NVO3 Working Group
INTERNET-DRAFT
Intended Status: Informational
Expires: January 7, 2016

Y. Li
L. Yong
Huawei Technologies

July 6, 2015

VLAN Configuration Considerations in Split-NVE
draft-yizhou-nvo3-vlan-config-in-split-nve-00

Abstract

In a Split-NVE structure, a control plane protocol between a hypervisor and its associated external NVE(s) to distribute the virtual machine networking state and the relevant attributes. One of the key attributes to be negotiated is VLAN ID which is the most common locally-significant tag for carrying traffic associated with a specific virtual network. This document provides the informational guides on how to configure the VLAN IDs to local networks in Split-NVE structure.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

Table of Contents

1. Introduction

The problem statement [RFC7364], discusses the needs for a control plane protocol (or protocols) to populate each NVE with the state needed to perform the required functions in Split-NVE scenario. The protocol requirement [I-D.ietf-nvo3-hpvr2nve-cp-req] presents one of the key requirements which allows the negotiation on a locally-significant tag for carrying traffic associated with a specific virtual network. The tag is commonly a VLAN ID [IEEE 802.1Q]. This document uses the term "VLAN ID" or VID to cover the locally-significant tag. Traffic isolation in overlay network is based on virtual network ID. Before the traffic entering the ingress point of the overlay network, isolation is based on VLAN ID.

A bridged network may connect end Devices to external NVE. We refer it as indirect connection. Another case is direction connect which means end device directly connects to the external NVE without going through any intermediate device. Figure 1 shows the two connection types in local network.

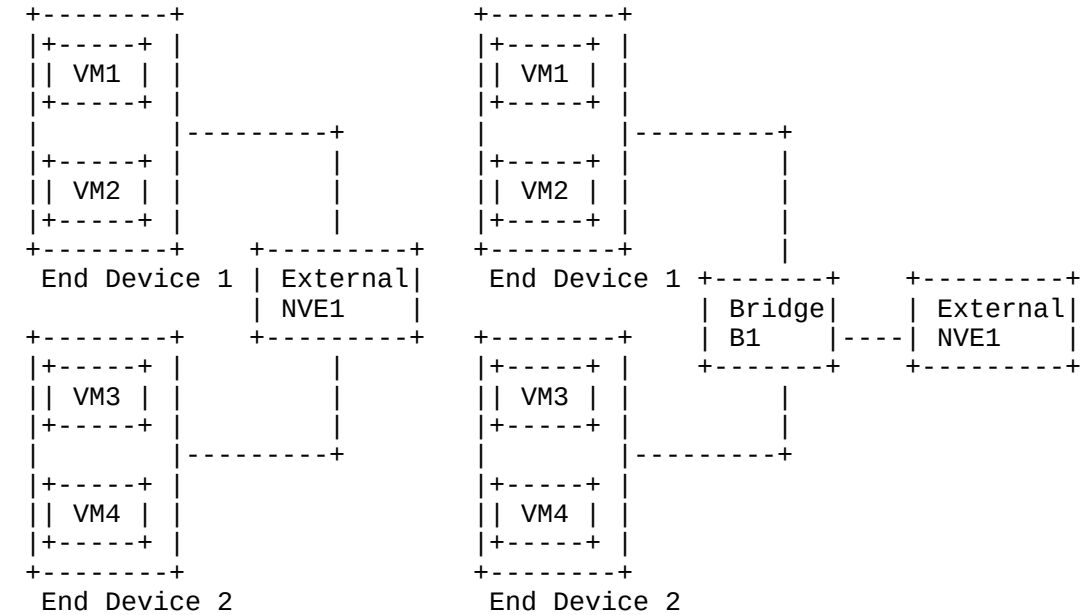

Figure 1 Direct Connection (left) and Indirect Connection (right)

Some scenarios require the switching among virtual machines to be always performed on the external NVE rather than on end device or an intermediate bridge (if any). It helps to ease the policy enforcement. Such forwarding mode is called reflective relay (RR) or hairpin forwarding. A received frame on a port that supports reflective relay mode can be forwarded on the same port on which it was received. Figure 2 and 3 show the expected traffic flow when RR mode is used in direct and indirect connection respectively. The numbers in brackets indicate the expected sequence and the number with a prime indicates simultaneous sequence when the multicast traffic is considered. To achieve the expected local traffic isolation could be tricky especially for that shown in figure 3 if we consider the intermediate bridge is a traditional switch that is only able to identify VLAN tags.

