Bit Index Explicit Replication (BIER) Multicasting in Transport Networks [Invited]

A. Giorgetti, A. Sgambelluri, F. Paolucci, N. Sambo, P. Castoldi
Scuola Superiore Sant'Anna
Pisa, Italy
Email: a.giorgetti@santannapisa.it

F. Cugini
CNIT
Pisa, Italy
Email: filippo.cugini@cnit.it

***Abstract*—Live events and high quality video delivery are pushing the introduction of dynamic multicasting solutions.**

In this paper we discuss the most recent technologies enabling effective multicasting in packet and optical transport networks. In particular, both point-to-multi-point (P2MP) optical coherent technology as well as the newly proposed Bit Index Explicit Replication (BIER) packet technology are presented.

Experimental demonstrations on both technologies are then reported, showing the effective capability to flexibly and dynamically configuring P2MP communications.

I. Introduction

The effective utilization of dynamic multicasting communications may significantly improve the transport network performance, successfully addressing the expected increase of video traffic and multimedia content from cloud systems located in the backbone/metro to a wide audience of subscribers.

Both optical and packet networks may significantly benefit from effective point-to-multi-point (P2MP) solutions.

In the context of optical networks, P2MP (i.e., *light-trees*) is still confined to the research level. Many valuable studies have investigated both routing and spectrum assignment schemes as well as experimental demonstrations, also focusing on the recently introduced Elastic Optical Networks (EONs) [1]–[8].

In this paper, we first review the required data and control plane technologies enabling effective optical P2MP. Then, we propose an effective networking strategy which takes advantage of the Optical Signal to Noise Ratio (ONSR) margin on less impaired branches to reduce the overall amount of spectrum resources occupied by light-trees. The proposed strategy is then validated in a P2MP network testbed including a commercial 100G transmission system.

In the context of packet metro/backbone networks, so far, dynamic multicasting solutions have not been widely deployed. One of the major motivations stands in the high complexity and low scalability of current P2MP distributed solutions in packet networks. For example, the currently adopted solutions based on Internet Group Management Protocol (IGMP) and Protocol Independent Multicast (PIM) require explicit signaling operations and intermediate routers to store multicast groups in their forwarding tables. The introduction of centralized solutions based on Software Defined Networking (SDN) have triggered new interest on P2MP packet solutions [9]. However, forwarding state information still need to be configured and maintained in each intermediate nodes. In addition, fully centralized SDN solutions may not be suitable in the context of large packet transport networks.

More recently, IETF has proposed a novel architecture for P2MP communication called Bit Index Explicit Replication (BIER) [10]–[12]. Using the BIER protocol, the ingress router applies a BIER header containing a bit string, called *BitString*, in which each bit represents exactly one egress router in the domain. Packet forwarding is then performed by each intermediate node by processing and updating the BitString, such that all indicated egress routers correctly receive the multicast packet. BIER enables simplified and dynamic P2MP networking operations since it does not require a per-flow state in intermediate nodes and, most important, it does not require signaling protocols for explicitly building the multicast distribution trees. Moreover, BIER supports a hierarchical structure of the packet header, such that it can be adopted in large packet transport networks.

In this paper, we first summarize the BIER technology. Then, we describe and experimentally validate our BIER implementation.

II. Optical P2MP technology

Effective optical P2MP requires adequate data plane solutions at both node and interface level, supported by specifically enhanced control plane solutions.

At the node level, two main node architectures are typically deployed in today's optical networks.

The first architecture is the broadcast and select (B&S). Such architecture natively supports optical P2MP. Indeed, optical multicast is achieved by the combined utilization of light splitters and bandwidth variable wavelength selective switches (BV-WSSs). In particular, an incoming optical signal is split and internally broadcasted to all node ports. Then, the transmission along each outgoing link is either enabled or disabled by properly configuring the BV-WSSs at each outgoing node port.

The second node architecture is the switch and select (S&S). S&S is typically preferred in case of optical nodes with large nodal degree. Indeed, this architecture, although more expensive, guarantees better crosstalk performance than the B&S architecture. In S&S, due to the presence of BV-WSS at both incoming and outgoing ports, optical multicasting would not be supported. However, the recent evolutions of BV-WSS technology have introduced the capability to provide a replica of the incoming signal to more than one outgoing BV-WSS port, practically solving the issue that affected the support of light-trees by previous S&S architectures.

