NVO3 Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 11, 2021

X. Min
G. Mirsky
ZTE Corp.
S. Pallagatti
VMware
J. Tantsura
Juniper Networks
March 10, 2021

BFD for Geneve
draft-ietf-nvo3-bfd-geneve-02

Abstract

This document describes the use of the Bidirectional Forwarding Detection (BFD) protocol in point-to-point Generic Network Virtualization Encapsulation (Geneve) tunnels used to make up an overlay network.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

"Generic Network Virtualization Encapsulation" (Geneve) [RFC8926] provides an encapsulation scheme that allows building an overlay network by decoupling the address space of the attached virtual hosts from that of the network.

This document describes the use of Bidirectional Forwarding Detection (BFD) protocol [RFC5880] to enable monitoring continuity of the path between two Geneve tunnel endpoints, which may be NVE (Network Virtualization Edge) or other device acting as a Geneve tunnel endpoint. Specifically, the asynchronous mode of BFD, as defined in [RFC5880], is used to monitor a p2p Geneve tunnel, and support for BFD Echo function is outside the scope of this document. For simplicity, in this document, NVE is used to represent Geneve tunnel endpoint, TS (Tenant System) is used to represent the physical or virtual device attached to a Geneve tunnel endpoint from the outside. VAP (Virtual Access Point) is the NVE side of the interface between the NVE and the TS, and a VAP is a logical network port (virtual or physical) into a specific virtual network. For detailed definitions and descriptions of NVE, TS and VAP, please refer to [RFC7365] and [RFC8014].

The use cases and the deployment of BFD for Geneve are consistent with what's described in Section 1 and 3 of [RFC8971] ("Bidirectional

Forwarding Detection (BFD) for Virtual eXtensible Local Area Network (VXLAN)"), except for the usage of Management VNI, which in the case of Geneve is described in [I-D.ietf-nvo3-geneve-oam], and outside the scope of this document. The major difference between Geneve and VXLAN [RFC7348] is that Geneve supports multi-protocol payload and variable length options.

2. Conventions Used in This Document

2.1. Abbreviations

BFD: Bidirectional Forwarding Detection

EVPN: Ethernet Virtual Private Networks

Geneve: Generic Network Virtualization Encapsulation

NVE: Network Virtualization Edge

TS: Tenant System

VAP: Virtual Access Point

VNI: Virtual Network Identifier

VXLAN: Virtual eXtensible Local Area Network

2.2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. BFD Packet Transmission over Geneve Tunnel

Concerning whether the Geneve data packets include an Ethernet frame or an IP packet, this document defines two formats of BFD packet encapsulation in Geneve. BFD session is originated and terminated at VAP of an NVE, selection of the BFD packet encapsulation is based on how the VAP encapsulates the data packets.

3.1. BFD Encapsulation With Inner Ethernet/IP/UDP Header

If the VAP that originates the BFD packets is used to encapsulate Ethernet data frames, then BFD packets are encapsulated in Geneve as described below. The Geneve packet format over IPv4 is defined in

Section 3.1 of [RFC8926]. The Geneve packet format over IPv6 is defined in Section 3.2 of [RFC8926]. The Outer IP/UDP and Geneve headers MUST be encoded by the sender as defined in [RFC8926]. Note that the outer IP header and the inner IP header may not be of the same address family, in other words, outer IPv6 header accompanied with inner IPv4 header and outer IPv4 header accompanied with inner IPv6 header are both possible.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 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                    ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                        Outer IPvX Header                      ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                        Outer UDP Header                       ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                          Geneve Header                        ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                      Inner Ethernet Header                    ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                        Inner IPvX Header                      ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                         Inner UDP Header                      ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                        BFD Control Packet                     ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Outer Ethernet FCS                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: Geneve Encapsulation of BFD Control Packet With the Inner Ethernet/IP/UDP Header

The BFD packet MUST be carried inside the inner Ethernet frame of the Geneve packet. The inner Ethernet frame carrying the BFD Control packet has the following format:

Ethernet Header:

Source MAC: MAC address of a VAP of the originating NVE.

