Dynamic Advance Reservation Multicast in Data Center Networks over Elastic Optical Infrastructure

Sheng Shen(1), Wei Lu(1), Xiahe Liu(1), Long Gong(1), Zuqing Zhu(1)*

(1) University of Science and Technology of China, Hefei, Anhui 230027, China. *Email: zqzhu@ieee.org

Abstract *We investigate dynamic advanced reservation (AR) multicast in data center networks over elastic optical infrastructure, and propose several algorithms to realize all-optical AR multicasting for data center backup and migration by considering request scheduling and RSA jointly.*

Introduction

Nowadays, with the boosting up of bandwidth-intensive applications, such as cloud computing, e-Science, and etc., data center networks have exhibited the characteristics of huge throughput and high traffic burstiness. Meanwhile, recent advances on the optical orthogonal frequency-division multiplexing (O-OFDM) technology have demonstrated over Tb/s transmission capacity and flexible bandwidth allocation with granularity at 12.5 GHz or less[1]. Therefore, the elastic optical infrastructure based on O-OFDM becomes a promising candidate for carrying data center networks, as it achieves agile optical spectrum management and can facilitate seamless integration of physical transmission and upper-layer applications.

In data center networks, multicast is widely used to support point-to-multipoint applications, such as distributed computing, data center backup, and data center migration[2]. Moreover, these applications can require advanced reservation (AR) services[3], in which the requests allow certain setup delay, as long as the network resources are allocated before a preset deadline. With a static traffic model, previous work has studied AR multicast in wavelength-division multiplexing (WDM) networks[3]. However, WDM may not be a proper technology to support data center networks directly due to its coarse spectrum allocation granularity. Also, the highly dynamic traffic in data center networks makes the static traffic model too simple to be practical.

For the first time, this paper investigates dynamic AR multicast in data center networks over elastic optical infrastructure. We propose several algorithms to realize all-optical AR multicasting over elastic optical infrastructure for data center backup and migration. The proposed algorithms consider request scheduling and routing and spectrum assignment (RSA) jointly and optimize the network performance in terms of blocking probability and request setup delay.

AR Multicast over Elastic Optical Infrastructure

We consider the physical topology of the data center network as a directed graph $G(V,E)$, where V is the node set, and E is the set of fiber links. The data center network is built over an elastic optical infrastructure, which is interconnected with O-OFDM transponders. On each fiber link $e \in E$, the transponders can utilize B frequency slots (FS') at most, under the operation principle of O-OFDM[1]. We assume that all nodes in V can realize all-optical multicasting with light-splitting, since it is known that supporting multicast at the optical layer has a few benefits, such as low latency, low power consumption and etc[4]. The tasks of data center backup or migration can invoke an AR multicast request denoted as $R(s,D,t_a,t_b,t_h,n)$, where s is the source node, D is the set of the destination nodes, t_a is the arrival time, t_b is the book-ahead time, t_h is the holding time, and n refers to the bandwidth requirement in terms of FS'. Specifically, the definition of $R(s,D,t_a,t_b,t_h,n)$ determines that the multicast session should start no later than t_a+t_b, and will be active for a period of t_h.

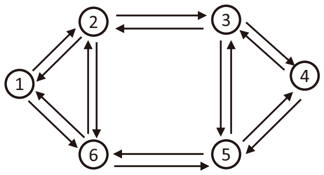
In order to serve the AR multicast request, we need to construct a light-tree from s to all destinations in D, and assign n contiguous FS' (i.e., the FS window) on it. Meanwhile, its “service time window” has a length of t_h and can slide from $[t_a,\ (t_a+t_h-1)]$ to $[(t_a+t_b),\ (t_a+t_b+t_h-1)]$. The actual start time is determined by the request scheduling algorithm. An intuitive example is shown in Fig. 1. With the six-node physical topology in Fig. 1(a), we have an AR multicast request arrives at $t=1$, with s as *Node* 1, D as *Nodes* 2 and 5, $t_b=1$, $t_h=2$ and $n=3$. As shown in Fig. 1(b), the RSA algorithm finds the light-tree that consists of *Links* 1-2, 2-6, and 6-5, and allocates the FS window [3, 5] on it. Meanwhile, the scheduling algorithm determines the service time window as [2, 3].

