Internet Engineering Task Force (IETF)
Request for Comments: 9521
Category: Standards Track
ISSN: 2070-1721

X. Min
ZTE Corp.
G. Mirsky
Ericsson
S. Pallagatti
VMware
J. Tantsura
Nvidia
S. Aldrin
Google

January 2024

Bidirectional Forwarding Detection (BFD) for Generic Network Virtualization Encapsulation (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 is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9521.

Table of Contents

1. Introduction
2. Conventions Used in This Document
   2.1. Abbreviations
   2.2. Requirements Language
3. BFD Packet Transmission over a Geneve Tunnel
4. BFD Encapsulation with the Inner Ethernet/IP/UDP Header
   4.1. Demultiplexing a BFD Packet When the Payload Is Ethernet
5. BFD Encapsulation with the Inner IP/UDP Header

5.1. Demultiplexing a BFD Packet When the Payload Is IP
6. Security Considerations
7. IANA Considerations
8. References
8.1. Normative References
8.2. Informative References
Acknowledgements
Authors' Addresses

1. Introduction

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

This document describes the use of the Bidirectional Forwarding Detection (BFD) protocol [RFC5880] to enable monitoring the continuity of the path between two Geneve tunnel endpoints, which may be a Network Virtualization Edge (NVE) or another device acting as a Geneve tunnel endpoint. Specifically, the asynchronous mode of BFD, as defined in [RFC5880], is used to monitor a point-to-point (P2P) Geneve tunnel. The support for the BFD Echo function is outside the scope of this document. For simplicity, an NVE is used to represent the Geneve tunnel endpoint. A Tenant System (TS) is used to represent the physical or virtual device attached to a Geneve tunnel endpoint from the outside. A Virtual Access Point (VAP) 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 Sections 1 and 3 of [RFC8971]. One exception is the usage of the Management Virtual Network Identifier (VNI), which is described in [GENEVE-OAM] and is outside the scope of this document.

As specified in Section 4.2 of [RFC8926], Geneve MUST be used with congestion controlled traffic or within a Traffic-Managed Controlled Environment (TMCE) to avoid congestion; that requirement also applies to BFD traffic. Specifically, considering the complexity and immaturity of the BFD congestion control mechanism, BFD for Geneve MUST be used within a TMCE unless BFD is really congestion controlled. As an alternative to a real congestion control, an operator of a TMCE deploying BFD for Geneve is required to provision the rates at which BFD is transmitted to avoid congestion and false failure detection.