Interface technologies, thanks to the introduction of coherent detection strategies, enable the full programmability of multiple transmission parameters, including modulation format and forward error correction (FEC) type. In the context of optical multicast, this allows the configuration of the transparent light-tree in a flexible way, according to the requested quality of transmission. For example, if multiple formats are supported, the most spectrally efficient one guaranteeing adequate quality of transmission to all tree destinations is selected. That is, the transmission parameters are dimensioned according to the most impaired branch of the tree. Thus, other branches may experience OSNR margin larger than necessary, possibly occupying an excessive amount of spectrum resources.

Fig. 1a shows a portion of an optical network where two different trees are activated. Both trees are generated at node A. The first tree has destinations G and H while the second tree has destinations D and E. Both trees also have node B as one additional destination. Fig. 1b shows a typical approach of allocating spectrum resources according to the most impaired branch of the tree (e.g., A-F-H and A-C-E respectively). For example, both trees are configured with PM-QPSK over 50GHz of spectrum resources in all branches. The two central frequencies are assigned with a channel spacing of 50GHz. This occurs also along the shorter link A-B, in common for both trees. However, on this link, both transmissions experience larger than necessary ONSR margin.

Fig. 1c shows the proposed routing and spectrum assignment strategy where reduced amount of spectrum resources are assigned to either branches along A-B. In particular, the two central frequencies are assigned with a channel spacing lower than 50GHz (e.g., 37.5GHz) and narrow filtering is applied on both PM-QPSK signals, such that an overall amount of reduced spectrum resources is occupied along A-B (i.e., saving one slice of 12.5GHz on A-B).

Specifically designed control plane solutions are then required to effectively support optical multicasting. Centralized SDN control enables the dynamic routing and spectrum assignment of the light-tree, also accounting for the required QoT on all branches as well as the actual spectrum needs on all occupied links. In the proposed strategy, light-tree computation is based on the enhanced minimum bandwidth tree (eMBT) strategy. In particular, eMBT targets the minimization of the total occupied spectrum resources of the whole light-tree (as in [1]), trying to exploit the most efficient modulation format in terms of bandwidth on the most impaired branch of the tree. In addition, eMBT applies narrow filtering when adequate OSNR margin is available, further improving the overall network resource utilization.

III. BIER P2MP TECHNOLOGY

The topology scenario shown in Fig. 1a is here utilized to describe the behavior of the BIER P2MP solution. In this case, the nodes of Fig. 1a have packet switching capabilities and are provided with routing information identifying shortest routes in the network. Such routing table can be either generated in advance through a distributed routing protocol (e.g., OSPF) or by a centralized SDN controller with visibility on the whole routing area/domain. When a multicast flow is requested from node A to B, D and E, the SDN controller has only to configure the ingress node A to encapsulate the incoming flow packets within a BIER header. This configuration, relying e.g. on OpenFlow communication, enforces an header including a BitString. In the BitString, each bit represents a single node in the BIER domain. The BitString enforced at the ingress node has the bits representing the destinations of the requested tree set to one, while all other bits are clear. A replica of the packet is then forwarded along the outgoing links included in the shortest route towards the involved destinations (e.g., towards nodes B and C and not towards F). The outgoing BitString is updated in such a way that only the bits along each outgoing branch are set. For example, the BitString included in the packet header sent to C only has set to one the bits representing nodes D and E. Similarly, the BitString sent to B only has set to one the bit representing node B itself. Transit nodes, as C, do not have to store and maintain per-tree status information: forwarding decisions are performed on the basis of the incoming BitString only, which is updated according to the active destinations towards the outgoing links (e.g., the BitString sent to E only has set to one the bit representing node E itself).

