Software-Defined Multicast Network Overlay Framework
draft-qi-bitar-intarea-sdn-multicast-overlay-01

Abstract
This document describes a framework for IP multicast based on the software defined networking paradigm. The framework enables flexible allocation of multicast responsibilities to designated control and forwarding elements, and provides for multicast path computation and programming of the forwarding plane. The objective of this framework is to enable network operators to support IP multicast in an IP/MPLS core free of Protocol-Independent-Multicast (PIM) and MPLS multicast control. The forwarding paradigm provides for multicast replication over unicast and multicast overlays at designated edges, and for native IP and MPLS multicast replication and forwarding in the core.

Status of this Memo
This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.
Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts. Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."


This Internet-Draft will expire on September 13, 2017.

Copyright Notice
Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.
This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Conventions used in this document
The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

1. Introduction .................................................. 4
2. Terminology .................................................. 4
3. Acronyms ........................................................ 5
4. Problem Statement .............................................. 6

5. Multicast SDN Framework Benefits ............................. 8
6. Multicast SDN Framework Guiding Principles and Requirements .10
7. Multicast SDN Framework Architectural Components ............12
8. SDN Multicast Framework ......................................15
8.1. SDN Multicast Operational Models ..........................16
  8.1.1. Full Multicast SDN Model Operational Overview ......16
  8.1.2. Hybrid Distributed-SDN Multicast Model Operational
         Overview ..................................................18
  8.1.3. Cut-Through Operational Mode ..........................18
  8.1.4. Summary of Control an Data Planes Applicable to the
         SDN Multicast Models and Gaps ............................19
8.2. SDN Multicast Overlay Control Plane Functionalities  ...21
  8.2.1. Unicast Topology Discovery ............................21
  8.2.2. Multicast Topology Discovery ..........................22
  8.2.3. Resource Utilization Monitoring ........................22
  8.2.4. Multicast Group Membership Discovery ...................22
  8.2.5. Multicast Sender Discovery .............................23
  8.2.6. Multicast distribution tree creation and
         maintenance ................................................23
  8.2.7. Multicast forwarding programmability ..................24
  8.2.8. Entitlement admission control .........................25
  8.2.9. Bandwidth Admission control ...........................25
8.3. Multicast Functional Components Implementation ..........25
8.4. SDN Multicast Management Interfaces .......................26
  8.4.1. North-Bound Interface ................................26
  8.4.2. Multicast SDN Functional Component Facing Interfaces
         .........................................................26
  8.4.3. Traditional Underlay Network Element Interfaces ....26
  8.4.4. Information Model and Data Model of Management
         Interfaces ................................................27
9. SDN Control Plane ...............................................27
10. Multi-Party Multicast SDN Overlay Interaction ...............27
11. Security Considerations .......................................27
12. IANA Considerations ..........................................28
13. References ..................................................28
  13.1. Normative References ....................................28
  13.2. Informative References ..................................28
14.0 Acknowledgments .............................................30
Authors Addresses ...............................................30
1. Introduction

Operators that enable multicast services in their network have traditionally relied on host-based multicast protocols such as Internet Group Management Protocol (IGMP), and core-based multicast protocols such as Protocol Independent Multicast (PIM) with its variants, and more recently MPLS multicast protocols. Some operators have encountered issues with these protocols that have ensued from protocol activities, and perhaps implementation behavior under some scale, that overloaded the control plane of the routing nodes, sometimes effecting scalability, and often leading to network instability. There are four sources for this activity: (1) high dynamicity of group membership, resulting in a high rate of join/leave, (2) refresh of multicast soft state, (3) faulty implementation, and (4) malicious attacks.

This document describes a framework for IP multicast based on the software-defined networking (SDN) paradigm. The objective of this framework is to allow an operator to flexibly design IP multicast networks to fit its needs, and minimize network risks. Specifically, it enables an operator to provide for multicast services in its network without PIM or MPLS multicast signaling protocols (i.e., point to multipoint Resource Reservation protocol - Traffic Engineering (RSVP-TE), or Multicast Label Distribution Protocol (MLDP)). The elimination of these protocols from network elements alleviates a good amount of control plane activity often consumed by these protocols to establish and maintain multicast soft state, minimizing network stability risks and providing for better scale.

The SDN control plane described in this framework, provides for (1) processing join and leave requests from receivers (e.g., IGMP, PIM) or applications that request the addition or pruning of receivers, (2) multicast entitlement and bandwidth admission control, (3) the computation of a multicast path if necessary, and (4) the establishment of the necessary forwarding states in the nodes on the computed forwarding path.

2. Terminology

This document adopts the following SDN terminology from RFC7426 (SDN: Layers and Architecture Terminology):

- Software-Defined Networking (SDN): A programmable network approach that supports the separation of control and forwarding planes via standardized interfaces.
- **Forwarding Plane (FP):** The collection of resources across all network devices responsible for forwarding traffic.

- **Control Plane (CP):** The collection of functions responsible for controlling one or more network devices. CP instructs network devices with respect to how to process and forward packets. The control plane interacts primarily with the forwarding plane and, to a lesser extent, with the operational plane.

Furthermore, this document adopts the following terminology from [RFC 7365](https://tools.ietf.org/html/rfc7365), Framework for Data Center (DC) Network Virtualization:

**Tenant:** The customer using a virtual network and any associated resources (e.g., compute, storage, and network). A tenant could be an enterprise or a department/organization within an enterprise.

**Tenant System:** A physical or virtual system that can play the role of a host or a forwarding element such as a router, switch, firewall, etc. It belongs to a single tenant and connects to one or more virtual networks (VNs) of that tenant.