2. Conventions Used in This Document

2.1. Abbreviations

BFD: Bidirectional Forwarding Detection

FCS: Frame Check Sequence

Geneve: Generic Network Virtualization Encapsulation

NVE: Network Virtualization Edge

TMCE: Traffic-Managed Controlled Environment

TS: Tenant System

VAP: Virtual Access Point

VNI: Virtual Network Identifier

VXLAN: Virtual eXtensible Local Area Network

2.2. Requirements Language

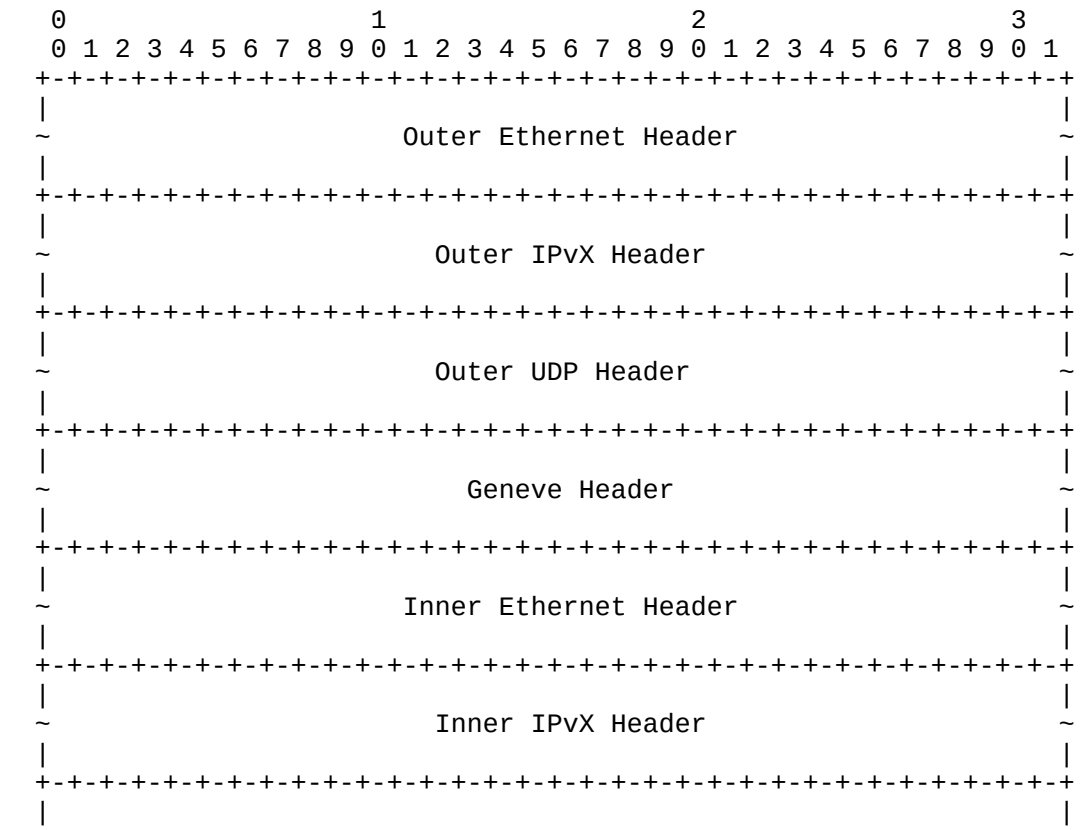
The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. BFD Packet Transmission over a 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. This occurs 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 the 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 Sections 3.1 and 3.2 of [RFC8926], respectively. The outer IP/UDP and Geneve headers are 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 by an inner IPv4 header and an outer IPv4 header accompanied by an inner IPv6 header are both possible.



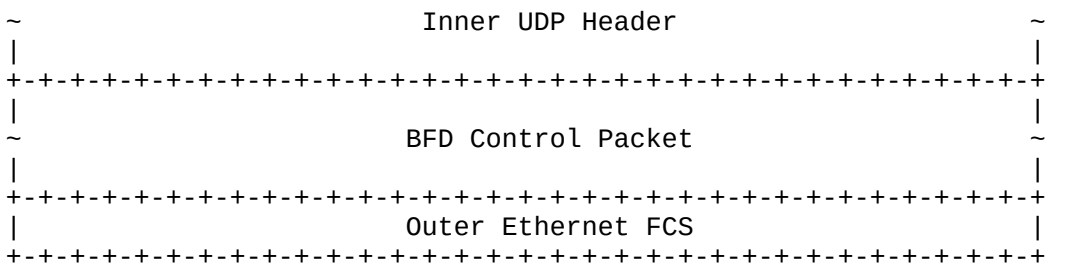


Figure 1: Geneve Encapsulation of a BFD Control Packet with the Inner Ethernet/IP/UDP Header

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:

Destination MAC: Media Access Control (MAC) address of a VAP of the terminating NVE.

Source MAC: MAC address of a VAP of the originating 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: The TTL for IPv4 or Hop Limit for IPv6 MUST be set to 255 in accordance with [RFC5881], which 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.

* The Opt Len field MUST be set as 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.

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

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

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

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

4.1. Demultiplexing a BFD Packet When the Payload Is Ethernet

Once a packet is received, the NVE validates the packet as described in [RFC8926]. When the payload is Ethernet, the Protocol Type field equals 0x6558. The destination MAC address of the inner Ethernet frame matches the MAC address of a VAP, which is mapped to the same VNI as the 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 (i.e., the three field values of the inner IP packet MUST be in compliance with what's defined in Section 4 of this document, as well as Section 4 of [RFC5881]). If the validation fails, the received packet MUST NOT be processed by BFD.