Destination MAC: MAC address of a VAP of the terminating NVE.

IP Header:

Source IP: IP address of a VAP of the originating NVE. If the VAP of the originating NVE has no IP address, then the IP address 0.0.0.0 for IPv4 or ::/128 for IPv6 MUST be used.

Destination IP: IP address of a VAP of the terminating NVE. If the VAP of the terminating NVE has no IP address, then the IP address 127.0.0.1 for IPv4 or ::1/128 for IPv6 MUST be used.

TTL or Hop Limit: MUST be set to 255 in accordance with [RFC5881].

The fields of the UDP header and the BFD Control packet are encoded as specified in [RFC5881].

When the BFD packets are encapsulated in Geneve in this way, the Geneve header defined in [RFC8926] follows the value set below.

Opt Len field SHOULD be set to 0, which indicates there isn't any variable length option.

O bit MUST be set to 1, which indicates this packet contains a control message.

C bit MUST be set to 0, which indicates there isn't any critical option.

Protocol Type field MUST be set to 0x6558 (Ethernet frame).

Virtual Network Identifier (VNI) field SHOULD be set to the VNI number that the originating VAP is mapped to.

3.2. BFD Encapsulation With Inner IP/UDP Header

If the VAP that originates the BFD packets is used to encapsulate IP data packets, then BFD packets are encapsulated in Geneve as described below. The Geneve packet format over IPv4 is defined in

Section 3.1 of [RFC8926]. The Geneve packet format over IPv6 is defined in Section 3.2 of [RFC8926]. The Outer IP/UDP and Geneve headers MUST be encoded by the sender as defined in [RFC8926]. Note that the outer IP header and the inner IP header may not be of the same address family, in other words, outer IPv6 header accompanied with inner IPv4 header and outer IPv4 header accompanied with inner IPv6 header are both possible.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                      Ethernet Header                          ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                     Outer IPvX Header                         ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                     Outer UDP Header                          ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                       Geneve Header                           ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                     Inner IPvX Header                         ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                      Inner UDP Header                         ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                     BFD Control Packet                        ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             FCS                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Geneve Encapsulation of BFD Control Packet With the Inner IP/UDP Header

The BFD packet MUST be carried inside the inner IP packet of the Geneve packet. The inner IP packet carrying the BFD Control packet has the following format:

IP header:

Source IP: IP address of a VAP of the originating NVE.

Destination IP: IP address of a VAP of the terminating NVE.

TTL or Hop Limit: MUST be set to 255 in accordance with [RFC5881].

The fields of the UDP header and the BFD Control packet are encoded as specified in [RFC5881].

When the BFD packets are encapsulated in Geneve in this way, the Geneve header defined in [RFC8926] follows the value set below.

Opt Len field SHOULD be set to 0, which indicates there isn't any variable length option.

O bit MUST be set to 1, which indicates this packet contains a control message.

C bit MUST be set to 0, which indicates there isn't any critical option.

Protocol Type field MUST be set to 0x0800 (IPv4) or 0x86DD (IPv6), depending on the address family of the inner IP packet.

Virtual Network Identifier (VNI) field SHOULD be set to the VNI number that the originating VAP is mapped to.

4. Reception of BFD packet from Geneve Tunnel

Once a packet is received, the NVE MUST validate the packet as described in [RFC8926].

If the Protocol Type field equals 0x6558 (Ethernet frame), and the Destination MAC of the inner Ethernet frame matches the MAC address of a VAP which is mapped to the same as received VNI, then the Destination IP, the UDP destination port and the TTL or Hop Limit of the inner IP packet MUST be validated to determine whether the received packet can be processed by BFD.

If the Protocol Type field equals 0x0800 (IPv4) or 0x86DD (IPv6), and the Destination IP of the inner IP packet matches the IP address of a VAP which is mapped to the same as received VNI, then the UDP destination port and the TTL or Hop Limit of the inner IP packet MUST be validated to determine whether the received packet can be processed by BFD.

