Network Working Group
Internet Draft
Intended status: Informational
Expires: April 2015

Z. Qiang
Ericsson

October 27, 2014

Supporting Applications Specific Multicast in NVO3
draft-zu-nvo3-mcast-considerations-00.txt

Status of this Memo

This Internet-Draft will expire on April 27, 2015.

Copyright Notice

Abstract

The framework of supporting applications specific multicast traffic in a network using Network Virtualization using Overlays over Layer 3 (NVO3) has been discussed. The various mechanisms and considerations that can be used for delivering those application specific multicast traffic in networks that use NVO3 have been considered.

This draft discusses some additional considerations on how to support applications specific multicast traffic in NVO3.

Table of Contents

1. Introduction

Network virtualization using Overlays over Layer 3 (NVO3) is a technology that is used to address issues that arise in building Large, multitenant data centers that make extensive use of server virtualization [RFC7364].

The framework of supporting application specific multicast traffic in a network that uses Network Virtualization using Overlays over Layer 3 (NVO3) is discussed in the draft [NVO3-MC]. It describes 4 mechanisms and some considerations that can be used for delivering those application specific multicast traffic in networks that use NVO3.

This draft discusses some additional considerations on how to support applications specific multicast traffic in NVO3.

The reader is assumed to be familiar with the terminology as defined in the NVO3 Framework document [RFC7365] and NVO3 Architecture document [NVO3-ARCH].

2. 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].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying RFC-2119 significance.

3. Terminology

This document uses the same terminology as found in [NVO3-MC].

4. Considerations

While the mechanisms discussed in Section 3 of [NVO3-MC] have been discussed individually, it is important for a development to support one or more methods in a large, multitenant data centers. It is hard to say which method is better than the others without considerations of the tenant system overlay network architecture. This document attempts to provide considerations on each mechanism detailed in section 3 of [NVO3-MC].

The source NVE replication method descripted in Section 3.2 of the [NVO3-MC] may be more attractive for a tenant system which only has a few NVEs participating the overlay forwarding, even the tenant system may have hundreds of VMs. As the number of participating NVE is not that many, the source NVE replication will not generate too many duplicated traffic traverse the same overlay network connection. Additional overlay control signaling may provide some level of optimizations. Furthermore, it is not a big scaling issue for a multi-tenant data center which the number of tenant networks can be hundreds.

However, if a tenant system contains many VMs attaching to a large number of NVEs, using source NVE replication method may generate large amount of duplicated traffic on the same overlay network connection.

Another alternative, which can be used to avoid network connections overloading due to the duplicated traffic, is to use an IP multicast in underlay method descripted in Section 3.4 of [NVO3-MC] to handle the application multicast traffic. However, additional NVA-NVE control signaling may be needed in order to enable the multicast address mapping at source NVE. This method however introduces an additional scaling problem if there are too many multicast groups within a multi-tenancy data center which may have a large number of tenant networks.

The service node method descripted in Section 3.3 of [NVO3-MC] may be preferable for a tenant network where the multicast source VM and the receiver VMs are distributed and the service node can be placed closer to the multicast receiving VMs. Otherwise, this alternative will have same duplicated traffic overloading issue as the source NVE replication mechanism. In order to enable the service node function at the property location, the NVA may need to know the participants list of each multicast group in advance. Ideally, the service node function shall be placed closer to the receivers attached NVEs. Indeed, if the service node function and one of the receivers attached NVEs are collocated, this method is very similar to the multicast distribution tree mechanism described in MVPN [RFC6513].

5. Optimizations

In order to provide optimal routing for a particular multicast flow and to improve the multicast scalability as unicast traffic, the source NVE shall forward the received multicast packet to the destination NVEs only if the destination NVE has at least one participating VM of that multicast group. Duplication of the

multicast packet to the destination NVEs on the same network connection shall be avoided.

One alternative for the above optimization is to use the multicast distribution proxy for each multicast group, which is similar to the multicast distribution tree mechanism described in MVPN [RFC6513]. Instead of using BGP as control plane signaling, the NVA-NVE signaling can be used to setup this multicast distribution proxy architecture per multicast group.

To avoid generating large amount of duplicated traffic on the same overlay network connection, the source NVE may duplicate the multicast traffic and only send it to a limited number of proxy NVEs. Then the proxy NVE can further duplicate the multicast traffic and forward it to the rest of the receiving NVEs.

6. Procedures

In this document, the optimization alternative of how the multicast distribution proxy can be applied in the NVO3 architecture is discussed.

The procedure is to make use of the NVA, which is the centralized control node of the NVO3 architecture, to select the proxy NVEs from the participating NVEs and setup a distribution path for each multicast group.