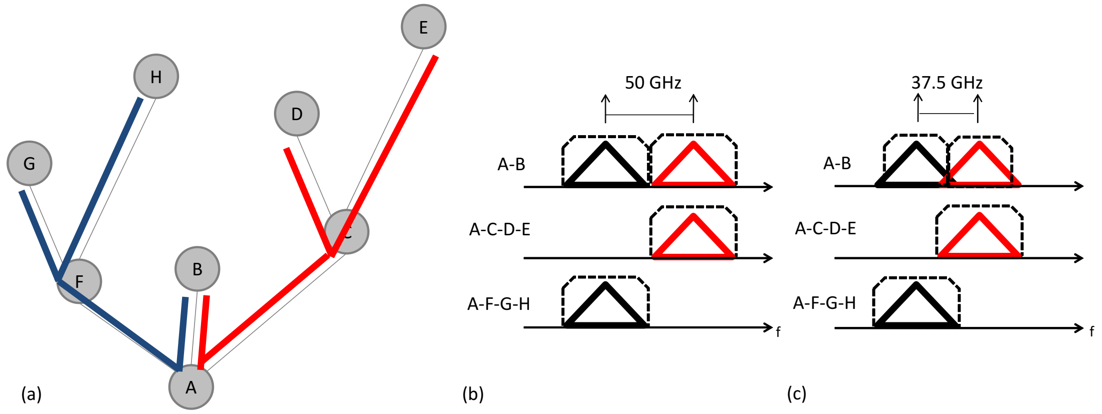


Fig. 1. (a) reference network including two light-trees; (b) spectrum assignment of the two light-trees along the reference network; (c) proposed spectrum assignment applying narrow filtering on less impaired branches (e.g., along link A-B)

At each egress node, e.g. B, D and E, the packet exits the BIER domain and the BIER header is removed. In general, a node can be both egress and transit. In this case, while a packet exits the BIER domain, other replicas are forwarded in the BIER domain with the proper BitString.

In case of changes of the multicast destination set, the controller has to reconfigure only the ingress node by enforcing a different BitString. Indeed, the transit nodes just forward the replicas along the shortest paths according to the BitString, without the need to handle signaling sessions or to maintain state information. For example, if node G subscribes to the multicast flow, the ingress A enforces a BitString with also the bit representing G set to 1. Similarly, if node D unsubscribes, the transit nodes process the header accordingly: the packet is forwarded by C only to E.

IV. Experimental Demonstration

In this section we report on the experimental demonstration of the multicasting solution based on optical P2MP and BIER.

First the light-tree is dynamically configured on an optical network testbed including a portion of the topology shown in Fig. 1a and derived from the one exploited in [1]. The testbed has been enhanced with optical nodes based on the last generation of BV-WSS technologies. That is, either B&S and S&S node architectures can be successfully utilized in this multicasting experiment. Note that either architectures can be configured, from the SDN Controller perspective, in the same way, just enforcing the configuration of the computed frequency slots on the identified nodes and ports. The testbed also includes commercial 100Gbs interfaces (Baud rate of around 30GBaud) supporting coherent detection.

The centralized SDN controller applies the proposed (eMBT) strategy targeting the minimization of the total occupied spectrum resources of the whole light-tree. Given the long route towards nodes D and E, the more robust PM-QPSK modulation format is selected, reserving a frequency slot of 50GHz. The same signal is sent also towards node B, which experiences larger than necessary OSNR margin if configured with the same frequency slot of 50GHz.

