INTERNET-DRAFT
Intended Status: Proposed Standard

D. Eastlake
W. Hao
S. Zhuang
Z. Li
Huawei Technologies
R. Gu
China Mobil

Expires: September 3, 2019
March 4, 2019

BGP Dissemination of Network Virtualization Overlays (NVO3) Flow Specification Rules

<draft-ietf-idr-flowspec-nvo3-04.txt>

Abstract

This draft specifies a new subset of component types to support the (Network Virtualization Overlays (NVO3)) flow-spec application.

Status of This Document

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

Distribution of this document is unlimited. Comments should be sent to the authors or the TRILL Working Group mailing list <dnsext@ietf.org>.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html. The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Table of Contents

1. Introduction

BGP Flow-spec is an extension to BGP that supports the dissemination of traffic flow specification rules. It uses the BGP Control Plane to simplify the distribution of Access Control Lists (ACLs) and allows new filter rules to be injected to all BGP peers simultaneously without changing router configuration. A typical application of BGP Flow-spec is to automate the distribution of traffic filter lists to routers for Distributed Denial of Service (DDOS) mitigation.

[RFC5575] defines a new BGP Network Layer Reachability Information (NLRI) format used to distribute traffic flow specification rules. NLRI (AFI=1, SAFI=133) is for IPv4 unicast filtering. NLRI (AFI=1, SAFI=134) is for BGP/MPLS VPN filtering. [IPv6-FlowSpec] and [Layer2-FlowSpec] extend the flow-spec rules for IPv6 and layer 2 Ethernet packets respectively. All these previous flow specifications match only single layer IP/Ethernet information fields like source/destination MAC, source/destination IP prefix, protocol type, ports, and the like.

In the cloud computing era, multi-tenancy has become a core requirement for data centers. Since Network Virtualization Overlays (NVO3) can satisfy multi-tenancy key requirements, this technology is being deployed in an increasing number of cloud data center networks. NVO3 is an overlay technology and VXLAN [RFC7348] and NVGRE [RFC7367] are two typical NVO3 encapsulations. GENEVE [GENEVE], GUE [GUE] and GPE [GPE] are three emerging NVO3 encapsulations. Because it is an overlay technology involving an additional level of encapsulation, flow specification matching on the inner header as well as the outer header, as specified below, is needed.

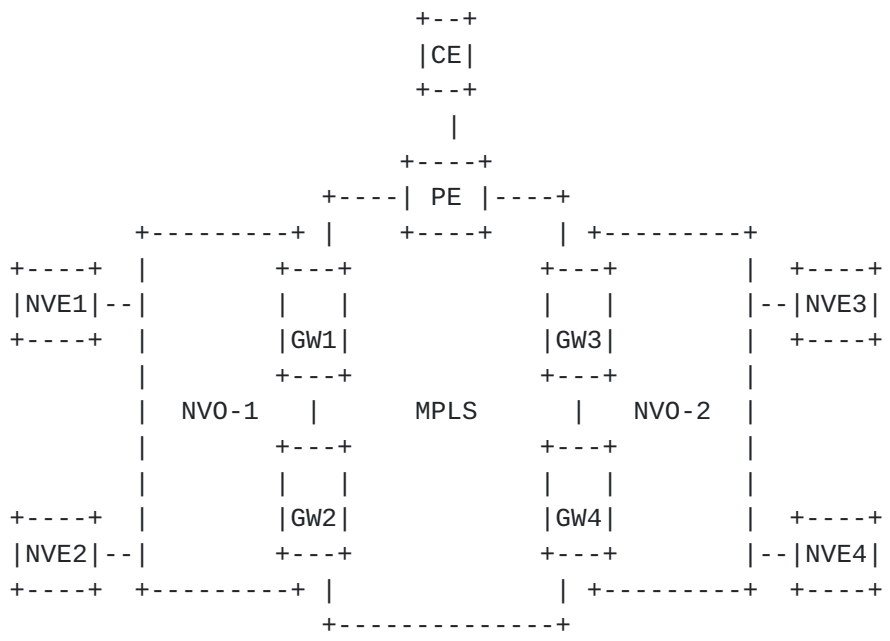


Figure 1. NVO3 Data Center Interconnection

The MPLS L2/L3 VPN in the WAN network can be used for NVO3 based data center network interconnection. When the Data Center (DC) and the WAN are operated by the same administrative entity, the Service Provider can decide to integrate the gateway (GW) and WAN Edge PE functions in the same router for capital and operational cost saving reasons. This is illustrated in Figure 1. There are two interconnection solutions as follows:

1. End-to-end NVO3 tunnel across different data centers: NVE1 performs NVO3 encapsulation for DC interconnection with NVE3. The destination VTEP IP is NVE3's IP. The GW doesn't perform NVO3 tunnel termination. The DC interconnect WAN is pure an underlay network.