6.1. Source NVE

Source NVE is the NVE which the multicast application VM is attached. It is assumed that the source NVE knows if a multicast application VM is attached. The NVE may know it via overlay network provisioning or using some kind multicast monitoring functions, e.g. IGMP snooping.

How the multicast monitoring is performed is out of the scope of this document.

The source NVE needs to register itself to the NVA in order to enable the multicast distribution mechanism.

Once the source NVE registration is done, based on the multicast registration information and the multicast proxy function role selection decision, a multicast proxy NVE list and a Participant NVE list are created by the NVA.

Multicast Proxy NVE list is a list of Proxy NVEs. It is created by the NVA and saved in the source NVE. It is used by the source NVE for multicast traffic duplication and distribution. And the source NVE is updated with the multicast proxy NVE list by NVA using the NVA-NVE control plane signaling.

6.2. Receiving NVE

Receiving NVEs is the NVE where a multicast group participanting VMs attached.
It is assumed that the Receiving NVE knows if an attached VM would like to join a multicast group. The NVE may know it via overlay network provisioning or using some kind multicast monitoring functions, e.g. IGMP snooping.

How the multicast monitoring is performed is out of the scope of this document.

The Receiving NVE shall inform the NVA if there is any VM multicast registrations, or if all the registered VMs of a given multicast group discontinue participating to the multicast group. The NVA uses this information to create / update the Participant NVE list of the multicast group.

Participant NVE list is a list of receiving NVEs. It is created by the NVA and saved in the Proxy NVE. It is used by the Proxy NVE for multicast traffic duplication and distribution.

6.3. Multicast Proxy function

Proxy NVE is one of the receiving NVEs. The proxy NVE is selected by the NVA. It has the responsibility of distributing the received multicast traffic to the participant NVEs.
Multicast Proxy function role is assigned to one of the Receiving NVEs. The multicast proxy function role is dynamic allocated by the NVA at per multicast group and per tenant system. The proxy NVE will receive the multi-cast traffic from the source NVE and distribute it to other receiving NVEs in the receiving NVE list.

When receiving the multicast registration information, the NVA selects one NVE as multicast proxy. The selection may be based on several conditions, such as forwarding capability, location, network segmentations, etc.

A Participant NVE list will be sent to the Multicast Proxy NVE for multicast traffic duplication and distribution. The Participant NVE list is created at per Proxy NVE based. This is to avoid any unnecessary traffic duplication and looping.

6.4. Receiving multicast packets at Source NVE

When receiving a multicast packet from the attached VM, the source NVE shall handle the packet as followings:

If the multicast Proxy NVE list of a given multicast group is empty, the source NVE shall not forward any multicast packet of the given multicast group when receiving it from the attached VM.

If the multicast Proxy NVE list of a given multicast group is not empty, the source NVE shall duplicate, encapsulate, and forward the received multicast packet to each Proxy NVEs based on the multicast Proxy NVE list received from NVA.

6.5. Receiving multicast packets at Proxy NVE

The NVE which has been assigned with the multicast proxy function role will have the responsibility to distribute the multicast traffic based on the received multicast Participated NVE list.

When receiving a multicast packet from the source NVE, the proxy NVE shall dencapsulate, duplicate, encapsulate, and forward the multicast packet to the destination NVEs based on the multicast Participated NVE list received from NVA.

6.6. Receiving multicast packets at Receiving NVE

When receiving a multicast packet from a Proxy NVE, the receiving NVE shall dencapsulate multicast packet, and forward to the attached VM which has registered to the multicast group.

7. Security Considerations

This is a discussion paper which provides inputs for the NVO3 requirement documents and in itself does not introduce any new security concerns.
TBD

8. IANA Considerations

No actions are required from IANA for this informational document.
TBD

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2234] Crocker, D. and Overell, P.(Editors), "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, Internet Mail Consortium and Demon Internet Ltd., November 1997.

9.2. Informative References

[NVO3-MC] "Framework of Supporting Applications Specific Multicast in NVO3", June 6, 2014.

[RFC6513] "Multicast in MPLS/BGP IP VPNs", February 2012

[RFC7364] "Problem statement: Overlays for network virtualization", October 2014.

[RFC7365] "Framework for Data Center (DC) Network Virtualization", October 2014.

[NVO3-ARCH] Narten, T. et al.," An Architecture for Overlay Networks (NVO3)", work in progress, Feb 2014.

10. Acknowledgments

Many people have contributed to the development of this document and many more will probably do so before we are done with it. While we cannot thank all contributors, some have played an especially prominent role. The following have provided essential input: Suresh Krishnan.

Authors' Addresses

Zu Qiang
Ericsson
8400, boul. Decarie
Ville Mont-Royal, QC,
Canada

Email: Zu.Qiang@Ericsson.com