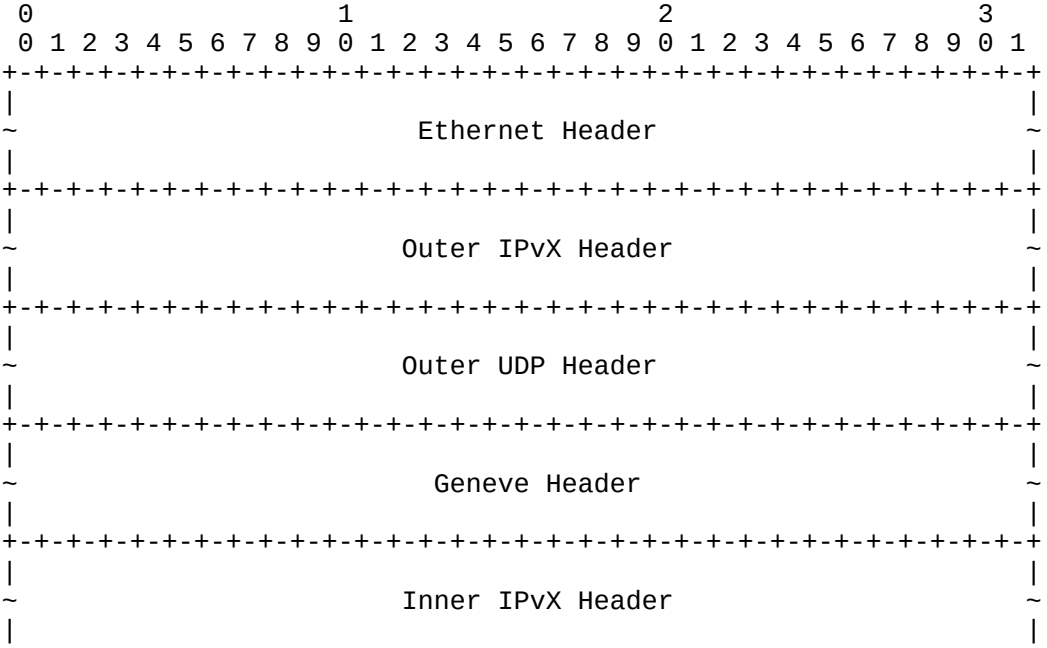
In BFD over Geneve, a BFD session is originated and terminated at a VAP. Usually one NVE owns multiple VAPs. Since multiple BFD sessions may be running between two NVEs, there needs to be a mechanism for demultiplexing received BFD packets to the proper session. Furthermore, due to the fact that [RFC8014] allows for N-to-1 mapping between VAPs and VNIs 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 that use the same way to encapsulate data packets.

If the BFD packet is received with the value of the Your Discriminator field set to 0, then the BFD session SHOULD be identified using the VNI number and the inner Ethernet/IP header. The inner Ethernet/IP header stands for the source MAC, the source IP, the destination MAC, and the destination IP. An implementation MAY use the inner UDP port source number to aid in demultiplexing incoming BFD Control packets. If it fails to identify the BFD session, the incoming BFD Control packets MUST be dropped, and an exception event indicating the failure should be reported to the management.

If the BFD packet is received with a non-zero Your Discriminator, then the BFD session MUST be demultiplexed only with the Your Discriminator as the key.

5. BFD Encapsulation with the 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 Sections 3.1 and 3.2 of [RFC8926], respectively. The outer IP/UDP and Geneve headers are 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 by an inner IPv4 header and an outer IPv4 header accompanied by an inner IPv6 header are both possible.



