NVO3 Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 29, 2021

G. Mirsky
ZTE Corp.
S. Boutros
Ciena
D. Black
Dell EMC
S. Pallagatti
VMware

September 25, 2020

OAM for use in GENEVE
draft-mmbb-nvo3-geneve-oam-04

Abstract

This document lists a set of general requirements for active OAM protocols in the Geneve overlay network. Based on the requirements, IP encapsulation for active Operations, Administration, and Maintenance protocols in Geneve protocol is defined. Considerations for using ICMP and UDP-based protocols are discussed.

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 March 29, 2021.

Copyright Notice

Table of Contents

1. Introduction

Geneve [I-D.ietf-nvo3-geneve] is intended to support various scenarios of network virtualization. In addition to carrying multi-protocol payload, e.g., Ethernet, IPv4/IPv6, the Geneve message includes metadata. Operations, Administration, and Maintenance (OAM) protocols support fault management and performance monitoring functions necessary for comprehensive network operation. Active OAM protocols, as defined in [RFC7799], use specially constructed packets that are injected into the network. To ensure that the measured performance metric or the detected failure of the transport layer are related to a particular Geneve flow, it is critical that these test packets share fate with overlay data packets for that flow when traversing the underlay network.

A set of general requirements for active OAM protocols in the Geneve overlay network is listed in Section 2. IP encapsulation conforms to these requirements and is defined as a suitable encapsulation of active OAM protocols in a Geneve overlay network. Note that the IP encapsulation of OAM is applicable to those VNIs that support the use of the necessary values of the Protocol Type field in the Geneve

header, i.e., Ethertypes of IPv4 or IPv6. It does not apply to VNIs that lack that support, e.g., VNIs that only support Ethernet Ethertypes. Analysis and definition of other types of OAM encapsulation in Geneve are outside the scope of this document.

1.1. Conventions used in this document

1.1.1. Acronyms

CC Continuity Check

CV Connectivity Verification

FM Fault Management

Geneve Generic Network Virtualization Encapsulation

NVO3 Network Virtualization Overlays

OAM Operations, Administration, and Maintenance

VNI Virtual Network Identifier

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. Active OAM Protocols in Geneve Networks

OAM protocols, whether part of fault management or performance monitoring, are intended to provide reliable information that can be used to detect a failure, identify the defect and localize it, thus helping to identify and apply corrective actions to minimize the negative impact on service. Several OAM protocols are used to perform these functions; these protocols require demultiplexing at the receiving instance of Geneve. To improve the accuracy of the correlation between the condition experienced by the monitored Geneve tunnel and the state of the OAM protocol the OAM encapsulation is required to comply with the following requirements:

REQ#1: Geneve OAM test packets MUST share the fate with data traffic of the monitored Geneve tunnel, i.e., be in-band with the monitored traffic, follow the same overlay and transport path as

packets with data payload, in the forward direction, i.e. from ingress toward egress endpoint(s) of the OAM test.

An OAM protocol MAY be used to monitor the particular Geneve tunnel as a whole. In that case, test packets could be fate-sharing with a sub-set of tenant flows transported over the Geneve tunnel. If the goal is to monitor the condition experienced by the flow of a particular tenant, the test packets MUST be fate-sharing with that specific flow in the Geneve tunnel. In the latter case, the test packet MUST use the same Geneve encapsulation as the data packet (except for the value in the Protocol Type field [I-D.ietf-nvo3-geneve]), including the value in the Virtual Network Identifier (VNI) field. Both scenarios are discussed in detail in Section 2.1.

REQ#2: Encapsulation of OAM control message and data packets in underlay network MUST be indistinguishable from the underlay network IP forwarding point of view.

REQ#3: Presence of OAM control message in Geneve packet MUST be unambiguously identifiable to Geneve functionality, e.g., at endpoints of Geneve tunnels.

REQ#4: OAM test packets MUST NOT be forwaded to a tenant system.

A test packet generated by an active OAM protocol, either for a defect detection or performance measurement, according to REQ#1, MUST be fate-sharing with the tunnel or data flow being monitored. In an environment where multiple paths through the domain are available, underlay transport nodes can be programmed to use characteristic information to balance the load across known paths. It is essential that test packets follow the same route, i.e., traverses the same set of nodes and links, as a data packet of the monitored flow. Thus, the following requirement to support OAM packet fate-sharing with the data flow:

REQ#5: It MUST be possible to express entropy for underlay Equal Cost Multipath in the Geneve encapsulation of OAM packets.

2.1. Defect Detection and Troubleshooting in Geneve Network with Active OAM

Figure 1 presents an example of a Geneve domain. In this section, we consider two scenarios of active OAM being used to detect and localize defects in the Geneve network.