### 3. Acronyms

- ASM: Any-Source Multicast
- MSN: Multicast Service Node
- MSE: Multicast Service Edge
- MBG: Multicast Border Gateway
- SSM: Source-Specific Multicast
- TES: Tenant End System
- VM: Virtual Machine

### 4. Problem Statement

IP multicast in core IP networks has prevalingly relied on PIM-ASM to establish a multicast tree between the first hop router connected to a sender and the edge routers connected to receivers, and relied on multicast source discovery protocols to discover multicast sources. PIM is a soft state protocol that relies on periodic join refreshes to keep the multicast
states alive and its mode of forwarding requires data-driven events. While PIM-SSM removed some of the above issues for PIM-ASM, it is IGMPv3 dependent and requires out-of-band source discovery mechanism. MPLS multicast signaling protocols, namely point to multipoint RSVP-TE, imposes similar control load on MPLS routers. On access links, edge routers process IGMP messages from attached receivers or CPEs, or PIM messages from attached CPEs. These messages signal to edge routers the multicast membership state of the attached devices. IGMP is also soft state, requiring periodic refreshes of join/prune messages to refresh multicast membership state, or queries to clean up stale state.

Multicast group activity (joining/leaving a multicast group) could be very dynamic, resulting in a high control message rate that increases with the number of multicast receivers, and is highly dependent on the behavior of these receivers in leaving/joining multicast groups.

All this control plane activity is performed on (all) nodes in the network to support multicast, risking network control plane scale and stability.

Routers not only run multicast control protocols, but also perform data plane multicast, holding multicast forwarding state and performing packet replication and forwarding. Multicast forwarding state is dependent on the number of multicast groups ((*,G) or (S,G)) and number of outgoing interfaces per multicast group. As receivers join or leave multicast groups, these multicast forwarding entries need to be updated, further adding to a router control plane load. A large number of (*,G)/(S, G) multicast states and control information push against the resource limit of network routing hardware (e.g., TCAM, CPU). On an edge node where the number of outgoing interfaces could be very large, this activity is further exacerbated. In addition, data replication shares forwarding capacity (processing cycles and bandwidth) with unicast, often mandating the need to control the capacity allocated for multicast and enforcing it.

Native multicast protocol procedures do not address the protection of a router control plane and data plane resources from the negative impacts that multicast enablement may have on these resources. Such negative impacts could be the result of malicious activities, misconfigurations, or simply lack of control on legitimate activities. For instance, a receiver may maliciously join a large number of multicast groups and
create shared bandwidth bottlenecks if allowed to successfully join these groups. Such a receiver may also generate IGMP reports at a very high frequency, creating an attack on the access router control plane if not properly controlled. A host may also masquerade itself as a source for many multicast groups, resulting in the creation of a large number of forwarding state on the access router and other core nodes. Some router vendors have implemented mechanisms to protect against such attacks (malicious or not) on the router control and data plane, but often these mechanisms are not implemented, partially implemented or not implemented with a consistent behavior let alone configuration as they are vendor proprietary. Lack of consistency poses challenges to network operators when multicast is to be enabled in a multi-vendor environment.

Without some type of entitlement that could be defined and uniformly enforced across all edge routers, a rogue multicast sender that knows active multicast groups can send traffic to these groups. The same goes for multicast receivers who may snoop on multicast traffic that they are not entitled to by simply joining the corresponding multicast group(s). Thus, there is a need to define and enforce entitlement for receivers to join multicast groups and for senders to be able to send traffic to multicast groups to protect against the negative impacts that sender and receiver misconfigurations or malicious activities could inflict.

There is need to (1) protect against multicast control plane message attacks that may or may not result in a receiver successfully joining a multicast group, (2) be able to define and enforce entitlement for both receivers of and senders to multicast groups, and (3) to protect against the oversubscription of resources beyond set engineering targets by performing resource (bandwidth and other resources) admission control.

Multicast entitlement, admission control and control DDoS protection are hard to implement in existing networks when relying on router vendor implementations as they do not holistically or uniformly address all needs, if any.

Finally, there is no multicast management plane or north-bound interface to harness multicast control state and telemetry information that enable monitoring the state of the network and enable controlling the usage of network resources.
5. Multicast SDN Framework Benefits

The multicast SDN framework is composed of an overlay control plane that computes multicast distribution paths between senders and receivers per multicast group/channel, and establishes the forwarding path on forwarding elements, enabling the distribution of multicast traffic between these senders and receivers, and provides for increased functionality and flexibility. The multicast overlay control plane on centralized multicast SDN controller(s) alleviates the multicast control plane activities performed by routing nodes in traditional networks. The controller could be an existing controller with multicast capability or a separate multicast controller.

The multicast SDN framework provides the following benefits:

- By alleviating multicast control protocols and associated processing and states off the distributed traditional routing nodes in a network, the routing control plane of these nodes will have an order of magnitude reduction in complexity and a significant reduction in processing load, improving network stability and scalability and faster time to recovery from failures.

- Much fine grained policy and advanced service enhancements such as multicast control policy (access control, admission control, entitlements and bandwidth), can be implemented directly and consistently in the SDN multicast control plane. This functionality can be and has been implemented in traditional routers but at the expense of additional processing requirements on the control plane and the burden of ensuring consistent implementation across various vendors.

- Much fine grained multicast path selection based on various attributes (e.g., bandwidth, diversity, latency) and programming, and various degree of replication, alleviating the current congruency between multicast and unicast topologies and providing flexibility in carving multicast paths for multicast groups and receivers.