4.1. Demultiplexing of the BFD packet

In BFD over Geneve, a BFD session is originated and terminated at VAP, usually one NVE owns multiple VAPs, so multiple BFD sessions may be running between two NVEs, there needs to be a mechanism for demultiplexing received BFD packets to the proper session. Furthermore, due to the fact that [RFC8014] allows for N-to-1 mapping between VAP and VNI at one NVE, multiple BFD sessions between two NVEs for the same VNI are allowed. Also note that a BFD session can only be established between two VAPs that are mapped to the same VNI and use the same way to encapsulate data packets.

If the BFD packet is received with Your Discriminator equals to 0, for different BFD encapsulation, the procedure for demultiplexing the received BFD packets is different.

> When the BFD Encapsulation With Inner Ethernet/IP/UDP Header is used, the BFD session MUST be identified using the VNI number, and the inner Ethernet/IP/UDP Header, i.e., the source MAC, the source IP, the destination MAC, the destination IP, and the source UDP port number present in the inner Ethernet/IP/UDP header.

> When the BFD Encapsulation With Inner IP/UDP Header is used, the BFD session MUST be identified using the VNI number, and the inner IP/UDP header, i.e., the source IP, the destination IP, and the source UDP port number present in the inner IP/UDP header.

If the BFD packet is received with non-zero Your Discriminator, then the BFD session MUST be demultiplexed only with Your Discriminator as the key. The exchange of BFD discriminators may be achieved by echo request/reply, EVPN, etc. The detailed mechanism on how to exchange the BFD discriminators is outside the scope of this document.

5. Security Considerations

This document does not raise any additional security issues beyond those of the specifications referred to in the list of references.

6. IANA Considerations

This document has no IANA action requested.

7. Acknowledgements

The authors would like to acknowledge Reshad Rahman, Jeffrey Haas and Matthew Bocci for their guidance on this work.

The authors would like to acknowledge David Black for his explanation on the mapping relation between VAP and VNI.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <https://www.rfc-editor.org/info/rfc5880>.

[RFC5881] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)", RFC 5881, DOI 10.17487/RFC5881, June 2010, <https://www.rfc-editor.org/info/rfc5881>.

[RFC7365] Lasserre, M., Balus, F., Morin, T., Bitar, N., and Y. Rekhter, "Framework for Data Center (DC) Network Virtualization", RFC 7365, DOI 10.17487/RFC7365, October 2014, <https://www.rfc-editor.org/info/rfc7365>.

[RFC8014] Black, D., Hudson, J., Kreeger, L., Lasserre, M., and T. Narten, "An Architecture for Data-Center Network Virtualization over Layer 3 (NVO3)", RFC 8014, DOI 10.17487/RFC8014, December 2016, <https://www.rfc-editor.org/info/rfc8014>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8926] Gross, J., Ed., Ganga, I., Ed., and T. Sridhar, Ed., "Geneve: Generic Network Virtualization Encapsulation", RFC 8926, DOI 10.17487/RFC8926, November 2020, <https://www.rfc-editor.org/info/rfc8926>.

8.2. Informative References

[I-D.ietf-nvo3-geneve-oam] Mirsky, G., Boutros, S., Black, D., and S. Pallagatti, "OAM for use in GENEVE", draft-ietf-nvo3-geneve-oam-01 (work in progress), November 2020.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <https://www.rfc-editor.org/info/rfc7348>.

[RFC8971] Pallagatti, S., Ed., Mirsky, G., Ed., Paragiri, S., Govindan, V., and M. Mudigonda, "Bidirectional Forwarding Detection (BFD) for Virtual eXtensible Local Area Network (VXLAN)", RFC 8971, DOI 10.17487/RFC8971, December 2020, <https://www.rfc-editor.org/info/rfc8971>.

Authors' Addresses

Xiao Min
ZTE Corp.
Nanjing
China

Phone: +86 25 88013062
Email: xiao.min2@zte.com.cn

Greg Mirsky
ZTE Corp.
USA

Email: gregimirsky@gmail.com

Santosh Pallagatti
VMware

Email: santosh.pallagatti@gmail.com

Jeff Tantsura
Juniper Networks

Email: jefftant.ietf@gmail.com