1354–1366, May 2011.
- [18] Y. Wang, X. Cao, and Y. Pan, “A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks,” in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1503–1511.

- [19] Z. Zhu, W. Lu, L. Zhang, and N. Ansari, "Dynamic service provisioning in elastic optical networks with hybrid single-/multi-path routing," *J. Lightw. Technol.*, vol. 31, no. 1, pp. 15–22, Jan. 2013.
- [20] S. Shen *et al.*, "Dynamic advance reservation multicast in data center networks over elastic optical infrastructure," presented at the Eur. Conf. Optical Communication, London, U.K., Sep. 2013.
- [21] Y. Yin *et al.*, "Spectral and spatial 2D fragmentation-aware routing and spectrum assignment algorithms in elastic optical networks," *J. Opt. Commun. Netw.*, vol. 5, pp. A100–A106, Oct. 2013.
- [22] W. Lu and Z. Zhu, "Malleable reservation based bulk-data transfer to recycle spectrum fragments in elastic optical networks," *J. Lightw. Technol.*, vol. 33, no. 10, pp. 2078–2086, May 2015.
- [23] N. Charbonneau and V. Vokkarane, "A survey of advance reservation routing and wavelength assignment in wavelength-routed WDM optical networks," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 1037–1064, Oct. 2012.
- [24] J. Kuri, N. Puech, M. Gagnaire, and E. Dotaro, "Routing and wavelength assignment of scheduled lightpath demands," *IEEE/ACM Trans. Netw.*, vol. 21, no. 8, pp. 1231–1240, Oct. 2003.
- [25] C. Saradhi and M. Gurusamy, "Scheduling and routing of sliding scheduled lightpath demands in WDM optical networks," presented at the Optical Fiber Communication Conf., Anaheim, CA, USA, Mar. 2007.
- [26] A. Jaekel and Y. Chen, "Demand allocation without wavelength conversion under a sliding scheduled traffic model," in *Proc. BroadNets*, Sep. 2007, pp. 495–503.
- [27] D. Andrei *et al.*, "Integrated provisioning of sliding scheduled services over WDM optical networks," *J. Opt. Commun. Netw.*, vol. 1, pp. A94–A105, Oct. 2009.
- [28] W. Su, G. Sasaki, C. Su, and A. Balasubramanian, "Scheduling of periodic connections with flexibility," *Opt. Switching Netw.*, vol. 3, pp. 158–172, Dec. 2006.
- [29] A. Mahimkar *et al.*, "Bandwidth on demand for inter-data center communication," in *Proc. ACM HotNets*, Nov. 2011, pp. 24–29.
- [30] J. Yao, P. Lu, L. Gong, and Z. Zhu, "On fast and coordinated data backup in geo-distributed optical inter-datacenter networks," *J. Lightw. Technol.*, vol. 33, no. 14, pp. 3005–3015, Jul. 2015.
- [31] W. Shi, Z. Zhu, M. Zhang, and N. Ansari, "On the effect of bandwidth fragmentation on blocking probability in elastic optical networks," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2970–2978, Jul. 2013.
- [32] Z. Zhu, X. Chen, F. Ji, L. Zhang, F. Farahmand, and J. P. Jue, "Energy-efficient translucent optical transport networks with mixed regenerator placement," *J. Lightw. Technol.*, vol. 30, no. 19, pp. 3147–3156, Oct. 2012.

Authors' biographies not available at the time of publication.