BFD Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 27, 2019

X. Min
G. Mirsky
ZTE

November 23, 2018

BFD for Geneve
draft-xiao-bfd-geneve-00

Abstract

This document describes the use of the Bidirectional Forwarding Detection (BFD) protocol in Generic Network Virtualization Encapsulation (Geneve) overlay networks.

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 May 27, 2019.

Table of Contents

1. Introduction

"Generic Network Virtualization Encapsulation" (Geneve) [I-D.ietf-nvo3-geneve] provides a generic tunneling protocol that is applicable to many scenarios, including an encapsulation scheme that allows virtual machines (VMs) to communicate in a data center network.

This document describes the use of Bidirectional Forwarding Detection (BFD) protocol for Geneve to enable monitoring continuity of the path between Network Virtualization Edges (NVEs) and/or availability of a replicator service node using BFD.

The use cases and the deployment of BFD for Geneve are consistent with what's described in Section 3 and Section 4 of [I-D.ietf-bfd-vxlan]. The main difference between Geneve and "Virtual eXtensible Local Area Network" (VXLAN) [RFC7348] encapsulation is that Geneve supports multi-protocol payload and variable length options.

1.1. Conventions Used in This Document

1.1.1. Terminology

BFD: Bidirectional Forwarding Detection

Geneve: Generic Network Virtualization Encapsulation

NVE: Network Virtualization Edge

VFI: Virtual Forwarding Instance

VM: Virtual Machine

VNI: Virtual Network Identifier

VXLAN: Virtual eXtensible Local Area Network

1.1.2. 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.

2. BFD Packet Transmission over Geneve Tunnel

BFD packet MUST be encapsulated and sent to a remote NVE using one of the options described in Section 2.1. Implementations SHOULD ensure that the BFD packets follow the same lookup path as Geneve data packets within the sender system.

2.1. BFD Packet Encapsulation in Geneve

Concerning whether or not the Geneve data packets include an IP protocol data unit, this document defines three options of BFD packet encapsulation in Geneve.

2.1.1. BFD Encapsulation With IP/UDP Header

If the Protocol Type field (as defined in Section 3.4 of [I-D.ietf-nvo3-geneve]) of data packets indicates that there exists an inner IP header, i.e., the Protocol Type equals to 0x6558 (Ethernet frame), or 0x0800 (IPv4), or 0x86DD (IPv6), or 0x8847 (MPLS), or 0x8848 (MPLS with the upstream-assigned label), then BFD packets are encapsulated in Geneve as described below. The Geneve packet format over IPv4 is defined in Section 3.1 of [I-D.ietf-nvo3-geneve]. The Geneve packet format over IPv6 is defined in Section 3.2 of [I-D.ietf-nvo3-geneve]. The Outer IP/UDP and Geneve headers MUST be encoded by the sender as defined in [I-D.ietf-nvo3-geneve]. 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.

