Workgroup: NVO3 Working Group
Internet-Draft: draft-ietf-nvo3-bfd-geneve-09
Published: 28 November 2022
Intended Status: Standards Track
Expires: 1 June 2023
Authors: X. Min
ZTE Corp.

G. Mirsky
Ericsson

S. Pallagatti
VMware

J. Tantsura
Microsoft

S. Aldrin
Google

BFD for Geneve

Abstract

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

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

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 1 June 2023.

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 the continuity of the path between two Geneve tunnel endpoints, which may be a NVE (Network Virtualization Edge) or an 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. The support for BFD Echo function is outside the scope of this document. For simplicity, NVE is used to represent the 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 mostly consistent with what's described in Section 1 and 3 of [RFC8971] ("Bidirectional Forwarding Detection (BFD) for Virtual eXtensible Local Area Network (VXLAN)"). One exception is on the usage of Management VNI, which 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

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

Since the Geneve data packet payload may be either an Ethernet frame or an IP packet, this document defines two formats of BFD packet encapsulation in Geneve. The BFD session is originated and terminated at the VAP of an NVE. The selection of the BFD packet encapsulation is based on how the VAP encapsulates the data packets. If the payload is IP, then BFD over IP is carried in the payload. If the payload is Ethernet, then BFD over IP over Ethernet is carried in the payload, in the same manner as BFD over IP in the IP payload case, regardless of what the Ethernet payload might normally carry.

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

If the VAP that originates the BFD packets is used to encapsulate Ethernet data frames, then the BFD packets are encapsulated in Geneve as described below. The Geneve packet formats over IPv4 and IPv6 are defined in Section 3.1 and 3.2 of [RFC8926] respectively. 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, an

outer IPv6 header accompanied with an inner IPv4 header and an outer IPv4 header accompanied with an inner IPv6 header are both possible.