- Multicasting across Campus, Datacenter, Wide Area Network, and multiple administrative boundaries under the same enterprise control can be implemented with a unified and
integrated Management Plane and Control Plane.

- Multicasting across administrative boundaries under the control of different enterprises/operators, can be implemented leveraging SDN mechanisms and/or traditional distributed control mechanisms, depending on the capabilities of the neighboring enterprise/operator domains.

- Control and/or management plane applications can compute, modify, steer, and program forwarding plane multicast paths, and query and harness state information from the routing nodes performing multicast forwarding using north bound interfaces from these nodes.

- Flexible design of the multicast services in the operator network to address operator design principles and the capabilities of the forwarding elements.

6. Multicast SDN Framework Guiding Principles and Requirements

Following are the requirements that MUST be satisfied by the Multicast SDN framework:

- No constraint on the network topology. The SDN framework will need to be aware of the unicast and multicast network topology when it needs to compute unicast and multicast paths that satisfy certain constrains and/or to optimize for network usage resources on these paths.

- No constraint on the unicast routing protocol selection in the underlay network. The network may consist of multiple domains with different unicast routing protocols.

- Be agnostic to other services in the network (e.g., BGMP/MPLS unicast and multicast Layer2 and Layer3 services, other types of VPN services). That is, it MUST allow new multicast services to be enabled based on this framework in
the presence of other services.

- Support IPv4, IPv6, and MPLS unicast underlay data plane, providing for edge replication over IPv4, IPv6, or MPLS unicast overlays.

- Support IPv4, IPv6, and MPLS multicast data planes, providing for replication over IPv4, IPv6, and MPLS multicast overlays, and for the computation and programmability of the overlay data paths on the underlay as needed.

- Support a BIER [BIER-ARCH] underlay data plane to support multicast replication of multicast packets over BIER type of multicast overlays.

- Support a segment-routing (SR) [SR-ARCH] underlay data plane to support edge replication of multicast packets over unicast SR overlays with both IP and MPLS data planes.

- Support stitching of multicast packets across domains with the different control and data planes stated in the earlier requirements.

- Integrate with network multicast (control and data plane) where operationally feasible resulting in optimum bandwidth utilization in data plane.

- Enable decoupling of multicast topology from unicast topology, providing the ability to enable multicast selectively on nodes, and/or carving multicast paths.

- Enable select edge data replication nodes, specifically when the first hop router connected to the Multicast sender is not best suited for data replication.

- Support existing multicast applications without requiring any modification to these applications, e.g., applications that use IGMP for multicast membership.

- Support the tunable aggregation of multicast groups on multicast tunnels (where tunnels may be stateful tunnels, or simply encapsulation headers)

- Enabling operators to traded-off multicast data plane bandwidth efficiency with multicast data plane and control plane state
- Support multicast admission control on unicast tunnels/paths, multicast tunnels, and path (re-) computation on the underlay.

- Support multicast admission control based on entitlement of the receiver to a multicast group or channel

- Support programmability of multicast traffic policers and filter policies

7. Multicast SDN Framework Architectural Components

Multicast Service Edge (MSE):
An MSE connects to tenant end systems (TESs) (hosts) on the access side and to an underlay IP network or IP/MPLS network on the network side. In the forwarding plane, when an MSE receives a user IP multicast packet from a TES on the access side, it replicates the packet on local ports with TES receivers, and encapsulates the IP multicast packet in a unicast packet to a Multicast Service Node (MSN) or a Multicast Border Gateway (MBG) to reach remote multicast receivers. It also receives IP multicast packets encapsulated in unicast packets on the network side, decapsulates the multicast packets and replicates them on local ports to corresponding TES receivers. In the control plane, an MSE connects to a multicast controller that programs the MSE with multicast forwarding entries triggered by management entities or dynamic control protocols. An MSE may receive IGMP or Multicast Listener Discovery protocol (MLD) report/leave packets on the access side. The MSE acts as an IGMP/MLD proxy and sends messages to the multicast SDN controller to inform the controller of local multicast receivers for a multicast group or channel. It may also be statically programmed with multicast entries. An MSE supporting multi-tenancy provides for multicast control and forwarding in a tenant specific virtual routing and forwarding context. It can be implemented in hardware or software. For example, it can be part of a virtual router in a virtualized environment, a kernel or user-plane routing function in physical servers, a physical router, or appliance which performs multicast service functions.
There are multiple ways an MSE can intercept multicast control messages and multicast data packets from TESs when the TESs are receivers and senders, respectively:
- MSE is a multicast router on local VLAN
- MSE is a designated multicast router for a LAN router. A LAN router (e.g., a Top of rack (ToR) router in a DC) forwards all multicast packets it receives to a designated MSE
- MSE is part of virtual switch/router on a virtualized host
- MSE is a kernel function and intercepts data frames
- MSE resides in a NIC of the TES and intercepts all frames in and out of the NIC.
- MSE is a customer edge (CE) router for a local site.

Multicast Tunnel TES (MT-TES):
An MT-TES is a tenant end system (host), virtual or physical, that may send or receive multicast packets encapsulated in unicast IP/UDP packets and delivers them to local applications based on the UDP port number in the encapsulating header or multicast destination address.