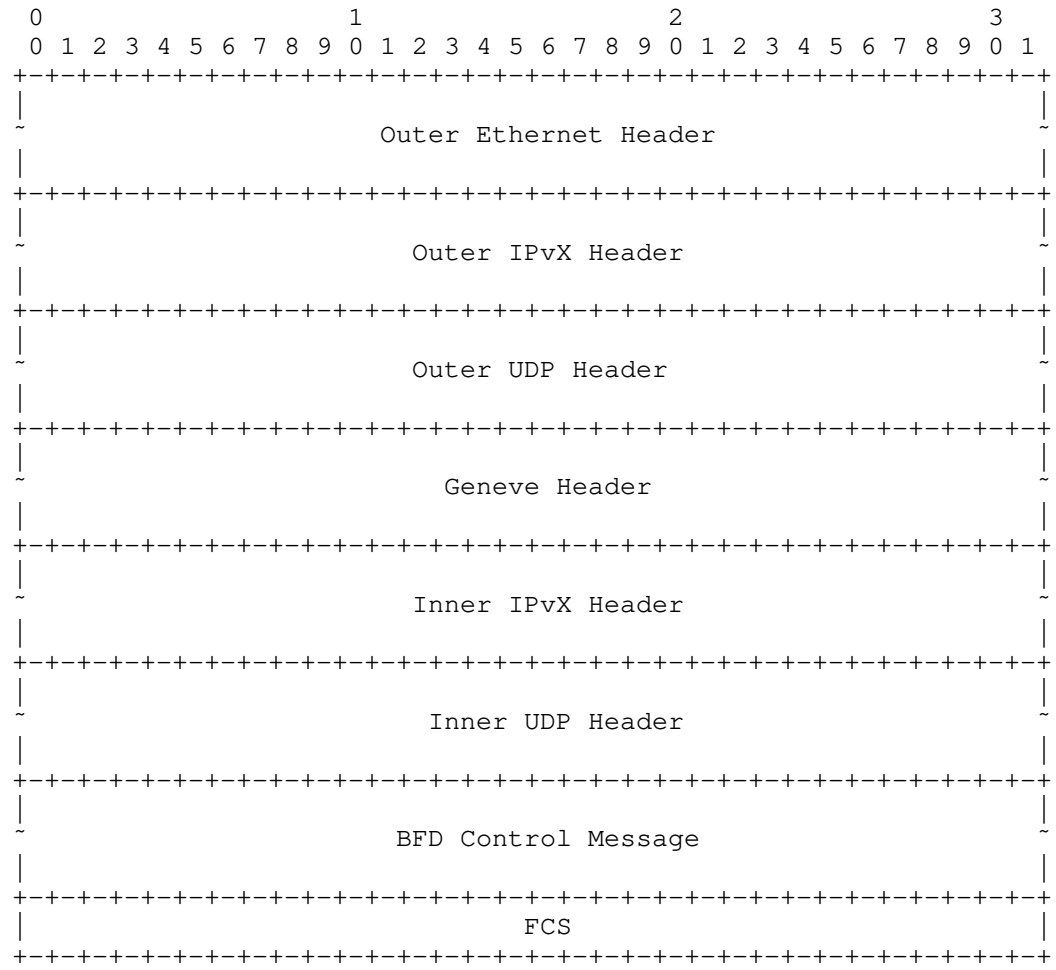

Figure 1: Geneve Encapsulation of BFD Control Message With the Inner IP/UDP Header

When the BFD packets are encapsulated in Geneve in this way, the BFD packet MUST be carried inside the inner IP packet of the Geneve packet. The inner IP packet carrying the BFD payload has the following format:

IP header:

Source IP: IP address of the originating NVE.

Destination IP: IP address of the terminating NVE.

TTL: MUST be set to 1 to ensure that the BFD packet is not routed within the L3 underlay network.

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 SHOULD follow the value set below.

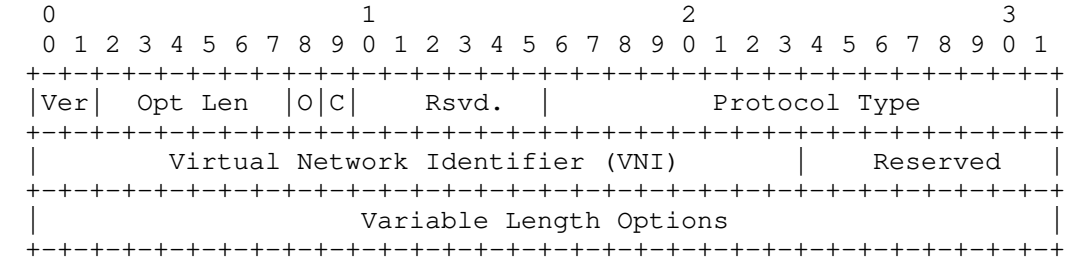

Figure 2: Geneve Header

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

[Ed.Note]: Use of O bit is still being discussed in the NVO3 WG, so the value is undetermined.

C bit SHOULD be set to 0.

Protocol Type field SHOULD be set to 0x0800 (IPv4) or 0x86DD (IPv6).

2.1.2. BFD Encapsulation Without IP/UDP Header

Alternatively to the use of the inner IP/UDP header to demultiplex BFD control packet by the value of the destination UDP port, BFD control packet MAY be encapsulated without the inner IP/UDP header. The BFD control packet MAY be identified directly in the Geneve header or through Geneve OAM shim. In either case, the Outer IP/UDP and Geneve headers MUST be encoded by the sender as defined in [I-D.ietf-nvo3-geneve].

Figure 3 displays the layout of the Ethernet frame with BFD control packet encapsulated in Geneve without the use of IP/UDP header and identified by the value TBA1 (to be assigned by IANA) of the Protocol Type field.