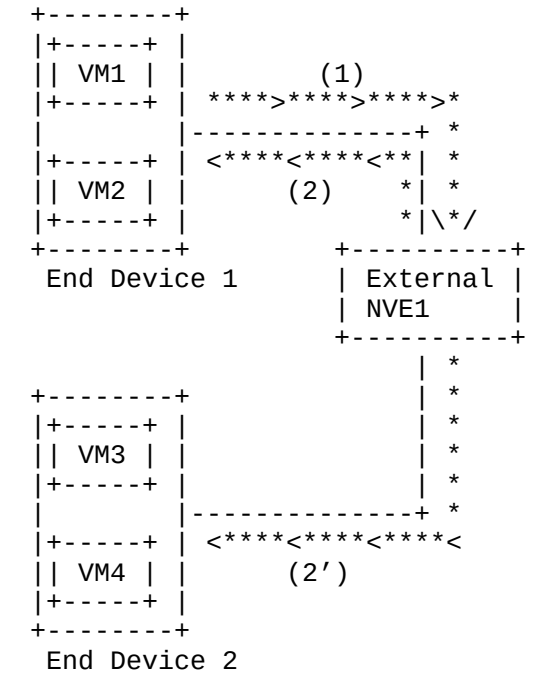

Figure 2 Reflective Relay Mode in Direct Connection

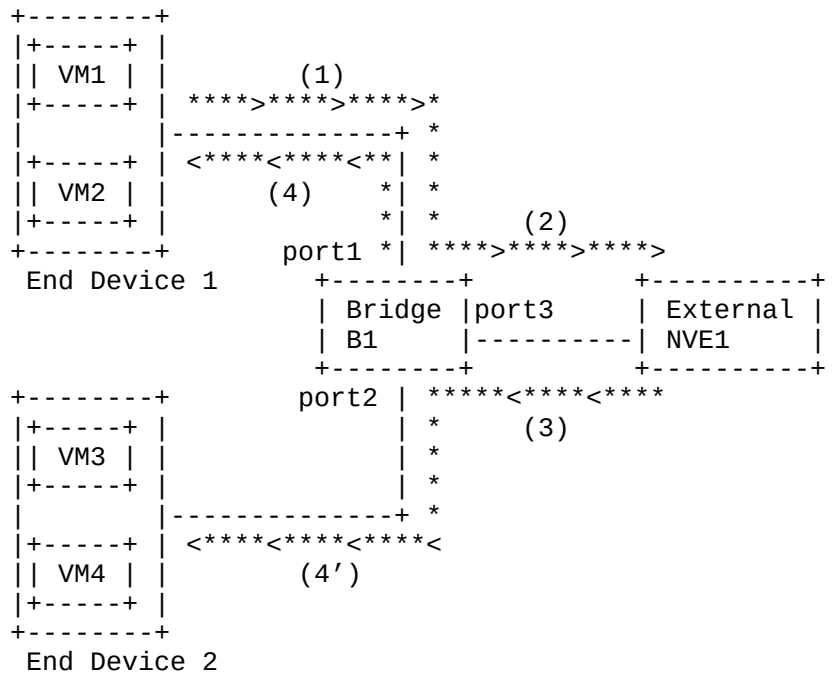

Figure 3 Reflective Relay Mode in Indirect Connection

This document provides the information on how to correctly configure the VLAN IDs to achieve the traffic isolation in local network for either direct or indirect connection and for either RR forwarding or normal forwarding mode.

1.1 Terminology

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].

This document uses the same terminology as found in [RFC7365] and [I-D.ietf-nvo3-hpvr2nve-cp-req].

RR - Reflective Relay. A received frame on a port that supports reflective relay mode can be forwarded on the same port on which it was received.

2. VLAN ID Configurations

The most common approach is to configure VLAN on per VN base in the

local network. It works well for most scenarios. If we examine the scenarios from two dimensions, direct or indirect connection, and RR mode or traditional forwarding mode, such VLAN configuration is not applicable to indirect connection and RR mode case for both unicast and multicast. Take figure 3 as example, we assume VM1, VM2 and VM3 are all belonging to the same VN, say VN100. When local VLAN ID is configured based on per VN, the packet from VM1 to VM3 will be forwarded by intermediate bridge B1 directly without NVE1 involved. It violates the expected behaviors in RR mode. If VM1 sends a multicast packet in VN100, intermediate Bridge B1 will forward to port 2 and port 3, NVE1 receives it and hairpins it back to B1. B1 will replicate it to port 1 and port 2. Then VM3 will receive duplicate copies which is not a correct behavior expected.