Multicast Service Node (MSN):
An MSN is the network entity that receives and replicates IP/Multicast packets from/to others MSNs, MBGs, MSEs and MT-TESs in the data plane. An MSN is often designated to serve an MSE with multicast receivers and senders, and/or MT-TESs. There could be more than one MSN serving the same MSE or the same MT-TES. In that case, the selection of the MSN can be based on whether the MSE or MT-TES is sending or receiving multicast packets to/from the network, and/or multicast group address or multicast channel based on the tuple multicast source (S) and multicast group address (G). An MSN may tunnel packets over the underlay network to other MSNs and MBGs using unicast tunnels/unicast header encapsulation, or may forward multicast packets natively on the underlay. An MSN always forwards multicast traffic encapsulated in unicast packets to MSEs and MT-TESs as programmed. On the network side, an MSN may participate in native IP multicast control and forwarding if so configured. The multicast controller MAY also program underlay multicast forwarding entries and the forwarding unicast entries that correspond to the unicast forwarding information in the encapsulating header of the packet carrying multicast packets on the MSN for policy-based routing and forwarding purposes. Alternatively, the unicast packets, encapsulating the multicast
packets, can be forwarded based on underlay dynamic unicast routing, the default behavior. Unlike MSE, MSN does not connect to access networks or TESs. MSN can optimize replication based on latency, bandwidth, QoS, number of multicast states desired or even more specific multicast application requirements.

MSN can be implemented in hardware or software. For example, it can be a physical router, or a software appliance which performs multicast service functions. Each MSN has a forwarding plane and a control plane, albeit new control plane functionality and traditional control plane functionality implemented today are relegated to an SDN multicast controller.

Multicast SDN Domain (MSD):
A multicast SDN domain encompasses forwarding nodes (MSEs, MSNs, and MBGs) and a cluster of one or more multicast SDN controllers that control the programming of tenant multicast forwarding entries on these nodes in addition to underlay core nodes that provide for IP connectivity among these nodes.

Multicast Border Gateway (MBG):
An MBG is a network forwarding node that sits at the border of a multicast SDN domain and connects to other networks in different domains, which could be either multicast SDN domains or traditional networks with multicast capabilities in the same or different administrative domains. An MBG forwards multicast packets from one domain to another across the boundary and forwards packets to MSNs within its domain. An MBG could also implement MSN functionality, serving MSEs and MT-TESs. As an example, in cloud networking whereby the cloud is composed of multiple data centers (DCs), an MBG may perform the function of a DC multicast gateway to the WAN to provide for the forwarding of multicast traffic originated by one TES in one DC to TESs located in other DCs. In this case, the MBG may use BGP/MPLS Multicast VPN (BGP-MVPN) [RFC 6513] across the WAN. Inside a domain, an MBG may be enabled with traditional IP multicast control and forwarding on the network underlay links to perform underlay multicast forwarding. Alternatively, underlay multicast forwarding entries reachable within the domain may be programmed by the corresponding multicast SDN control plane. Overlay traffic may be tunneled to/from the MBG using unicast tunnels/unicast-header encapsulation or multicast tunnels/multicast header encapsulation.

An MBG can be implemented in hardware or software. For example, it can be a physical router, or a software appliance which performs multicast service functions albeit new control plane
functionality and traditional control plane functional
implemented today are relegated to an SDN multicast controller.

Core Node (CN):
A CN is a routing node which provides the underlay
interconnectivity among the MSEs, MSNs and MBGs. It provides for
unicast IP, MPLS, and SR forwarding and may provide multicast IP,
MPLS and/or BIER forwarding.

8. SDN Multicast Framework
Figure 1 depicts the SDN multicast framework, specifically the
SDN management plane, control plane and data plane functional
elements and their interactions. In the full SDN model, control
of any data plane element for multicast is under the control of
the Multicast SDN control plane. Variants of this model that
leverage existing solutions (e.g., BGP-MVPN) are described in
this document to fit different deployment models.

Qi-Bitar Expires January 2018 [ Page 15]
8.1. SDN Multicast Operational Models
8.1.1. Full Multicast SDN Model Operational Overview

This section provides an overview of the full SDN multicast model and describes the transactions and information that needs to be exchanged across the different interfaces, including the interfaces between the multicast SDN controller and the various multicast nodes (MSE, MSN, CN, MBG). Throughout this section it should be assumed that all multicast operations on any of the edge nodes are done in a tenant context. In the case of multi-tenancy, MSEs, MSNs and MBGs must be able to perform destination lookup and forwarding in a virtual routing and forwarding (VRF) context. In addition, they must be able to encapsulate the original multicast packet with a unicast (tunnel) header and additional header information that carries a tenant identifier that can be used to identify the forwarding context at the receiving end.