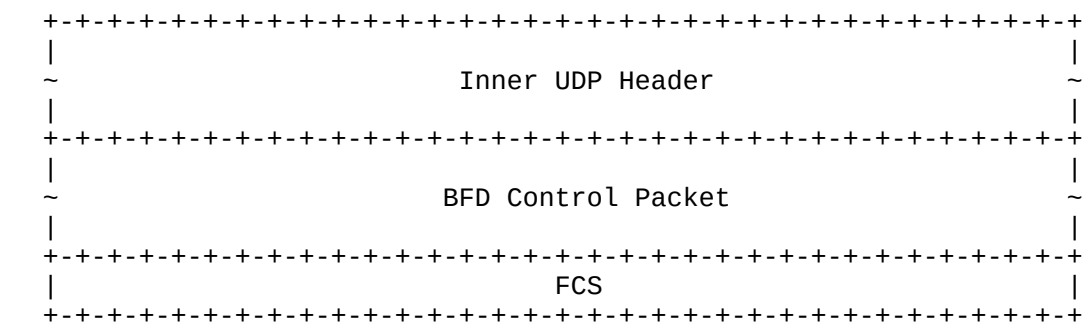


Figure 2: Geneve Encapsulation of a BFD Control Packet with the Inner IP/UDP Header

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: The TTL for IPv4 or Hop Limit for IPv6 MUST be set to 255 in accordance with [RFC5881], which 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.

* The Opt Len field MUST be set as 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.

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

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

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

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

5.1. Demultiplexing a BFD Packet When the Payload Is IP

Once a packet is received, the NVE validates the packet as described in [RFC8926]. When the payload is IP, the Protocol Type field equals 0x0800 or 0x86DD. The destination IP address of the inner IP packet matches the IP address of a VAP, which is mapped to the same VNI as the received VNI. Then, the UDP destination port and the TTL or Hop Limit of the inner IP packet MUST be validated to determine whether or not the received packet can be processed by BFD (i.e., the two field values of the inner IP packet MUST be in compliance with what's defined in Section 5 of this document as well as Section 4 of [RFC5881]). If the validation fails, the received packet MUST NOT be processed by BFD.

If the BFD packet is received with the value of the Your

Discriminator field set to 0, then the BFD session SHOULD be identified using the VNI number and the inner IP header. The inner IP header stands for the source IP and the destination IP. An implementation MAY use the inner UDP port source number to aid in demultiplexing incoming BFD Control packets. If it fails to identify the BFD session, the incoming BFD Control packets MUST be dropped, and an exception event indicating the failure should be reported to the management.

If the BFD packet is received with a non-zero Your Discriminator, then the BFD session MUST be demultiplexed only with the Your Discriminator as the key.

6. Security Considerations

Security issues discussed in [RFC8926] and [RFC5880] apply to this document. Particularly, the BFD is an application that is run at the two Geneve tunnel endpoints. The IP underlay network and/or the Geneve option can provide security between the peers, which are subject to the issue of overload described below. The BFD introduces no security vulnerabilities when run in this manner. Considering Geneve does not have any inherent security mechanisms, BFD authentication as specified in [RFC5880] is RECOMMENDED to be utilized.

This document supports establishing multiple BFD sessions between the same pair of NVEs. For 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 high 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 run for the pair of NVEs. Generally speaking, the number of BFD sessions is supposed to be enough as long as all VAPs of the pair of NVEs are covered.

7. IANA Considerations

This document has no IANA actions.

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>.

[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

[GENEVE-OAM]
Mirsky, G., Boutros, S., Black, D., and S. Pallagatti, "OAM for use in GENEVE", Work in Progress, Internet-Draft, draft-ietf-nvo3-geneve-oam-09, 6 December 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-nvo3-geneve-oam-09>.

[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>.

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 VAPs and VNIs.

The authors would like to acknowledge Stewart Bryant, Anoop Ghanwani, Jeffrey Haas, Reshad Rahman, Matthew Bocci, Andrew Alston, Magnus Westerlund, Paul Kyzivat, Sheng Jiang, Carl Wallace, Roman Danyliw, John Scudder, Donald Eastlake 3rd, ric Vyncke, Zaheduzzaman Sarker, and Lars Eggert for their thorough review and very helpful comments.

Authors’ Addresses

Xiao Min
ZTE Corp.
Nanjing
China
Phone: +86 18061680168
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

Nvidia
United States of America
Email: jefftant.ietf@gmail.com

Sam Aldrin
Google
United States of America
Email: aldrin.ietf@gmail.com