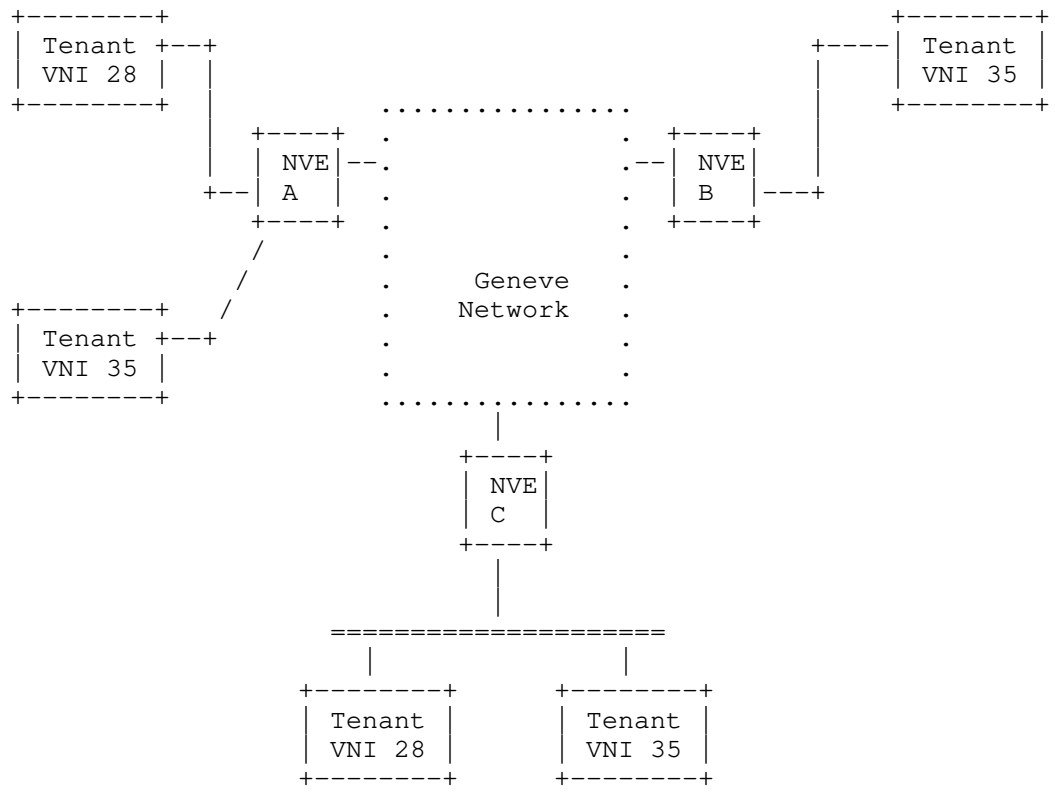

Figure 1: An example of a Geneve domain

In the first case, a communication problem between Network Virtualization Edge (NVE) device A and NVE C was observed. The underlay, e.g., IP network, forwarding is working well but the Geneve connection is unstable for all tenants of NVE A and NVE C. Troubleshooting and localization of the problem can be done irrespective of the VNI value.

In the second case, traffic on VNI 35 between NVE A and NVE B has no problems, as on VNI 28 between NVE A and NVE C. But traffic on VNI 35 between NVE A and NVE C experiences problems, for example, excessive packet loss.

The first case can be detected and investigated using any VNI value, whether it connects tenant systems or not. To conform to REQ#4 (Section 2) OAM test packets could be transmitted on VNI that doesn't have any tenant. That VNI in a Geneve tunnel is dedicated to carrying only control and management data between the tunnel

endpoints, hence communication that uses that VNI is also and referred to as a Geneve control channel. Thus, the control channel of a Geneve tunnel MUST NOT carry tenant data. As no tenants are connected using the control channel, a system that supports this specification, MUST NOT forward a packet received over the control channel to any tenant. A packet received over the control channel MAY be forwarded if and only if it is sent onto the control channel of the concatenated Geneve tunnel. A specific VNI MAY be used to identify the control channel. The value that is associated with this function is referred to as Management VNI. It is RECOMMENDED that the value 1 be used as the default value of Management VNI. Encapsulation of test packets, in this case, is discussed in Section 2.2. The Management VNI SHOULD be terminated on the tenant-facing side of the Geneve encap/decap functionality, not the DC-network-facing side (per definitions in Section 4 [RFC8014]) so that Geneve encap/decap functionality is included in its scope. This approach causes an active OAM packet, e.g., an ICMP echo request, to be decapsulated in the same fashion as any other received Geneve packet. In this example, the resulting ICMP packet is handed to NVE's local management functionality for the processing which generates an ICMP echo reply. The ICMP echo reply is encapsulated in Geneve for forwarding back to the NVE that sent the echo request. One advantage of this approach is that a repeated ping test could detect an intermittent problem in Geneve encap/decap hardware, which would not be tested if the Management VNI were handled as a "special case" at the DC-network-facing interface.