Senders and receivers for a multicast group may be static or dynamic. For each tenant and multicast group, the SDN multicast controller is informed of the sender(s) via the controller northbound management interface (MApp-C) or the MSE connected to the sender(s) via the MSE-C interface. The SDN controller, in the generic case, will select an MSN that will proxy the MSE connected to the sender, and programs the MSE to send the multicast group traffic it receives from the sender(s) to the selected MSN. The SDN controller may select more than one such MSN. The multicast group packets will be delivered to tenant end systems served by other MSEs and MSNs from these selected MSNs. For static TES multicast receivers on a particular group or channel, the SDN controller knows which TESs are members of a multicast group/channel via the MApp-C management interface, and learns the MSE serving the TES via the MApp-C management interface or via dynamic routing. Then for each TES on that multicast group, The SDN controller selects the MSN to serve the corresponding MSE. This selection could be based on certain
criteria that optimize for bandwidth utilization on the path between the MSN and the MSE, and/or resource utilization on the MSN among others. Once an MSN selection is done to serve the MSE, the SDN multicast controller needs to ensure that the selected MSN receives the multicast group/channel traffic. The SDN controller then selects the remote MSN from which it wants to receive the multicast group packets. This latter selection could be based on bandwidth utilization on the path between the two MSNs in question, remote MSN resources, and/or the transport method (multicast or unicast tunnel/tunneling encapsulation). In the case of unicast transport, the remote MSN proxying the sender MSE is programmed via the MSN-C interface with a forwarding entry that results in encapsulating the multicast group packets with a unicast SR, IP or MPLS tunnel or encapsulation header that delivers the packets to the MSN serving the receiver MSE. The MSN serving the receiver MSE is also programmed with an entry that results in forwarding the received multicast packet over another tunnel encapsulation to the destination MSE. Optionally, the MSN serving the receiver MSE may be programmed with the tunnel/encapsulation header information over which the multicast group packets will be delivered (i.e., the sender MSN IP address or MPLS tunnel). Unicast forwarding is based on dynamic routing information, static (policy-based) routing or MPLS signaling in the case of RSVP-TE. In the case of multicast transport, the SDN controller may choose to add a new multicast tree branch to an existing multicast tree or re-compute a new one rooted at the sender. In the case of IP multicast transport, and in the absence of PIM in the core, the SDN controller programs the CNs on the path with the appropriate multicast forwarding entries. The MSE could also optionally be programmed with the tunnel/header encapsulation information over which packets are to be received.

For dynamic receivers, a TES may initiate an IGMP report or MLD join message for a multicast group. That message should be received by the MSE. The MSE may act as an IGMP proxy/MLD proxy in which case the MSE, upon receiving the first join message from a connected TES for a multicast group/channel, sends a join message for that multicast group/channel with tenant context to the SDN controller over the MSE-C interface. In case there is requirement for entitlement or bandwidth admission control, the MSE always sends a message to the SDN controller over the MSE-C interface, requesting admission control for the requesting TES for the particular group/channel, including the tenant context. If the request is admitted, the MSE is informed of the admission decision and the transport path setup proceeds as described for the static
receiver case. In the proxy case, when the last receiver leaves a multicast group on the MSE either via an explicit leave, failure to respond to a query or failing, the MSE sends a message to the SDN controller that it no longer has receivers for that group/channel via the MSE-C interface, and the controller may remove the MSN as being a receiver for that group/channel and take appropriate action removing associated forwarding state on other nodes (MSNs, CNs).

8.1.2. Hybrid Distributed-SDN Multicast Model Operational Overview
In this model, MSNs and MBGs perform the role of an BGP-MVPN PE with a distributed BGP-MPLS control plane for exchanging receiver information in a tenant network (MVPN) with other MSNs and MBGs. The programmability of the MSNs with the receiver information is done via the SDN controller to create the relevant BGP policies and enable tunneling of multicast traffic to served MSEs.

8.1.3. Cut-Through Operational Mode
In the cut-through model, a sender MSE can send multicast packets directly to remote receiver MSE(s), bypassing MSNs and MBGs. That is, MSNs do not proxy for downstream MSEs. In this case, the multicast SDN controller programs a sender MSE via the MSE-C interface with a forwarding entry that results in encapsulating the multicast group packets received at the MSE from the access network with a unicast IP header and a tenant identifier that delivers these packets to remote receiver MSE(s). The receiver MSE is also programmed with an entry that results in forwarding the received multicast packet to downstream multicast receivers. Similarly, Sender MSE can send multicast packets directly to receiver MT-TESs with the same operational procedures.

It should be noted that upon operator configuration, a multicast deployment MAY include one or more of the operational models described in the SDN multicast framework section.

8.1.4. Summary of Control an Data Planes Applicable to the SDN Multicast Models and Gaps

Table 1: Underlay and overlay control planes applicable
to the various SDN multicast models, and gaps (new).

<table>
<thead>
<tr>
<th>Model</th>
<th>Multicast Underlay</th>
<th>Multicast Overlay</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control Plane</td>
<td></td>
</tr>
<tr>
<td>Full SDN</td>
<td>MSE LAN access options:</td>
<td>MSE: MSE-MSDNC (new)</td>
</tr>
<tr>
<td></td>
<td>IGMP, MLD, PIM, static</td>
<td>MSN: MSN-MSDNC (new)</td>
</tr>
<tr>
<td></td>
<td>MSE/MSN/MBG/CN-MSDNC (new)</td>
<td></td>
</tr>
<tr>
<td>Hybrid</td>
<td>MSE LAN access options:</td>
<td>MSE: MSE-MSDNC (new)</td>
</tr>
<tr>
<td></td>
<td>IGMP, MLD, PIM, static</td>
<td>MSN: MSN-MSDNC (new)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSN-MSE: None</td>
<td>MSN &amp; MBG options:</td>
</tr>
<tr>
<td></td>
<td>Core (MSN-CN-MBG) options:</td>
<td>MBGP (membership)</td>
</tr>
<tr>
<td></td>
<td>MSDN Control</td>
<td>LISP[LISP][LISP-mcst]</td>
</tr>
<tr>
<td></td>
<td>IGP (for BIER data plane)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PIM (core multicast IP)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MPLS multicast (MPLS core)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IGP (SR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MSDNC-MSDNC (new)</td>
<td></td>
</tr>
<tr>
<td>Cut-Through</td>
<td>MSE LAN access options:</td>
<td>MSE options:</td>
</tr>
<tr>
<td></td>
<td>IGMP, MLD, PIM</td>
<td>MSE-MSDNC (new)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MBGP (membership)</td>
</tr>
<tr>
<td></td>
<td>MSE options:</td>
<td>LISP</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Core (MSN-CN-MBG) options:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IGP (BIER data plane)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PIM (core multicast IP)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MPLS multicast (MPLS core)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multicast SDN Controller</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(MSDNC)-MSDNC (new)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Underlay Control Plane Protocol Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicast: BGP OSP/ISIS (SR support) RSVP-TE/LDP</td>
</tr>
<tr>
<td>Multicast PIM IGMP/MLD pt-mpt RSVP-TE/mLDP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overlay Control Plane Protocol Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBP LISP NVo3 (TBD) Multicast SDN (TBD)</td>
</tr>
</tbody>
</table>

Figure 2: Summary of underlay and overlay control protocol options.