2. Segmented NVO3 tunnels across different data centers: NVE1 doesn't perform end-to-end NVO3 encapsulation to NVE3 for DC interconnection. The GW performs NVO3 tunnel encapsulation termination, and then transmits the inner original traffic through an MPLS network to the peer data center GW. The peer data center GW again terminates MPLS encapsulation, and then performs NVO3 encapsulation to transmit the traffic to the local NVE3.

In the first solution, to differentiate bandwidth and Quality of Service (QoS) among different tenants or applications, different TE tunnels in the WAN network will be used to carry the end-to-end NVO3 encapsulation traffic using VN ID, NVO3 outer header DSCP, and other fields as the traffic classification match part. The BGP Flow-spec

protocol can be used to set the traffic classification on all GWs simultaneously.

In the second solution, a centralized BGP speaker can be deployed for DDOS mitigation in the WAN network. When the analyzer detects abnormal traffic, it will automatically generate Flow-spec rules and distribute them to each GW through the BGP Flow-spec protocol, the match part should include matching on inner or outer L2/L3 layer or NVO3 headers.

In summary, the Flow specification match part on the GW/PE should be able to include inner layer 2 Ethernet header, inner layer 3 IP header, outer layer 2 Ethernet header, outer layer 3 IP header, and/or NVO3 header information. Because the current flow-spec matching facilities lack a layer indicator and NVO3 header information, those facilities can't be used directly for traffic filtering based on NVO3 headers or on a specified layer header directly. This draft specifies a new subset of component types to support the NVO3 flow-spec application.

1.1 Terminology

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.

The reader is assumed to be familiar with BGP and NVO3 terminology. The following terms and acronyms are used in this document with the meaning indicated:

ACL - Access Control List

DC - Data Center

DDOS - Distributed Denial of Service (Attack)

GW - gateway

VN - virtual network

VTEP - Virtual Tunnel End Point

WAN - wide area network

2. NVO3 Flow Specification Encoding

The current Flow-spec rules can only recognize flows based on the outer layer header of NVO3 encapsulation data packets. To enable traffic filtering based on an NVO3 header and on an inner header of NVO3 packets, a new component type acting as a delimiter is introduced. The delimiter type is used to indicate the boundary between the inner and outer layer component types for NVO3 data packets. All the component types defined in [RFC5575], [IPv6-FlowSpec], [Layer2-FlowSpec], and the like can be used for the inner or outer header as indicated by the use of delimiters.

Because the NVO3 outer layer address normally belongs to a public network, the "Flow Specification" NLRI for the outer layer header doesn't need to include a Route Distinguisher field (8 bytes). If the outer layer address belongs to a VPN, the NLRI format for the outer header should consist of a fixed-length Route Distinguisher field (8 bytes) corresponding to the VPN. This Route Distinguisher is followed by the detail flow specifications for the outer layer.

The VN ID is the identification for each tenant network. The "Flow Specification" NLRI for an NVO3 header part should always include the VN ID field but a Route Distinguisher field does not need to be included.

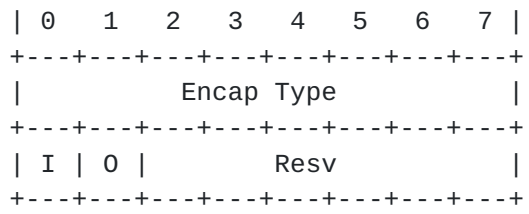
The inner layer MAC/IP address is always associated with a VN ID. Thus the NLRI format for the inner header should consist of a fixed-length VN ID field (4 bytes). The VN ID is followed by the detailed flow specifications for the inner layer. The NLRI length field shall include both the 4 bytes of the VN ID as well as the subsequent flow specification. In the NVO3 terminating into a VPN scenario, if multiple access VN IDs map to one VPN instance, one shared VN ID can be carried in the Flow-Spec rule to enforce the rule on the entire VPN instance and the shared VN ID and VPN correspondence should be configured on each VPN PE beforehand. In this case, the function of the layer3 VN ID is the same as a Route Distinguisher: it acts as the identification of the VPN instance.