The second case requires that a test packet be transmitted using the VNI value for the traffic that is encountering problems. The encapsulation of the test packet, i.e., inner encapsulation, MUST be the same as the tenant packet's encapsulation in the Geneve tunnel and use the same VNI number. Encapsulation of test packets in this case, is discussed in Section 2.2.

2.2. OAM Encapsulation in Geneve

Active OAM in Geneve network uses an IP encapsulation. Protocols such as BFD [RFC5880] or STAMP [RFC8762] use UDP transport. Destination UDP port number in the inner UDP header (Figure 2) identifies the OAM protocol. This approach is well-known and has been used, for example, in MPLS networks [RFC8029]. The UDP source port can be used to provide entropy, e.g., for Equal Cost Multipath. To use IP encapsulation for an active OAM protocol the Protocol Type field of the Geneve header MUST be set to the IPv4 (0x0800) or IPv6 (0x86DD) value.

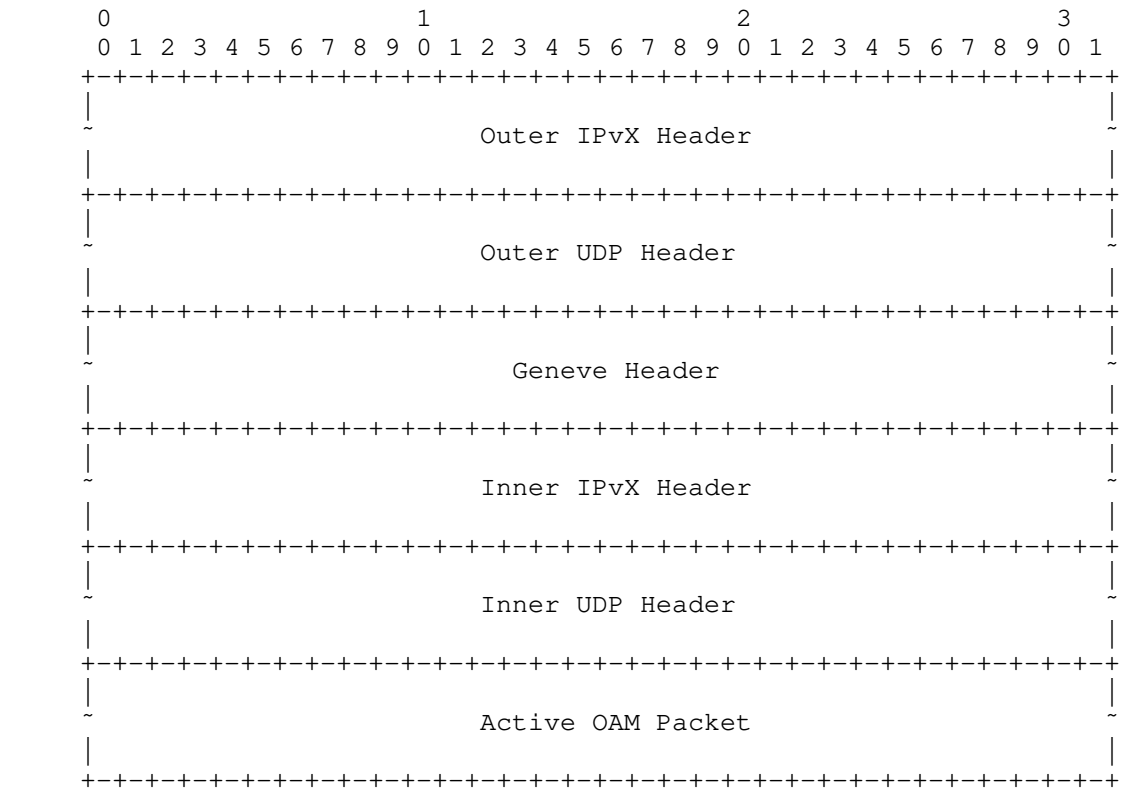

Figure 2: Geneve IP/UDP Encapsulation of an Active OAM Packet

Inner IP header:

Destination IP: IP address MUST NOT be of one of tenant's IP addresses. The IP address SHOULD be set to the loopback address 127.0.0.1/32 for IPv4, or the loopback address ::1/128 for IPv6 [RFC4291]. Alternatively, the destination IP address MAY be set to VAP's IP address.

Source IP: IP address of the originating VAP.

TTL or Hop Limit: MUST be set to 255 per [RFC5082].