Qi-Bitar Expires January 2018 [ Page 19]
Table 2: Underlay and overlay data plane applicable to the various SDN multicast models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Multicast Underlay Data Plane</th>
<th>Multicast Overlay Data Plane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full SDN and Hybrid</td>
<td>MSE LAN access: IP multicast</td>
<td>MSE LAN access: Not applicable</td>
</tr>
<tr>
<td>MSN,MBG and CN options:</td>
<td>-IP multicast</td>
<td>MSE-MSN core options: -IP multicast in IP VXLAN or IP/GRE/VPN unicast</td>
</tr>
<tr>
<td></td>
<td>-BIER</td>
<td>-IP multicast in SR VPN</td>
</tr>
<tr>
<td></td>
<td>-IP unicast</td>
<td>MPLS encap</td>
</tr>
<tr>
<td></td>
<td>-SR</td>
<td>-IP multicast in LISP encap</td>
</tr>
<tr>
<td></td>
<td>-MPLS unicast</td>
<td>MSN/MBG options:: -IP multicast in IP VXLAN or IP/GRE/VPN o LISP unicast</td>
</tr>
<tr>
<td></td>
<td>-MPLS multicast</td>
<td>-IP multicast in BIER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-IP multicast in MPLS VPN multicast</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-IP multicast in SR VPN MPLS</td>
</tr>
</tbody>
</table>

Cut-Through
<table>
<thead>
<tr>
<th>Model</th>
<th>Multicast Underlay Data Plane</th>
<th>Multicast Overlay Data Plane</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE LAN access: IP multicast</td>
<td>MSE LAN access: Not applicable</td>
<td></td>
</tr>
<tr>
<td>MSN,MBG and CN options:</td>
<td>-IP multicast</td>
<td>Core(MSE-MSN) options: -IP multicast in IP VXLAN or IP/GRE/VPN unicast</td>
</tr>
<tr>
<td></td>
<td>-BIER</td>
<td>-IP multicast in SR VPN</td>
</tr>
<tr>
<td></td>
<td>-IP unicast</td>
<td>MPLS encap</td>
</tr>
<tr>
<td></td>
<td>-SR</td>
<td>-IP multicast in LISP uncap</td>
</tr>
<tr>
<td></td>
<td>-MPLS unicast</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-MPLS multicast</td>
<td></td>
</tr>
</tbody>
</table>

Underlay Data Plane Protocol Options

<table>
<thead>
<tr>
<th>IPv4 unicast/multicast</th>
<th>IPv6 unicast/multicast</th>
<th>MPLS</th>
<th>BIER</th>
<th>SR</th>
</tr>
</thead>
</table>

Shim header (carries tenant identifier):

<table>
<thead>
<tr>
<th>VXLAN</th>
<th>LISP</th>
<th>GRE/MPLS</th>
<th>MPLS</th>
</tr>
</thead>
</table>

Overlay Data Plane Protocols

<table>
<thead>
<tr>
<th>IPv4 multicast</th>
<th>IPv6 multicast</th>
</tr>
</thead>
</table>

Figure 3: Summary of underlay and overlay data plane protocol options.
8.2. SDN Multicast Overlay Control Plane Functionalities

Multicast control is traditionally distributed on network routing elements, and mainly responsible for building and maintaining multicast distribution trees or transport, and configuring the multicast forwarding plane. Control Plane entities may exchange multicast sender or receiver-membership information with each other using network protocols such as MP-BGP or PIM. The full multicast SDN model described in this document alleviates the distributed control plane functions from network elements but preserves the functionalities needed to discover multicast membership and source information, build multicast distribution trees and/or multicast transport tunnels.

The SDN Multicast Control plane in the full multicast SDN model Must include the following functionalities:

- Unicast topology discovery interconnecting MSEs, MBGs, MSNs and CNs
- Multicast topology discovery including MSEs, MSNs, MBGs, and CNs enabled for multicast
- Resource utilization monitoring on links and nodes
- Multicast group membership discovery and maintenance
- Multicast source discovery and maintenance
- Multicast distribution tree creation and maintenance
- Multicast path selection, instantiation, maintenance and failover
- Multicast forwarding programmability

The SDN Multicast control plane SHOULD include the following functionalities

- Entitlement admission control
- Bandwidth Admission control

8.2.1. Unicast Topology Discovery

The unicast Topology is needed to discover reachability among MSEs, MSNs and MBGs for MSN selection or MBG selection. In addition, the unicast topology will be needed if traffic is to be carried using any unicast tunneling technology (IP, MPLS, SR with IP or MPLS data planes), or pt-to-pt RSVP-TE. In this
latter case the TE tunnels may be computed by the underlying router. Alternatively, the TE tunnels paths MAY be computed by the SDN controller. The same may apply for Traffic engineered SR paths. The SDN control plane SHOULD always maintain a database of these TE tunnels/paths and associated capacity, irrespective how they are computed, enabling the SDN controller to selectively bind multicast traffic to TE paths. The SDN multicast control plane and network elements that will be used to learn the unicast topology SHOULD support BGP-LS [RFC 7752] or YANG topology data models [Yang-Topology-Model]. The SDN control plane SHOULD learn the unicast topology via BGP-LS or YANG from the network nodes. The SDN controller MUST also learn TES reachability via MSEs using same mechanisms.