There are two potential ways to configure VLAN IDs in indirect connection and RR forwarding mode case to fulfil the local traffic isolation requirement.

2.1 VLAN ID per VM

When configuring different VLAN IDs for each VM and let NVE associate these VLAN ID to the same VN, it naturally ensures that the frame from one VM to another is not locally switched at the intermediate bridges. It requires a lot of work at the external NVE. NVE needs to remember the VN to VLAN ID mappings and performs the VLAN ID translations for unicast packet. For multicast traffic, the external NVE needs to replicate the packet to each of the VLANs belonging to the same VN. One way to save such effort for multicast packets is to use per-VN based VLAN ID for downstream multicast traffic. Downstream traffic here refers the multicast packets forwarded by external NVE to potential recipient VMs. Per-VN based VLAN IDs should not overlap with per-VM based VLAN IDs with this approach. Number of VLANs are consumed very quickly in this case.

2.2 Private VLAN configuration per VN

The intermediate bridge can be configured as private VLAN [RFC5517] deployment. Each VN consumes two VLAN IDs in this case. Primary VLAN ID needs to be configured on the uplink port of the intermediate bridge and the port type is set to be Promiscuous Port. Secondary VLAN ID needs to be configured on the down link ports of the intermediate bridge and the port type is set to be Isolated Ports to prohibit the direct communicating between any ports of them. Such setting should be per VN base. The shared VLAN learning (SVL) [IEEE 802.1Q] needs to be enabled for primary and secondary VLAN per VN.

To support RR mode on NVE, the intermediate bridge MUST disable MAC learning on the uplink port. As a result, the frame from a down link

port of the intermediate bridge will be sent to the uplink port as an unknown unicast frame to the external NVE. Such configuration will prevent the MAC learning hopping between the uplink and downlink ports in shared VLAN learning case.

3. Summary

In indirect connection scenarios, the intermediate bridge has to be carefully configured with VLAN IDs especially when RR forwarding is enabled on the external NVE and end device. The protocol running between the hypervisor of the end device and the external NVE does not have the capability to configure the intermediate bridge. Therefore the network management system is required to configure the intermediate bridge when indirect connection has to be used. The MVRP [IEEE802.1ak] may facilitate the auto VLAN ID configuration at the intermediate bridge in some cases.

4. Security Considerations

TBD

5. IANA Considerations

No IANA action is required. RFC Editor: please delete this section before publication.

6. References

6.1 Normative References

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

[RFC5517] HomChaudhuri, S. and M. Foschiano, "Cisco Systems' Private VLANs: Scalable Security in a Multi-Client Environment", RFC 5517, February 2010.

6.2 Informative References

[RFC7364] Narten, T., Gray, E., Black, D., Fang, L., Kreeger, L., and M. Napierala, "Problem Statement: Overlays for Network Virtualization", October 2014.

[RFC7365] Lasserre, M., Balus, F., Morin, T., Bitar, N., and Y. Rekhter, "Framework for DC Network Virtualization", October 2014.

[I-D.ietf-nvo3-nve-nva-cp-req] Kreeger, L., Dutt, D., Narten, T., and D. Black, "Network Virtualization NVE to NVA Control Protocol Requirements", draft-ietf-nvo3-nve-nva-cp-req-01 (work in progress), October 2014.

[I-D.ietf-nvo3-arch] Black, D., Narten, T., et al, "An Architecture for Overlay Networks (NVO3)", draft-narten-nvo3-arch, work in progress.

[I-D.ietf-nvo3-hpvr2nve-cp-req] Yizhou, L., Yong, L., Kreeger, L., Narten, T., and D. Black, "Hypervisor to NVE Control Plane Requirements", draft-ietf-nvo3-hpvr2nve-cp-req-02 (work in progress), February 2015.

[IEEE 802.1Qbg] IEEE, "Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks - Amendment 21: Edge Virtual Bridging", IEEE Std 802.1Qbg, 2012.

[8021Q] IEEE, "Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks", IEEE Std 802.1Q-2011, August, 2011.

[802.1ak] IEEE, "IEEE Standard for Local and Metropolitan Area Networks - Virtual Bridged Local Area Networks - Amendment 07: Multiple Registration Protocol", IEEE Std 802.1ak-2007, 2007.

Authors' Addresses

Yizhou Li
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China

Phone: +86-25-56624629
EMail: liyizhou@huawei.com

Lucy Yong
Huawei Technologies, USA

Email: lucy.yong@huawei.com