This document specifies the following Flow-Spec Component Types for use with NVO3 flows:

Type TBD1 - Delimiter type
   Encoding: <type (1 octet), length (1 octet), Value>.

      When this delimiter type is present, it indicates the component types and layer for the NVO3 header fields immediately following. At the same time, it indicates the end of the component types belonging to the previous delimiter.

The value field defines encapsulation type and is encoded as:



This document defines the following Encap types:

- VXLAN: Tunnel Type = 0

- NVGRE: Tunnel Type = 1

I: If I is set to one, it indicates the component types for the inner layer of NVO3 headers immediately follow.

O: If O is set to one, it indicates the component types for the outer layer of NVO3 headers immediately follow.

For the NVO3 header part, the following additional component types are introduced.

Type TBD2 - VN ID

Encoding: <type (1 octet), [op, value]+>.

Defines a list of {operation, value} pairs used to match the 24-bit VN ID that is used as the tenant identification in NVO3 networks. For NVGRE encapsulation, the VN ID is equivalent to VSID. Values are encoded as 1- to 3-byte quantities.

Type TBD3 - Flow ID

Encoding: <type (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match 8-bit Flow ID fields which are only useful for NVGRE encapsulation. Values are encoded as 1-byte quantity.

3. NVO3 Flow Specification Traffic Actions

The current traffic filtering actions are used for NVO3 encapsulation traffic. For Traffic Marking, only the DSCP in the outer header can be modified.

4. Security Considerations

No new security issues are introduced to the BGP protocol by this specification.

5. IANA Considerations

IANA is requested to assign three new values in the "Flow Spec Component Types" registry as follows:

Type	Name	Reference
TBD1	Delimiter type	[this document]
TBD2	VN ID	[this document]
TBD3	Flow ID	[this document]

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>.

[RFC5575] - Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <https://www.rfc-editor.org/info/rfc5575>.

[RFC8174] - [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>.

[GENEVE] - J. Gross, T. Sridhar, etc, "Geneve: Generic Network Virtualization Encapsulation", draft-ietf-nvo3-geneve, work in progress.

[GUE] - T. Herbert, L. Yong, O. Zia, "Generic UDP Encapsulation", draft-ietf-nvo3-gue, work in progress.

Informative References

[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>.

[RFC7367] - Garg, P., Ed., and Y. Wang, Ed., "NVGRE: Network Virtualization Using Generic Routing Encapsulation", RFC 7637, DOI 10.17487/RFC7637, September 2015, <https://www.rfc-editor.org/info/rfc7637>.

[IPv6-FlowSpec] - R. Raszuk, etc, "Dissemination of Flow Specification Rules for IPv6", draft-ietf-idr-flow-spec-v6, work in progress.

[Layer2-FlowSpec] - W. Hao, etc, "Dissemination of Flow Specification Rules for L2 VPN", draft-ietf-idr-flowspec-l2vpn, work in progress.

[GPE] - P. Quinn, etc, "Generic Protocol Extension for VXLAN", draft-ietf-nvo3-vxlan-gpe, work in progress.

Acknowledgments

The authors wish to acknowledge the important contributions of Jeff Haas, Susan Hares, Qiandeng Liang, Nan Wu, Yizhou Li, and Lucy Yong.

Authors' Addresses

Donald Eastlake
Huawei Technologies
1424 Pro Shop Court
Davenport, FL 33896 USA

Tel: +1-508-333-2270
Email: d3e3e3@gmail.com

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012 China

Email: haoweiguo@huawei.com

Shunwan Zhuang
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095 China

Email: zhuangshunwan@huawei.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095 China

Email: lizhenbin@huawei.com

Rong Gu
China Mobile

Email: gurong_cmcc@outlook.com

Copyright, Disclaimer, and Additional IPR Provisions