8.2.2. Multicast Topology Discovery
The SDN multicast control plane MUST have knowledge of all the MSNs, MBGs and MSEs and their multicast resources. In case the core network is also enabled for multicast and the SDN control plane is to compute multicast trees, the SDN control plane MUST know which nodes are enabled for multicast forwarding and enabled for PIM. The Multicast topology is used by the multicast SDN control plane to select MSNs and MBGs and to compute multicast tree paths. The multicast tree could be based on native IP multicast or pt-to-mpt RSVP-TE. The SDN multicast control plane MUST maintain the computed paths and the receivers and senders mapped to them.

8.2.3. Resource Utilization Monitoring
The SDN control plane MUST maintain knowledge of the available resources on the nodes that may be impacted resource-wise by being on multicast paths selected by the SDN controller. For instance, the SDN controller MUST maintain the available resources for multicast forwarding entries on MSEs, MSNs and MBGs. In addition, if bandwidth and link utilization is to be considered as part of the path selection, the SDN controller MUST maintain knowledge of the utilization on network links.

8.2.4. Multicast Group Membership Discovery
The SDN Controller MUST learns via the MApp-C management interfaces or the MSE-C/MTTES-C interfaces the receivers and senders for each multicast group or channel.
When the list of multicast receivers and senders for each multicast group in a tenant network is known in advance and when multicast session establishment (set of senders and receivers for a multicast group) can be set up ahead of time of actual multicast data delivery, applications can use the multicast SDN controller northbound interface (MApp-C) to populate the list of receivers and senders for the multicast group in the SDN controller. The SDN controller may also have access to a database, or be programmed with information via MApp-C, that associates the receivers and senders with MSEs.

8.2.5. Multicast Sender Discovery
The multicast SDN controller MUST learn the set of senders for each multicast group. This information may be learned via the MSE-C/MTTES-C interfaces from MSEs or TESs, or via the MApp-C interface from a management application. The SDN controller could obtain the location of the sender (which MSE a sender is connected to) from routing information, or via MApp-C and MSE-C interfaces.

8.2.6. Multicast Distribution Tree Creation and Maintenance
The SDN controller MUST be able to compute a multicast distribution tree. There could be several segments for a multicast distribution tree depending on the capabilities of the network elements, and/or the operator design. Following are the distribution tree types:

1- Distribution Tree Type 1: The distribution tree is rooted at one MSN that receives a multicast group source traffic from an MSE. Then, that MSN encapsulates the multicast traffic with a unicast tunneling header to other MSNs connected to MSEs with receiver TESs for that group, and to other MBGs to reach TES receivers in other domains.

2- Distribution Tree Type 2: The distribution Tree is rooted at one MSN that receives a multicast group source traffic from an MSE. Then, the MSN encapsulates the traffic with a multicast tunneling header or in a multicast tunnel with leaves being other MSNs and MBGs to reach receivers. When an MSN receives the tenant multicast traffic on a multicast tunnel/encapsulated in a multicast tunneling header, it decapsulates the tenant multicast packets and unicast tunnel them to the MSEs with receiver TESs for that group.
In computing the multicast tree, the SDN controller SHOULD take into account constraints on bandwidth and other QoS parameters such as latency, jitter and packet loss in addition to the capabilities of the forwarding plane elements (e.g., native multicast replication support, multicast replication over unicast tunnels, number of multicast entries, bandwidth, etc.) and policies (e.g., which MSN is to serve which MSE, inter-SDN domain policies etc.). A multicast distribution tree is recomputed every time that there is a new edge node (MSE, MSN or MBG) to be added to the multicast tree, or there is an edge node that needs to be pruned out of the multicast tree), adding or removing multicast branches to/from the tree.

An SDN multicast controller can map tenant overlay multicast group addresses 1:1 to distinct multicast group addresses in native IP multicast underlay, and program MSNs and MBGs with the appropriate forwarding information. An SDN multicast controller can alternatively map each designated set of tenant overlay multicast group addresses to a unique underlay multicast IP address and accordingly program MSNs and MBGs with the appropriate forwarding information. In this latter case, these tenant overlay multicast groups are aggregated into distinct multicast group address(es) similar to mechanisms in Default or Data MDTs (Multicast Distribution Trees) [RFC 6037]. This aggregation technique can help minimize multicast routing states in native multicast network.

8.2.7. Multicast Forwarding Programmability
Subsequent to the (re-) computation of the multicast tree, the SDN multicast controller MUST be able to program the multicast forwarding entries on the forwarding elements: MSE, MSN and MBG. When supporting dynamic receivers and senders, multicast tree (re-)computation and programmability must be done relatively fast so that it does not impact the user experience, creating a significant delay in starting to receive a multicast channel that the user joins. In other cases, multicast receiver applications may also have a requirement on the latency between sending a join to a multicast channel and starting to receive data on that channel. Among other things, forwarding programmability SHOULD enable setting the TTL treatment behavior. Specifically, an ingress MBG SHOULD be configurable not to copy the Time to Live (TTL) field from SDN overlay multicast packets. Operators should take account of the end-to-end network topology and set the TTL field accordingly.