Proposed Algorithms

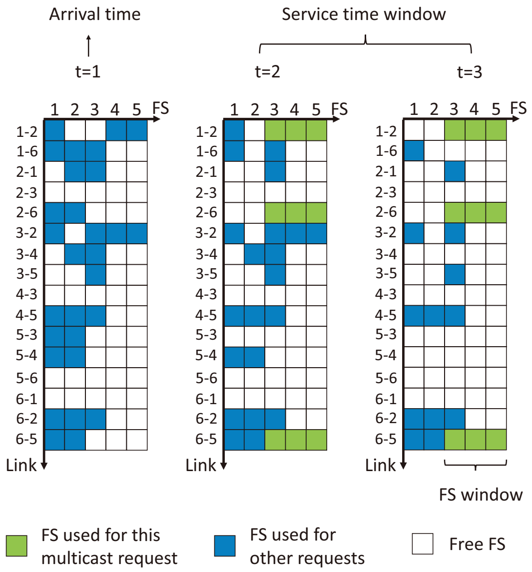
We adopt the minimum spanning tree (MST) algorithm[5] to calculate the light-tree for each AR multicast request. For the RSA of multicast requests, we incorporate two schemes together with MST, which are multicast-capable separated RSA (MC-S-RSA) and multicast-capable integrated RSA (MC-I-RSA).

For MC-S-RSA, we build a light-tree in $G(V,E)$ for $s \rightarrow D$ using MST, and then perform the first-fit spectrum assignment on it according to the spectrum utilization in the service time window determined by the scheduling algorithm.

The MC-I-RSA accomplishes the light-tree calculation and spectrum assignment in one step with the assistance of layered auxiliary graphs. According to the spectrum utilization in the service time window determined by the scheduling algorithm, we decompose $G(V,E)$ into layers. In order to construct the k-th layer,



(a) Physical topology.

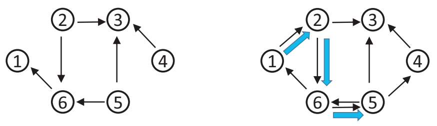


(b) Spectrum utilization.

Fig. 1: An example of dynamic AR multicast provisioning.

$G^k(V^k,E^k)$, we check the spectrum utilization, and insert a direct link $e^k=(u^k,v^k)$ in $G^k(V^k,E^k)$, if there are n available contiguous FS' starting from the k-th FS on $e=(u,v)$ in $G(V,E)$. Therefore, if we can obtain a light-tree for $s^k \rightarrow D^k$ in $G^k(V^k,E^k)$ with MST, the multicast request can be provisioned with the FS window $[k,\ (k+n-1)]$ on it in $G(V,E)$.

Fig. 2 shows an example of MC-I-RSA, in which the layers are constructed from the spectrum utilization at $t=2$ in Fig. 1(b). Since the multicast request needs 3 FS' and the FS window [1, 3] is available on *Links* 2-3, 2-6, 4-3, 5-3, 5-6 and 6-1, the *1*-st layer only consists of these links as shown in Fig. 2(a). Apparently, we cannot obtain a feasible light-tree in the *1*-st layer for the multicast session from *Node* 1 to *Nodes* 2 and 5. Therefore, we move on to construct the 2-nd and 3-rd layers. Finally, in the 3-rd layer, we obtain a light-tree as shown in Fig. 2(b), and hence the multicast request is provisioned with the FS window [3, 5] using *Links* 1-2, 2-6 and 6-5.



(a) Topology of the 1-st layer. (b) Topology of the 3-rd layer.

