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/ietf/1id-abstracts.txt.

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

This Internet-Draft will expire on March 19, 2013.

Copyright Notice

Copyright (c) 2009 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.
Abstract

This document specifies NVGRE and VXLAN encapsulation for L3VPN Extension. Both NVGRE and VXLAN are originally designed for L2 overlay. The draft proposes the enhancement on both to allow L3 overlay completely decoupled from the L2 overlay in terms of encoding schema and data processing.

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...................................................3
2. NVGRE and VXLAN Encapsulation for L3VPN Extension..............3
   2.1. NVGRE Enhancement for L3VPN Extension....................3
   2.2. VXLAN Enhancement for L3VPN Extension....................4
   2.3. Benefits................................................5
3. Security Considerations........................................5
4. IANA Considerations............................................6
5. References....................................................6
   5.1. Normative References......................................6
   5.2. Informative References...............................6
1. Introduction

L3VPN Extension [ESRQMT] requires that a BGP based L3VPN [RFC4364] is able to expand to the end systems in a DC, i.e. to a server where Guest OS and Host OS/Hypervisor reside. This implies that the CE and PE components in the L3VPN architecture may reside on a same device. To complement current practical server implementations, the L3VPN Extension further proposes utilizing NVGRE [NVGRE] and/or VXLAN [VXLAN] data encapsulation formats [DRAO]. The approach brings the advantage to use a unified solution for both L2 overlay [SALI] [DRAKE] and L3 overlay services. However, both NVGRE and VXLAN are originally designed as a L2 overlay data encapsulation in which the inner header MUST be an Ethernet header.

This document proposes an enhancement to NVGRE and VXLAN encapsulation formats that allow them to be used as the same data encapsulation semantics for both L2 overlay and L3 overlay services. The benefits of this usage include maintaining L3VPN natively and decoupling it from the L2 overlay completely.

2. NVGRE and VXLAN Encapsulation for L3VPN Extension

A BGP/MPLS L3VPN [RFC4364] requires that the inner header is IP header, either IPv4 or IPv6. Both NVGRE [NVGRE] and VXLAN [VXLAN] encapsulation schema require that the inner header MUST be an Ethernet header. The solution of [DRAO] suggests several options to fill the MAC address fields when using NVGRE and/or VXLAN in a BGP based L3VPN, although inner Ethernet address is not relevant to the L3VPN mechanism. These options make L3 overlay still coupled with L2 overlay one way or another, which does not meet the requirement of L3VPN Extension [ESRQMT].

2.1. NVGRE Enhancement for L3VPN Extension

NVGRE [NVGRE] leverages the GRE protocol [RFC2890] and specifies that the protocol type field in the GRE header MUST be filled with the value of 0x6558, which means for Transparent Ethernet.

This draft proposes to allow the protocol type field to be either the value of 0x6558 or 0x0800. The value 0x0800 means IP payload [RFC3232]. The value 0x6558 MUST be used if the inner header is Ethernet header. The value 0x0800 MUST be used if the inner header is IP header. Furthermore, the version field in IP header indicates if the IP header is IPv4 or IPv6. When L3VPN Extension uses NVGRE encapsulation, it MUST use the value of 0x0800 in the protocol type.
field and encode IP header as the inner header. Other fields in the outer header of the NVGRE remain the same.

2.2. VXLAN Enhancement for L3VPN Extension

This document proposes adding a protocol type field in the VXLAN header as indicated below. It takes 16 bits from the reserved 24 bits as the protocol type field. For L2 overlay encapsulation, the protocol type field MUST be filled with the value of 0x6558. For L3 overlay encapsulation, the protocol type field MUST be filled with the value of 0x0800, which means IP header [RFC3232]; inner header MUST be IP header as shown below. The remained 8 reserved bit MUST be filled with zero.

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
Outer Ethernet Header:
   As described in VXLAN [VXLAN]
Outer IP Header:
   As described in VXLAN [VXLAN]
Outer UDP Header:
   As described in VXLAN [VXLAN]
VXLAN Header:
   +-----------------------------------------------+
   | R | R | R | R | I | R | R | R | Reserved | Protocol Type (0x0800) |
   +-----------------------------------------------+
   | VNI | Reserved |
   +-----------------------------------------------+
Inner IPv4 Header:
   +-----------------------------------------------+
   | Version | IHL | Type of Service | Total Length |
   +-----------------------------------------------+
   | Identification | Flags | Fragment Offset |
   +-----------------------------------------------+
   | Time to Live | Protocol | Header Checksum |
   +-----------------------------------------------+
   | Source Address |
   +-----------------------------------------------+
   | Destination Address |
   +-----------------------------------------------+
   | Options | Padding |
   +-----------------------------------------------+
   | Original IP Payload |
   +-----------------------------------------------+
When L3VPN Extension uses VXLAN encapsulation, it MUST use the value of 0x0800 in the protocol type field and encode IP header as the inner header. Other fields in the outer header and the VXLAN header remain the same.

For an L2 overlay, the protocol type value MUST be 0x6558, and inner header MUST be Ethernet header and use the same format as in [VXLAN].

Another option to achieve this is to use one or more of reserved bits to indicate inner header type. For example, use the bit 6 in the VLAN header. However, this option has limited space for future expansion and makes NVGRE and VXLAN use of different semantics to indicate inner header.

2.3. Benefits

The enhancement on NVGRE and VXLAN encapsulation formats brings some beneficial for L3VPN Extension to use NVGRE and/or VXLAN as a data encapsulation format:

- Maintain L3VPN implementation natively and decouple it completely from L2 overlay implementation.
- Extend VXLAN and NVGRE encapsulation formats to support seamless L2 and L3 overlay interworking [E-IP-VPN].
- BGP control plane works consistently with the data plane in term of multiple protocol support, i.e. the inner header on a data packet matches the address family being advertised by BGP route UPDATE message.
- Save 12 bytes in every packet in a native L3 overlay and lower the probability of the packet fragmentation due to a shorter inner header for the L3 overlay.

It is worth mention that the protocol extension in this document does not impact any mechanism that is specified in NVGRE [NVGRE] and VXLAN [VXLAN].

3. Security Considerations

The mechanism proposed in this document does not add any additional security concern beside what has been described in the NVGRE [NVGRE] and VXLAN [VXLAN].
4. IANA Considerations

The document does not require any IANA action.

5. References

5.1. Normative References


[RFC2890] Dommety, G., "key and sequence Number Extentions to GRE", RFC2890, September 2000


5.2. Informative References


Authors’ Addresses

Lucy Yong
Huawei USA
4320 Legacy Driver
Plano, TX  75025
U.S.A

Phone:  469-277-5837
Email:  lucy.yong@huawei.com

Xiaohu Xu
Huawei Technologies,
Beijing, China

Phone: +86-10-60610041
Email: xuxiaohu@huawei.com