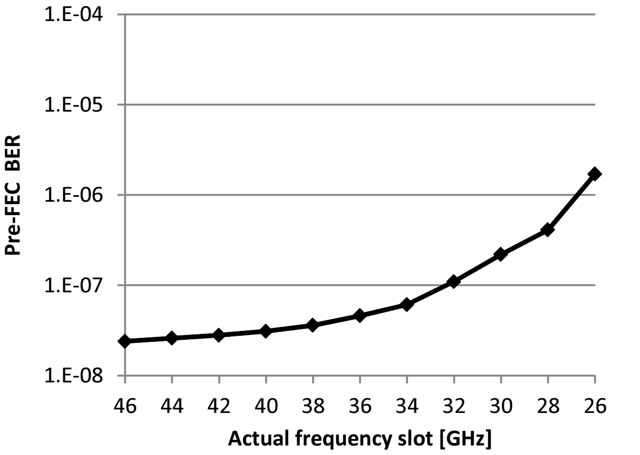
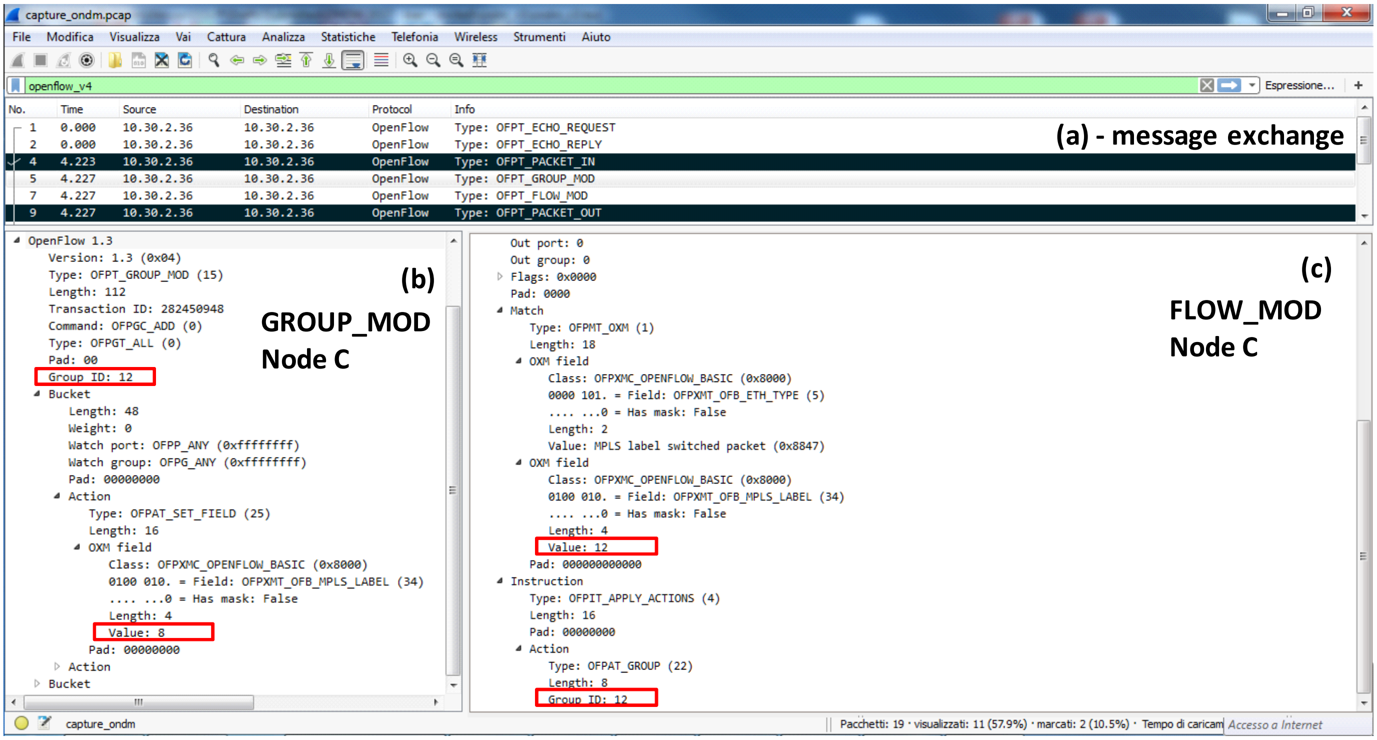


Fig. 2. Pre-FEC BER performance on the light-tree branch experiencing narrow filtering

The eMBT strategy then applies narrow filtering on A-B, further improving the overall network resource utilization.

Fig. 2 shows the performance of the commercial 100Gbs when narrow filtering is applied on a short-reach branch of the light-tree (i.e., link A-B). Results show that relevant narrow filtering can be supported without experiencing significant pre-FEC degradation, while always guaranteeing post-FEC error free conditions.

As a second experiment, BIER-capable packet nodes are introduced in the testbed, implementing a packet



```
openflow_v4
```

```
No.  Time   Source      Destination  Protocol  Info
1    0.000  10.30.2.36  10.30.2.36   OpenFlow  Type: OFPT_ECHO_REQUEST
2    0.000  10.30.2.36  10.30.2.36   OpenFlow  Type: OFPT_ECHO_REPLY
4    4.223  10.30.2.36  10.30.2.36   OpenFlow  Type: OFPT_PACKET_IN
5    4.227  10.30.2.36  10.30.2.36   OpenFlow  Type: OFPT_GROUP_MOD
7    4.227  10.30.2.36  10.30.2.36   OpenFlow  Type: OFPT_FLOW_MOD
9    4.227  10.30.2.36  10.30.2.36   OpenFlow  Type: OFPT_PACKET_OUT
```

```
OpenFlow 1.3
  Version: 1.3 (0x04)
  Type: OFPT_GROUP_MOD (15)
  Length: 112
  Transaction ID: 282450948
  Command: OFPGC_ADD (0)
  Type: OFPGT_ALL (0)
  Pad: 00
  Group ID: 12
  Bucket
    Length: 48
    Weight: 0
    Watch port: OFPP_ANY (0xffffffff)
    Watch group: OFPG_ANY (0xffffffff)
    Pad: 00000000
    Action
      Type: OFPAT_SET_FIELD (25)
      Length: 16
      OXM field
        Class: OFPXMC_OPENFLOW_BASIC (0x8000)
        0100 010. = Field: OFPXMT_OFB_MPLS_LABEL (34)
        .... ...0 = Has mask: False
        Length: 4
        Value: 8
      Pad: 00000000
    Action
  Bucket
```

```
  Out port: 0
  Out group: 0
  Flags: 0x0000
  Pad: 0000
  Match
    Type: OFPMT_OXM (1)
    Length: 18
    OXM field
      Class: OFPXMC_OPENFLOW_BASIC (0x8000)
      0000 101. = Field: OFPXMT_OFB_ETH_TYPE (5)
      .... ...0 = Has mask: False
      Length: 2
      Value: MPLS label switched packet (0x8847)
    OXM field
      Class: OFPXMC_OPENFLOW_BASIC (0x8000)
      0100 010. = Field: OFPXMT_OFB_MPLS_LABEL (34)
      .... ...0 = Has mask: False
      Length: 4
      Value: 12
    Pad: 000000000000
  Instruction
    Type: OFPIT_APPLY_ACTIONS (4)
    Length: 16
    Pad: 00000000
    Action
      Type: OFPAT_GROUP (22)
      Length: 8
      Group ID: 12
```

```
Pacchetti: 19 · visualizzati: 11 (57.9%) · marcati: 2 (10.5%) · Tempo di caricam
```