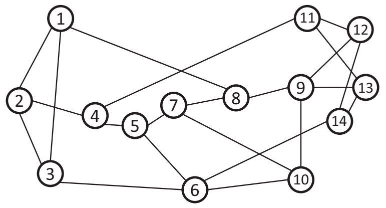
Fig. 2: Constructing layered auxiliary graphs in MC-I-RSA.

Basically, we need to find an FS window and a service time window to provision an AR multicast request. Therefore, together with the two RSA schemes, we adopt two request scheduling algorithms based on either the service time window or the FS window. The first one is the least time to wait (LTW) algorithm, which aims to provisioning the request at the earliest possible start time. The second one is the smallest FS starting index (SFSSI) algorithm, which tries to schedule the request with a FS window that has the smallest FS starting index, when the book-ahead time t_b permits. The rationale behind the SFSSI scheduling algorithm is to allocate FS' evenly on the links.

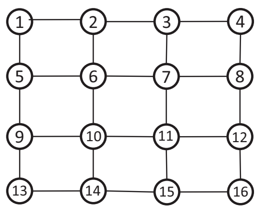
Performance Evaluation

We evaluate the four combinations of the RSA and scheduling algorithms, i.e., MC-S-RSA-LTW, MC-S-RSA-SFSSI, MC-I-RSA-LTW, and MC-I-RSA-SFSSI, with the 14-node NSFNET and 16-node Grid topologies in Fig. 3. For simplicity, we assume that the link lengths in the topologies are identical, i.e., the data center network is interconnected with fibers that have the same length. Note that the proposed algorithms can also work with topologies that have different link lengths, as the MST algorithm does not have any restriction on link lengths.

We assume that the elastic optical infrastructure is deployed in the C-band and the bandwidth of each FS is 12.5 GHz, corresponding to 358 FS' on each fiber link. The source and destinations of the AR multicast requests are randomly selected, and the number of destinations in each request is uniformly distributed within [2, 5]. The bandwidth requirement of each request is uniformly distributed within [1, 10] FS', corresponding to a throughput of [12.5, 125] Gb/s, and the book-ahead time t_b is randomly selected from [1, 5]. The requests are generated using the Poisson traffic model with an average holding time of $\overline{t_h}=10$.



(a) 14-node NSFNET topology.



(b) 16-node Grid topology.

Fig. 3: Simulation topologies.

Figs. 4 and 5 show the results on request blocking probability in the two topologies. We observe that in general, the algorithms with MC-I-RSA achieve much smaller blocking probabilities than those with MC-S-RSA. Also, it is interesting to notice that when the RSA schemes are the same, the LTW scheduling algorithm provides smaller blocking probabilities than SFSSI. We believe this observation can be explained as follows. As LTW tries to schedule the requests as early as possible, it avoids the spectrum fragmentation in the time domain, i.e., spectrum segments that are isolated along the time axis. Note that in the simulations, the book-ahead time t_b of the AR multicast requests is within [1, 5], while the average holding time t_h is 10. Therefore, if SFSSI delays the setup of an AR request for the sake of minimizing the FS starting index, it could generate spectrum segments that only have "short" availability along the time axis. Consequently, future requests can be blocked due to the low availability of spectrum resources. This explanation can be verified with Fig. 6, which plots the spectrum utilization in the Grid topology when the traffic load is 300 Erlangs. It can be seen that throughout the provision period, the spectrum utilization from the algorithms with SFSSI is lower than that from those using LTW.

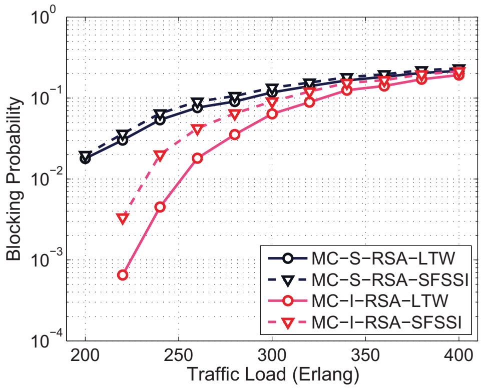


Fig. 4: Blocking probability in the NSFNET topology.