8.2.8. Entitlement Admission Control
When this function is enabled for a given tenant, the SDN multicast controller checks if the requesting receiver is entitled to receive the content of the multicast group/channel it is requesting to join. The SDN controller MAY be programmed with the list of receivers that are allowed to receive traffic from certain source(s), and whereby the receivers may be identified by IP host addresses, IP subnets or other types of identifiers. This entitlement information can be configured on the controller or it may be available through some database or system that has that information.

8.2.9. Bandwidth Admission Control
Bandwidth admission control prevents content or subscription overload from senders or receivers as well as provides important security protection by putting upper limits on how much traffic each sender or receiver can send or subscribe to. When this function is enabled, the SDN multicast controller may perform bandwidth admission control when permitting a channel to be delivered to a TES and/or when selecting the multicast channel path. The specification of bandwidth admission control mechanisms is beyond the scope of this framework.

8.3. Multicast Functional Components Implementation
MSE, MSN and MBG are functional components that may be co-located on the same router and implemented in different forms (physical, virtual, or applications). For example, MSEs and MSNs can co-locate together in the same physical or virtual router.

8.4. SDN Multicast Management Interfaces
8.4.1. North-Bound Interface
MApp-C is a northbound interface from the SDN controller that provides management applications the means to interact with the SDN controller to set and extract multicast configurations and information, respectively. It is independent of individual network devices, vendors or device types. MApp-C interface SHOULD provide a comprehensive and flexible data model and messaging methods that allow these applications to access Multicast SDN Controllers to request the addition or removal of a receiver to a multicast (S,G) channel, the establishment of a multicast distribution tree between specified senders and receivers with potential constraints on bandwidth and QoS for
instance, set entitlement rules, and obtain state information pertinent to a multicast group or participants in that multicast group among others. Applications SHOULD have a means to perform these actions across MSDs, each with an SDN controller, with one action rather than separate actions initiated with each of the MSD controllers.

8.4.2. Multicast SDN Functional Component Facing Interfaces

MTTES-C, MSE-C, MSN-C, MBG-C are interfaces between a Multicast SDN Controller on one hand and TESs and Multicast SDN forwarding functional components on the other hand. Each interface is specific to the associated functional component, and is used to provide control plane and/or data plane programming of that component and to harness information from that component.

8.4.3. Traditional Underlay Network Element Interfaces

CN-C is the interface between Multicast SDN controller and core network elements that provide core forwarding services, as opposed to edge services provided by MSE, MSN and MBG. The purpose of CN-C is to provide for programming of network elements in the core whenever it is possible or necessary, and to extract operational state information from these elements.

8.4.4. Information Model and Data Model of Management Interfaces

Both northbound interface and SDN functional component facing interfaces SHOULD be based on RESTCONF [RESTCONF] or NETCONF [RFC6241]. Their data models should be based on YANG.

9. SDN Control Plane

The SDN Control plane can be composed of one or more controllers. These controllers will need to cooperate in the establishment of a multicast distribution tree within MSDs and across MSDs. How they cooperate or interact is dependent on the relationship relative to MSDs, ASes, and administrative entities (operators). In this section, the following cases are considered:
1- Intra-MSD, Inter-SDN Controller Interactions
2- Inter-MSD, Intra-AS, Inter-SDN Controller Interactions
3- Inter-MSD, Inter-AS, Intra-operator, Inter-SDN Controller Interactions
4- Inter-MSD, Inter-AS, Inter-operator, Inter-SDN Controller Interactions

(this section will be completed in a future update).

10. Multi-Party Multicast SDN Overlay Interaction

If multicast senders and receivers are on different multicast SDN overlay vendor solutions, it SHOULD be possible to setup and exchange multicast control states as well as underlying tunneling information.

11. Security Considerations
Following are the security considerations that pertain to the multicast SDN framework in this document:
1- The communication channel between any data element and the SDN controller MUST be trusted and secure, with mutual authentication. This applies to all interfaces (MSE-C, MSN-C, CN-C, MTTES-C and MBG-C) in order to prevent rogue controllers from hijacking multicast traffic, mounting security attacks on network elements or leaking tenant multicast traffic to other tenants.
2- MSE must have means to cope with tenant end systems misbehavior, malicious or malfunctioning, that may result in control plane attack on the SDN controller, impacting the stability of the system and as a result impacting multicast services. The MSE MUST be able to limit the message rate that could be triggered by tenant end system activity, and the controller MUST be able to limit message rate and routing activity changes at the tenant level as that could result in a DDOS on the multicast services across tenants.
3- The association between a TES and tenant network must be authentic so that traffic eavesdropping is prevented and a TES is not allowed to illegally inject multicast traffic into another tenant network.
4- The SDN controller SHOULD have the means to authorize a TES to join a multicast channel (entitlement).
5- The SDN controller SHOULD have the means to exercise admission control to avoid congestion on network links as causing congestion could be a way to launch a DDOS.

12. IANA Considerations
The document does not require any IANA action.

13. References
13.1. Normative References
13.2. Informative References


Internet-Draft SDN Multicast Overlay July 2017


14. Acknowledgments

The authors thank Dino Farinacci for his comments and input.

Authors Addresses
David Qi
Bloomberg
EMail: dqi@bloomberg.net

Nabil Bitar
Nokia
EMail: nabil.bitar@nokia.com

Truman Boyes
Bloomberg
EMail: tboyes@bloomberg.net

Senad Palislamovic
Nokia
EMail: senad.palislamovic@nokia.com

Anil Lohiya
Juniper
EMail: alohiya@juniper.net