Fig. 3. Capture of local OpenFlow communications between the BIER agent and data plane at node C in Fig. 1: (a) message exchange; (b) details of OF_Group_Mod message; (c) details of OF_Flow_Mod message.

over optical multi-layer scenario.

Packet nodes are based on Open vSwitches software tool. The SDN centralized controller is based on Ryu SDN controller version 4.2 [13]. The considered BIER implementation, enhancing the work in [14], [15], also relies on a BIER agent in each node.

The agent, implemented by using a local instance of the Ryu controller, stores the list of nodes reachable with a shortest path using each of the outgoing links. Such list, stored within the so called Bit Indexed Forwarding Table (BIFT), provides good scalability performance since it scales on the (limited) number of outgoing links and not with the number of flows traversing the node. In this implementation, the BIFT table in each node is provided and updated (in case of network changes) by the global SDN controller. Note that, no signaling protocol is needed to configure the multicast trees.

With reference to the 8-node topology reported in Fig. 1a, the BitString is composed of 8 bits. The BITF tables stored at node A includes three entries. The first entry indicates that F, G and H are reachable through link A-F, the second entry that B is reachable through link A-B, and the last entry indicating the remaining nodes reachable through A-C.

In the considered BIER implementation, OpenFlow 1.3 is used between the BIER agent and the packet nodes to configure the required flow entries. When a packet addressed to a new multicast address arrives at the ingress node, the SDN controller only has to configure the specific ingress node.

In particular, a group is created with the *id* equal to the numerical value of the BitString including all recipients joining the multicast address.

When an intermediate node receives a BIER packet, the local BIER agent processes the BitString and, according to the stored BIFT, properly instructs the node to update the BitString and to perform packet replication and forwarding. Each data plane node communicates with its local agent through standard OpenFlow messages. Specifically, the agent installs a flow and a group to instruct the data plane to properly process all packets arriving with the same BitString. In this case the flow match is performed on the MPLS label representing the BitString, the action field refers to a group of type all, i.e., again the group is created with the id equal to the numerical value of the received BitString. If packets need to be replicated at the specific node, the group will include a number of buckets. This way, data plane nodes do not require one entry for each multicast flow traversing the node, because aggregation based on the destination set (i.e., the BitString value) is enforced along each branch of the multicast tree.

Fig. 3-a shows the sequence of OpenFlow messages exchanged between the BIER agent and the data plane at node C to configure the red multicast traffic flow as depicted in Fig. 1. The ingress node is A and the traffic is addressed to nodes B, D and E, that are respectively coded with the second, the third and the forth bit in the BitString. When the multicast traffic request arrives the centralized controller enforces the configuration of the ingress node A. Thus, node A forwards the traffic with two different BitStrings towards node B and node D, i.e., 0000010 and 00001100. At each transit node, the local agent then configures the forwarding table. For example, Fig. 3 shows the messages exchanged to build the forwarding table at node C: after a OF_Pck_In, a OF_Group_Mod and a OF_Flow_Mod messages are sent for data plane configuration, finally an OF_Pck_Out is used. Fig. 3-b and Fig. 3-c respectively expand the OF_Group_Mod and a OF_Flow_Mod messages. Fig. 3-b shows the group of type all (id is equal to 12, i.e., 00001100) enclosing two buckets to replicate the received packet on port 2 and 3 with proper BitString value (on port 2 the packet is forwarded with the BitString equal to 8, i.e., 00001000). Fig. 3-c shows the association of the multicast flow with the assigned group. The capture shows that the required time needed for the complete BIER configuration at node C is around 4 ms.