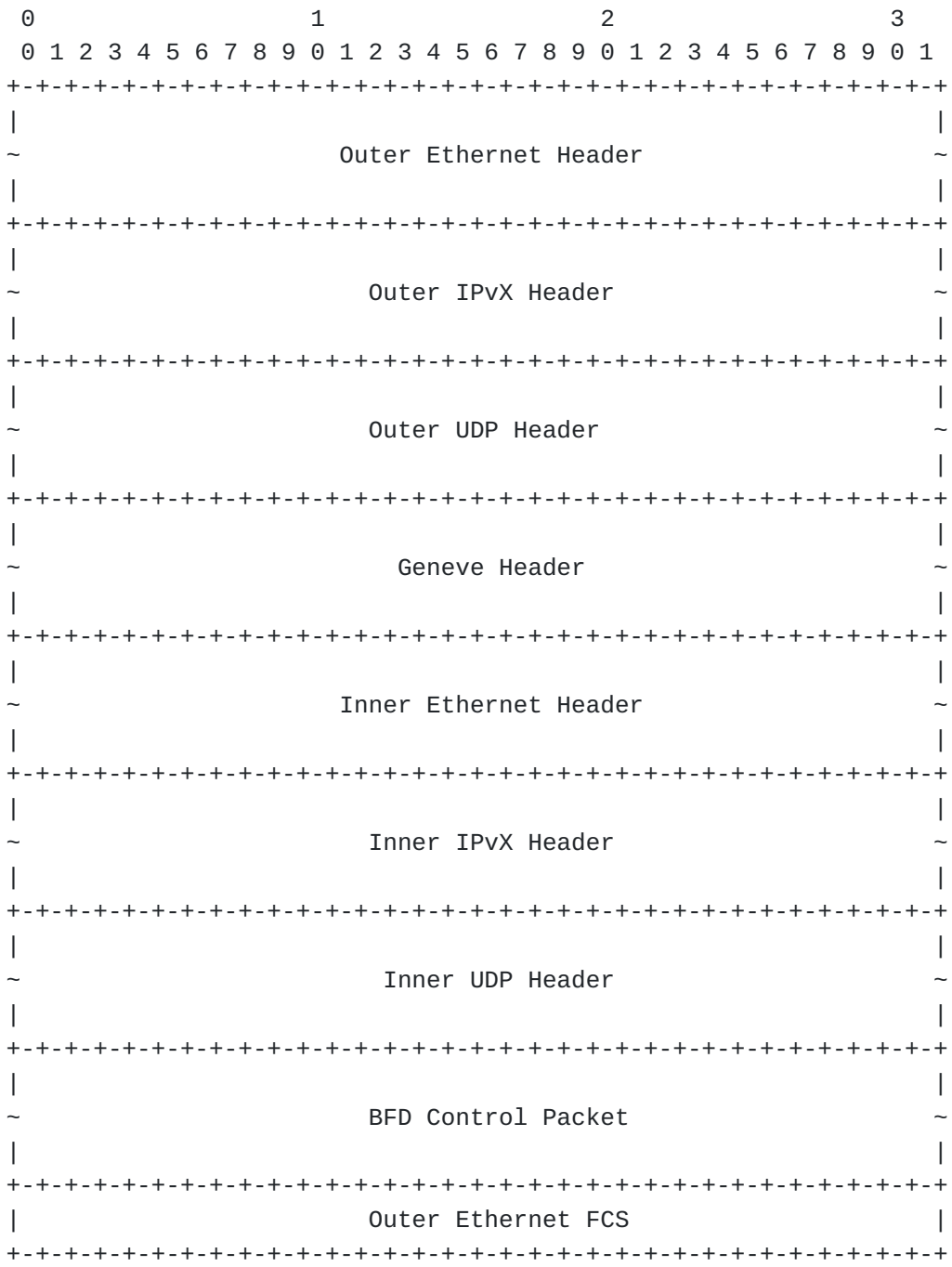

Figure 1: Geneve Encapsulation of BFD Control Packet With the Inner Ethernet/IP/UDP Header

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:

Inner 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] that specifies the IPv4/IPv6 single-hop BFD.

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 MUST be set consistent with the Geneve specification [RFC8926] depending on whether or not Geneve options are present in the frame. The use of Geneve options with BFD is beyond the scope of this document.

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. One exception is that the Management VNI is used.

4.1. Demultiplexing BFD packet when payload is Ethernet

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

Type field equals 0x6558. 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.

In BFD over Geneve, a BFD session is originated and terminated at a VAP. Usually one NVE owns multiple VAPs. Since 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, then the BFD session MUST be identified using the VNI number, and the inner Ethernet/IP/UDP header. The inner Ethernet/IP/UDP header stands for the source MAC, the source IP, the destination MAC, the destination IP, and the source UDP port number.

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.

5. BFD Encapsulation With Inner IP/UDP Header

If the VAP that originates the BFD packets is used to encapsulate IP data packets, then the BFD packets are encapsulated in Geneve as described below. The Geneve packet formats over IPv4 and IPv6 are defined in Section 3.1 and 3.2 of [RFC8926] respectively. 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, an outer IPv6 header accompanied with an inner IPv4 header and an outer IPv4 header accompanied with an inner IPv6 header are both possible.