3. Echo Request and Echo Reply in Geneve Tunnel

ICMP and ICMPv6 ([RFC0792] and [RFC4443] respectively) provide required on-demand defect detection and failure localization. ICMP control messages immediately follow the inner IP header encapsulated

in Geneve. ICMP extensions for Geneve networks use mechanisms defined in [RFC4884].

4. IANA Considerations

This document has no requirements for IANA. This section can be removed before the publication.

5. Security Considerations

As part of a Geneve network, active OAM inherits the security considerations discussed in [I-D.ietf-nvo3-geneve]. Additionally, a system MUST provide control to limit the rate of Geneve OAM packets punted to the Geneve control plane for processing in order to avoid overloading that control plane.

6. Acknowledgments

TBD

7. References

7.1. Normative References

[I-D.ietf-nvo3-geneve]
Gross, J., Ganga, I., and T. Sridhar, "Geneve: Generic Network Virtualization Encapsulation", draft-ietf-nvo3-geneve-16 (work in progress), March 2020.

[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>.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, DOI 10.17487/RFC4385, February 2006, <https://www.rfc-editor.org/info/rfc4385>.

[RFC5586] Bocci, M., Ed., Vigoureux, M., Ed., and S. Bryant, Ed., "MPLS Generic Associated Channel", RFC 5586, DOI 10.17487/RFC5586, June 2009, <https://www.rfc-editor.org/info/rfc5586>.

[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>.

7.2. Informative References

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, DOI 10.17487/RFC0792, September 1981, <https://www.rfc-editor.org/info/rfc792>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <https://www.rfc-editor.org/info/rfc4291>.

[RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", STD 89, RFC 4443, DOI 10.17487/RFC4443, March 2006, <https://www.rfc-editor.org/info/rfc4443>.

[RFC4884] Bonica, R., Gan, D., Tappan, D., and C. Pignataro, "Extended ICMP to Support Multi-Part Messages", RFC 4884, DOI 10.17487/RFC4884, April 2007, <https://www.rfc-editor.org/info/rfc4884>.

[RFC5082] Gill, V., Heasley, J., Meyer, D., Savola, P., Ed., and C. Pignataro, "The Generalized TTL Security Mechanism (GTSM)", RFC 5082, DOI 10.17487/RFC5082, October 2007, <https://www.rfc-editor.org/info/rfc5082>.

[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>.

[RFC7799] Morton, A., "Active and Passive Metrics and Methods (with Hybrid Types In-Between)", RFC 7799, DOI 10.17487/RFC7799, May 2016, <https://www.rfc-editor.org/info/rfc7799>.

[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>.

[RFC8029] Kompella, K., Swallow, G., Pignataro, C., Ed., Kumar, N., Aldrin, S., and M. Chen, "Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures", RFC 8029, DOI 10.17487/RFC8029, March 2017, <https://www.rfc-editor.org/info/rfc8029>.

[RFC8762] Mirsky, G., Jun, G., Nydell, H., and R. Foote, "Simple Two-Way Active Measurement Protocol", RFC 8762, DOI 10.17487/RFC8762, March 2020, <https://www.rfc-editor.org/info/rfc8762>.

Appendix A. Additional Considerations for OAM Encapsulation Method in Geneve

Several other options of OAM encapsulation were considered. Those are listed in the Appendix solely for the informational purpose.

A Protocol Type field might be used to demultiplex active OAM protocols directly. Such method avoids the use of additional intermediate header but requires that each active OAM protocol be assigned unique identifier from the Ether Types registry maintained by IANA.

The alternative to using the Protocol Type directly is to use a shim that, in turn, identifies the OAM Protocol and, optionally, includes additional information. [RFC5586] defines how the Generic Associated Channel Label (GAL) can be used to identify that the Associated Channel Header (ACH), defined in [RFC4385], immediately follows the Bottom-of-the-Stack label. Thus, the MPLS Generic Associated Channel can be identified, and protocols are demultiplexed based on the Channel Type field's value. Number of channel types, e.g., for continuity check and performance monitoring, already have been defined and are listed in IANA MPLS Generalized Associated Channel Types (including Pseudowire Associated Channel Types) registry. The value of the Protocol Type field in the Geneve header MUST be set to MPLS to use this approach. The Geneve header MUST be immediately followed by the GAL label with the S flag set to indicate that GAL is the Bottom-of-the-stack label. Then ACH MUST follow the GAL label and the value of the Channel Type identifies which of active OAM protocols being encapsulated in the packet.

Authors' Addresses

Greg Mirsky
ZTE Corp.

Email: gregimirsky@gmail.com

Sami Boutros
Ciena

Email: sboutros@ciena.com

David Black
Dell EMC
176 South Street
Hopkinton, MA 01748
United States of America

Email: david.black@dell.com

Santosh Pallagatti
VMware

Email: santosh.pallagatti@gmail.com