V. Conclusions

This paper first reviewed the most recent technologies enabling dynamic multicasting in transport networks, focusing on both optical and packet solutions. In terms of optical solutions, both data and control plane architectures and technologies for elastic optical P2MP networks are considered. Then, an innovative networking strategy is proposed, which takes advantage of the OSNR margin on less impaired branches to reduce the overall amount of spectrum resources occupied by light-trees. In terms of packet solutions, the recently introduced BIER technology is specifically considered. Then, an effective BIER implementation is presented and experimentally validated.

References

[1] N. Sambo, G. Meloni, G. Berrettini, F. Paolucci, A. Malacarne, A. Bogoni, F. Cugini, L. Potì, and P. Castoldi, “Demonstration of data and control plane for optical multicast at 100 and 200 gb/s with and without frequency conversion,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 5, no. 7, pp. 667–676, July 2013.

[2] Z. Fan, Y. Li, G. Shen, and C. K. Chan, “Distance-adaptive spectrum resource allocation using subtree scheme for all-optical multicasting in elastic optical networks,” *Journal of Lightwave Technology*, vol. PP, no. 99, pp. 1–1, 2016.

[3] Z. Zhu, X. Liu, Y. Wang, W. Lu, L. Gong, S. Yu, and N. Ansari, “Impairment- and splitting-aware cloud-ready multicast provisioning in elastic optical networks,” *IEEE/ACM Transactions on Networking*, vol. PP, no. 99, pp. 1–15, 2016.

[4] A. Cai, J. Guo, R. Lin, G. Shen, and M. Zukerman, “Multicast routing and distance-adaptive spectrum allocation in elastic optical networks with shared protection,” *Journal of Lightwave Technology*, vol. 34, no. 17, pp. 4076–4088, Sept 2016.

[5] D. D. Le, F. Zhou, and M. Molnár, “Minimizing blocking probability for the multicast routing and wavelength assignment problem in wdm networks: Exact solutions and heuristic algorithms,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 1, pp. 36–48, Jan 2015.

[6] L. Yang, L. Gong, F. Zhou, B. Cousin, M. Molnár, and Z. Zhu, “Leveraging light forest with rateless network coding to design efficient all-optical multicast schemes for elastic optical networks,” *Journal of Lightwave Technology*, vol. 33, no. 18, pp. 3945–3955, Sept 2015.

[7] L. Gifre, F. Paolucci, O. G. de Dios, L. Velasco, L. M. Contreras, F. Cugini, P. Castoldi, and V. López, “Experimental assessment of abno-driven multicast connectivity in flexgrid networks,” *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1549–1556, April 2015.

[8] R. Lin, M. Zukerman, G. Shen, and W. D. Zhong, “Design of light-tree based optical inter-datacenter networks,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 5, no. 12, pp. 1443–1455, Dec 2013.

[9] W. Gu, X. Zhang, B. Gong, and L. Wang, “A survey of multicast in software-defined networking,” in *Proc. 5th Int. Conf. Inf. Eng. Mech. Mater. (ICIMM)*, Hohhot, China, 2015.

[10] I. Wijnands *et al.*, “Multicast using bit index explicit replication,” *draft-ietf-bier-architecture-04*, July 2016.

[11] N. Kumar *et al.*, “BIER use cases,” *draft-ietf-bier-use-cases-03*, July 2016.

[12] I. Wijnands *et al.*, “Encapsulation for bit index explicit replication in MPLS networks,” *draft-ietf-bier-mpls-encapsulation-05*, July 2016.

[13] “Ryu sdn controller.” [Online]. Available: http://osrg.github.io/ryu/

[14] A. Sgambelluri, F. Paolucci, A. Giorgetti, F. Cugini, and P. Castoldi, “Experimental demonstration of segment routing,” *Journal of Lightwave Technology*, vol. 34, no. 1, pp. 205–212, Jan 2016.

[15] A. Giorgetti, A. Sgambelluri, F. Paolucci, F. Cugini, and P. Castoldi, “Segment routing for effective recovery and multi-domain traffic engineering,” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 2, pp. A223–A232, Feb. 2017.