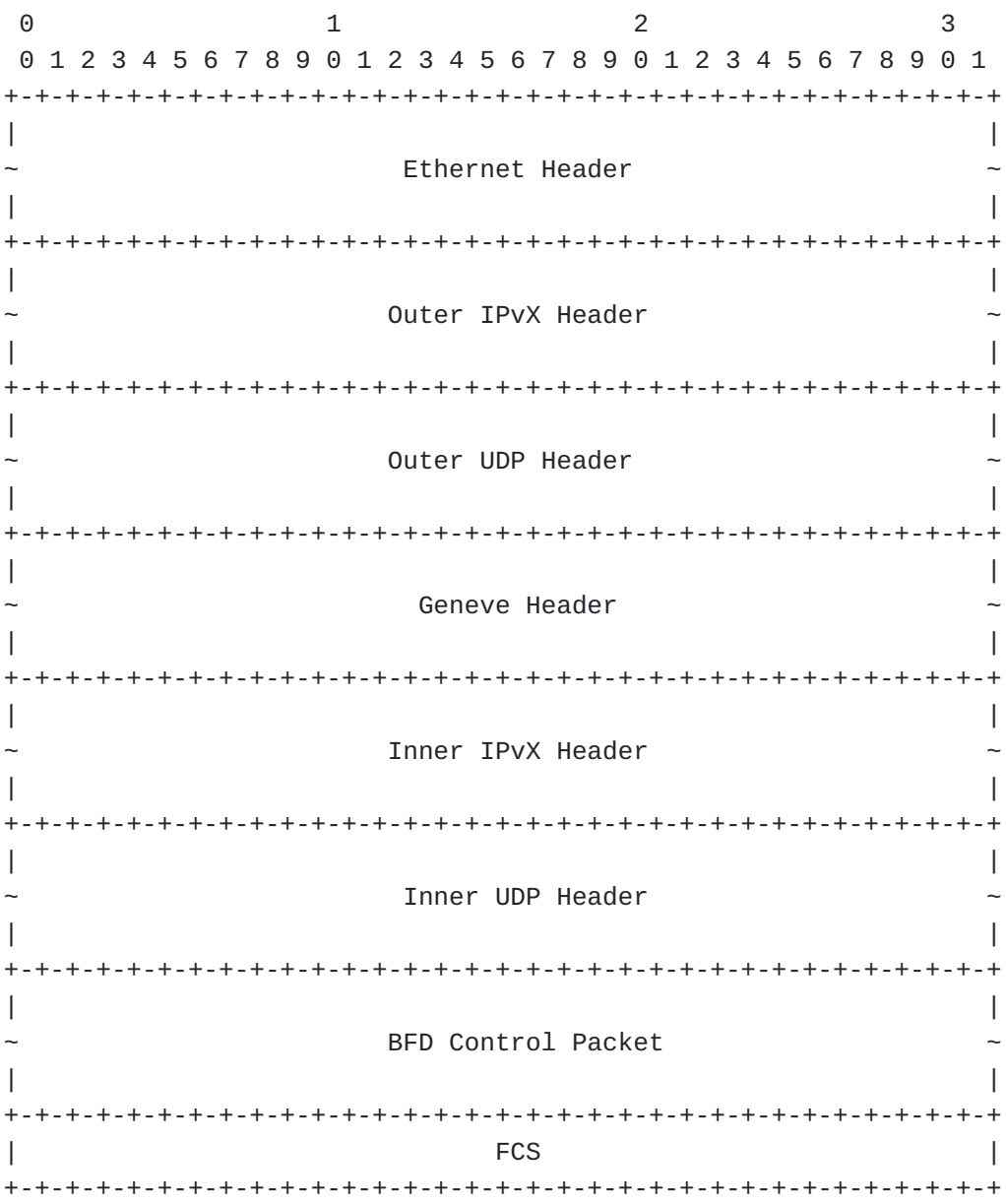

Figure 2: Geneve Encapsulation of BFD Control Packet With the Inner IP/ UDP Header

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:

Inner 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] that specifies the IPv4/IPv6 single-hop BFD.

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 MUST be set consistent with the Geneve specification [RFC8926] depending on whether or not Geneve options are present in the frame. The use of Geneve options with BFD is beyond the scope of this document.

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. One exception is that the Management VNI is used.

5.1. Demultiplexing BFD packet when payload is IP

Once a packet is received, the NVE MUST validate the packet as described in [RFC8926]. When the payload is IP, the Protocol Type field equals 0x0800 or 0x86DD. 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.

If the BFD packet is received with Your Discriminator equals to 0, then the BFD session MUST be identified using the VNI number, and the inner IP/UDP header. The inner IP/UDP header stands for the source IP, the destination IP, and the source UDP port number.

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.

6. Security Considerations

Security issues discussed in [RFC8926] apply to this document. Particularly, the BFD is an application that is run at the two Geneve tunnel endpoints. Geneve provides security between the peers and subject to the issue of overload described below. The BFD introduces no security vulnerabilities when run in this manner.

This document supports establishing multiple BFD sessions between the same pair of NVEs, each BFD session over a pair of VAPs residing in the same pair of NVEs, there SHOULD be a mechanism to control the maximum number of such sessions that can be active at the same time. Particularly, assuming an example that each NVE of the pair of NVEs has N VAPs using Ethernet as the payload, then there could be N squared BFD sessions running between the pair of NVEs. Considering N could be a not low number, the N squared BFD sessions could result in overload of the NVE. In this case, it's recommended that N BFD sessions covering all N VAPs are enough for the pair of NVEs.

7. IANA Considerations

This document has no IANA action requested.

8. 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.

The authors would like to acknowledge Stewart Bryant, Anoop Ghanwani, Jeffrey Haas, Reshad Rahman and Matthew Bocci for their thorough review and very helpful comments.

9. References

9.1. 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>.

[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>.

9.2. Informative References

[I-D.ietf-nvo3-geneve-oam] Mirsky, G., Boutros, S., Black, D. L., and S. Pallagatti, "OAM for use in GENEVE", Work in Progress, Internet-Draft, draft-ietf-nvo3-geneve-oam-05, 14 June 2022, <https://www.ietf.org/archive/id/draft-ietf-nvo3-geneve-oam-05.txt>.

[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>.

[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>.

[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
Ericsson
United States of America

Email: gregimirsky@gmail.com

Santosh Pallagatti
VMware
India

Email: santosh.pallagatti@gmail.com

Jeff Tantsura
Microsoft
United States of America

Email: jefftant.ietf@gmail.com

Sam Aldrin
Google
United States of America

Email: aldrin.ietf@gmail.com