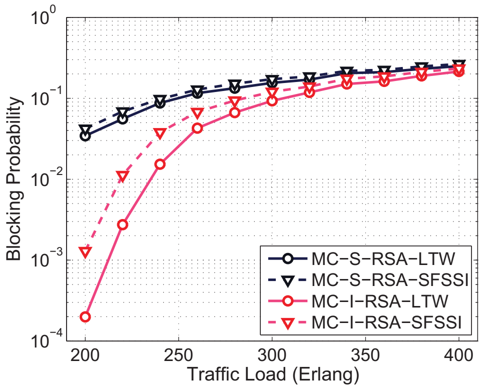


Fig. 5: Blocking probability in the Grid topology.

Fig. 7 indicates that the average setup delays from the algorithms with MC-I-RSA are comparable to those with MC-S-RSA, when the scheduling algorithm is the same. While when the RSA scheme is the same, the LTW scheduling achieves effective reduction on the average setup delay, compared with SFSSI. In all, MC-I-RSA-LTW achieves the smallest blocking probabilities for all traffic loads among the four algorithms, and the average setup delay from it is also reasonably small. We therefore consider MC-I-RSA-LTW as an effective algorithm for serving AR multicast requests in the elastic optical infrastructure.

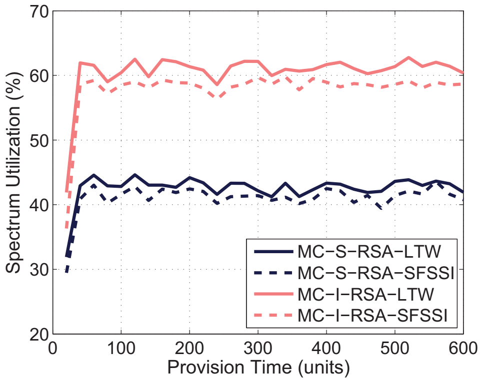


Fig. 6: Spectrum utilization versus provision time in the Grid topology with traffic load at 300 Erlangs.

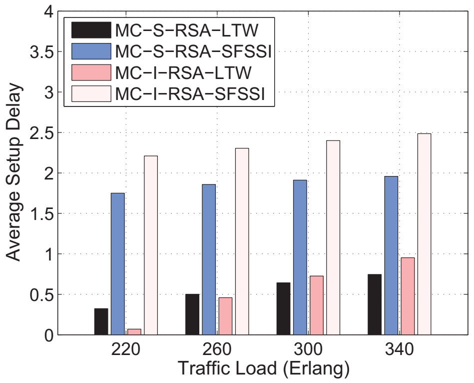


Fig. 7: Average setup delay in the Grid topology.

Conclusions

This paper investigated dynamic multicast AR in data center networks over elastic optical infrastructure. We proposed several algorithms to realize all-optical AR multicasting over elastic optical infrastructure for data center backup and migration. The simulation results indicated that the MC-I-RSA-LTW algorithm provides the smallest blocking probability and reasonably small average setup delay.

Acknowledgments

This work was supported in part by the Project NCET-11-0884, and the Natural Science Foundation of Anhui Province under Project 1208085MF88.

References

[1] O. Gerstel et al., IEEE Commun. Mag. **50**, S12 (2012).
[2] J. Zhang et al., Proc. OFC'13, PDP5B (2013).
[3] N. Charbonneau et al., IEEE/ACM Trans. Netw., **20**, 1 (2012).
[4] G. Rouskas, IEEE Network, **17**, 60 (2003).
[5] L. Kou et al., Acta Inform., **15**, 141 (1981).