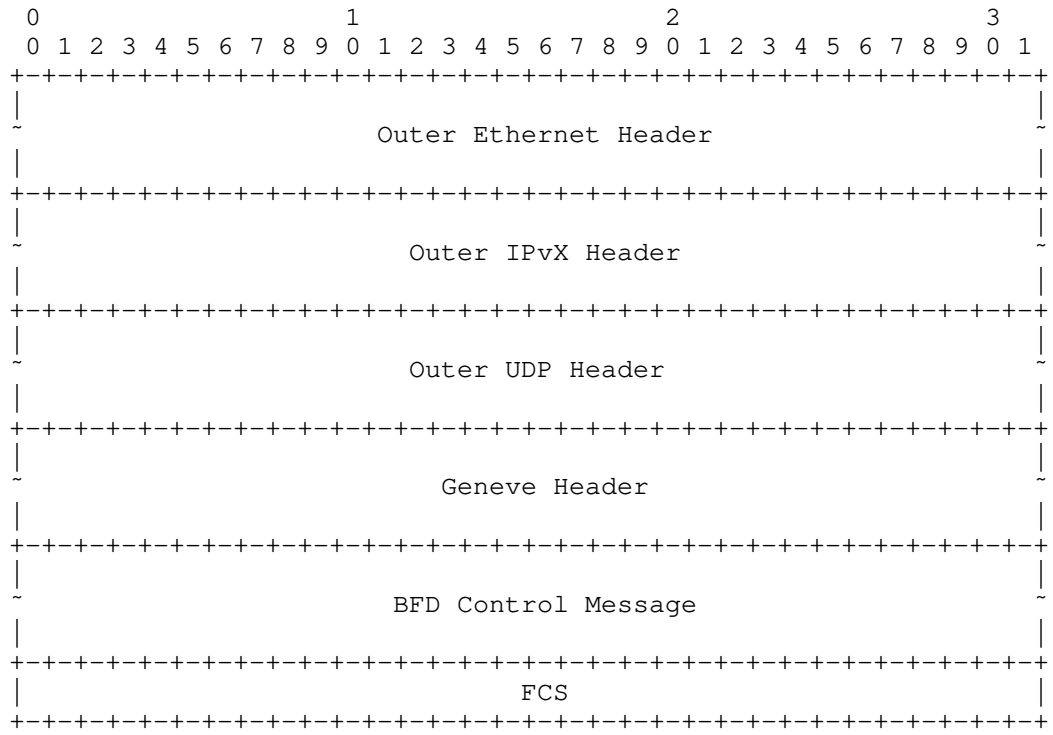

Figure 3: Geneve Encapsulation of BFD Control Message Without the Inner IP/UDP Header

When the BFD packets are encapsulated in Geneve in this way, the BFD packet MUST immediately follow the Geneve header, and the Geneve header SHOULD follow the value set below.

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

[Ed.Note]: Use of O bit is still being discussed in the NVO3 WG, so the value is undetermined.

C bit SHOULD be set to 0.

Also, if BFD control packet is encapsulated in Geneve without the use of IP/UDP header, the BFD control packet MAY be identified through the Geneve OAM shim. The layout of the Ethernet frame is shown in Figure 4. Protocol Type field MUST be set to the value TBA2 (to be assigned by IANA) which indicates a Geneve OAM shim that will have a field to indicate the inner BFD control packet. Definition of the

format of the Geneve OAM shim is outside the scope of this document. The Geneve OAM shim immediately follows the Geneve header, and the BFD control packet immediately follows the Geneve OAM shim.

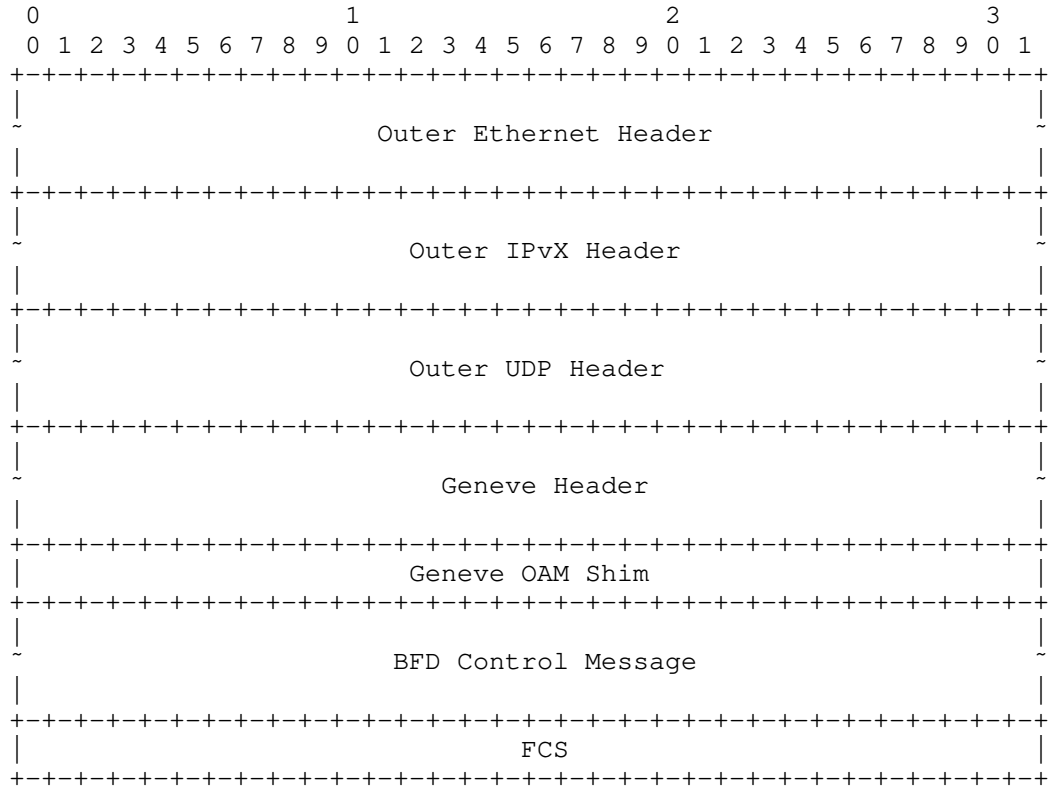

Figure 4: Geneve Encapsulation of BFD Control Message With Geneve OAM Shim

3. Reception of BFD packet from Geneve Tunnel

Once a packet is received, NVE MUST validate the packet as described in [I-D.ietf-nvo3-geneve].

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 the NVE, the UDP destination port and the TTL of the inner IP packet MUST be validated to determine whether BFD can process the received packet. BFD packet with inner IP set to NVE MUST NOT be forwarded to VMs.

If the Protocol Type field equals the value TBA1 (to be assigned by IANA) which indicates an inner BFD control message, the received packet MUST be processed by BFD and MUST NOT be forwarded to VMs.

If the Protocol Type field equals the value TBA2 (to be assigned by IANA) which indicates a Geneve OAM shim that will have a field to indicate the inner BFD control message, the received packet MUST be processed by BFD and MUST NOT be forwarded to VMs. This case is for further study.

To ensure BFD detects the proper configuration of Virtual Network Identifier (VNI) in a remote NVE, a lookup SHOULD be performed with the MAC-DA/IP-DA/MPLS-Label and VNI as key in the Virtual Forwarding Instance (VFI) table of the originating/terminating NVE to exercise the VFI associated with the VNI.

3.1. Demultiplexing of the BFD packet

If the Protocol Type field equals 0x0800 (IPv4) or 0x86DD (IPv6), demultiplexing of IP BFD packet has been defined in Section 3 of [RFC5881]. 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. The procedure for demultiplexing packets with Your Discriminator equal to 0 is different from [RFC5880]. For such packets, the BFD session MUST be identified using the inner headers, i.e., the source IP and the destination IP present in the IP header carried by the payload of the Geneve encapsulated packet. The VNI of the packet SHOULD be used to derive interface-related information for demultiplexing the packet. If BFD packet is received with non-zero Your Discriminator, then BFD session MUST be demultiplexed only with Your Discriminator as the key.

If the Protocol Type field equals the value TBA1 (to be assigned by IANA) which indicates an inner BFD control message, or the value TBA2 (to be assigned by IANA) which indicates a Geneve OAM shim that will have a field to indicate the inner BFD control message, the VNI of the packet SHOULD be used to derive interface-related information for demultiplexing the packet, demultiplexing of BFD packet MUST rely on non-zero Your Discriminator as the key.

4. Security Considerations

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

5. IANA Considerations

In the Geneve Protocol Type registry defined in [ETYPES], a new BFD Control Message or Geneve OAM Shim is requested from IANA as follows:

Geneve Protocol Type	Description	Semantics Definition	Reference
TBA1	BFD Control Message	Section 3.1	This Document
TBA2	Geneve OAM Shim	Section 3.1	This Document

Table 1: New BFD Control Message or Geneve OAM shim Ethertype

6. Acknowledgements

To be added.

7. Normative References

[ETYPES] The IEEE Registration Authority, "IEEE 802 Numbers", 2013, <http://www.iana.org/assignments/ieee-802-numbers/ieee-802-numbers.xml>.

[I-D.ietf-bfd-vxlan] Networks, J., Paragiri, S., Govindan, V., Mudigonda, M., and G. Mirsky, "BFD for VXLAN", draft-ietf-bfd-vxlan-03 (work in progress), October 2018.

[I-D.ietf-nvo3-geneve] Gross, J., Ganga, I., and T. Sridhar, "Geneve: Generic Network Virtualization Encapsulation", draft-ietf-nvo3-geneve-08 (work in progress), October 2018.

[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>.

[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>.

[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>.

Authors' Addresses

Xiao Min
ZTE
Nanjing
China

Phone: +86 25 88016574
Email: xiao.min2@zte.com.cn

Greg Mirsky
ZTE
USA

Email: gregimirsky@gmail.com