INTERNET-DRAFT
Intended Status: Proposed Standard

W. Hao
Huawei Technologies
D. Eastlake
Futurewei Technologies
J. Uttaro
AT&T
S. Litkowski
Cisco Systems
S. Zhuang
Huawei Technologies

Expires: May 2, 2020
November 3, 2019

BGP Dissemination of L2VPN Flow Specification Rules
draft-ietf-idr-flowspec-l2vpn-12

Abstract

This document defines a Border Gateway Protocol (BGP) Flow-spec extension to disseminate Layer 2 Virtual Private Network (L2VPN) Ethernet traffic filtering rules. AFI=25 SAFI=134 is used for this purpose. New component types and an extended community also are defined.

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 IDR Working Group mailing list <idr@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.....3
1.1 Terminology.....4

2. Layer 2 Flow Specification Encoding in BGP.....5

3. L2VPN Flow Specification Encoding in BGP.....6
3.1 Order of Traffic Filtering Rules.....8

4. Ethernet Flow Specification Traffic Actions.....10
4.1 VLAN-action.....10
4.2 TPID-action.....12

5. Flow Spec Validation.....13

6. IANA Considerations.....14
7. Security Considerations.....15

8. Acknowledgements.....15
9. Contributors.....15

Normative References.....16
Informative References.....16

Authors' Addresses.....17

1. Introduction

Border Gateway Protocol (BGP) Flow-spec [RFC5575bis] is an extension to BGP that supports the dissemination of traffic flow specification rules and actions to be taken on packets in a specified flow. It leverages the BGP Control Plane to simplify the distribution of ACLs (Access Control Lists). Using the Flow-spec extension new filter rules can be injected to all BGP peers simultaneously without changing router configuration. The typical application is to automate the distribution of traffic filter lists to routers for DDOS (Distributed Denial of Service) mitigation, access control, etc.

BGP Flow-spec [RFC5575bis] defines a 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 IPv4 BGP/MPLS VPN filtering. The Flow specification match part defined in [RFC5575bis] only includes L3/L4 information like IPv4 source/destination prefix, protocol, ports, and the like, so traffic flows can only be filtered based on L3/L4 information. This has been extended by [FlowSpecV6] to cover IPv6.

Layer 2 Virtual Private Networks (L2VPNs) have been deployed in an increasing number of networks. Such networks also have requirements to deploy BGP Flow-spec to mitigate DDoS attack traffic. Within an L2VPN network, both IP and non-IP Ethernet traffic maybe exist. For IP traffic filtering, the Flow specification rules defined in [RFC5575bis] and/or [FlowSpecV6], which include match criteria and actions, can still be used. Flow specification rules received via the new NLRI format apply only to traffic that belongs to the VPN instance(s) in which it is imported. For non-IP Ethernet traffic filtering, Layer 2 related information like source/destination MAC and VLAN need to be considered.

There are different kinds of L2VPN networks like EVPN [RFC7432], BGP VPLS [RFC4761], LDP VPLS [RFC4762] and border gateway protocol (BGP) auto discovery [RFC6074]. Because the Flow-spec feature relies on the BGP protocol to distribute traffic filtering rules, it can only be incrementally deployed in those L2VPN networks where BGP has already been used for auto discovery and/or signaling purposes such as BGP- based VPLS [RFC4761], EVPN and LDP-based VPLS [RFC4762] with BGP auto-discovery [RFC6074].

This draft defines new Flow-spec component types and two new extended communities to support L2VPN Flow-spec application. The Flow-spec rules can be enforced on all border routers or on some interface sets of the border routers. SAFI=134 in [RFC5575bis] and [FlowSpecV6] is extended for AFI=25 as specified in Section 2 to cover traffic

filtering information in an L2VPN environment.

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 following acronyms are used in this document:

ACL - Access Control List

DDOS - Distributed Denial of Service

EVPN - Ethernet VPN [RFC7432]

L2VPN - Layer 2 VPN

L3VPN - Layer 3 VPN

PCP - Priority Code Point [802.1Q]

TPID - Tag Protocol ID, typically a VLAN ID

VLAN = Virtual Local Area Network

VPLS - Virtual Private Line Service [RFC4762]

VPN - Virtual Private Network

2. Layer 2 Flow Specification Encoding in BGP

[RFC5575bis] defines SAFI 133 and SAFI 134, with AFI=1, for "dissemination of IPv4 flow specification rules" and "dissemination of VPNv4 flow specification rules", respectively. [FlowSpecV6]] extends [RFC5575bis] to also allow AFI=2 thus making it applicable to both IPv4 and IPv6 applications. This document further extends the SAFI=134 for AFI=25 and make it applicable to L2VPN applications.

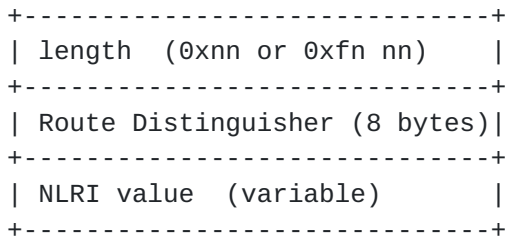
The following change is specified:

> "SAFI 134 for dissemination of L3VPN flow specification rules" in [FlowSpecV6] is defined as "SAFI 134 for dissemination of VPN flow specification rules"

The address family to which SAFI 134 refers is indicate by the AFI value (AFI=1 for VPNv4, AFI=2 VPNv6 and AFI=25 for L2VPN). Such extension is fully backwards compatible with existing implementation and production deployments.

3. L2VPN Flow Specification Encoding in BGP

The NLRI format for AFI=25/SAFI=134, as with the other VPN Flow-spec AFI/SAFI pairs, consists of an overall length encoded as provided in Section 4.1 of [RFC5575bis], then a fixed-length Route Distinguisher field (8 octets), then a flow specification [RFC5575bis] [FlowSpecV6] that may include the component types defined in this document. The length field includes both the 8 octets of the Route Distinguisher as well as the subsequent flow specification.



Flow specification rules received via this NLRI apply only to traffic that belongs to the VPN instance(s) into which it is imported. Flow rules are accepted as specified in Section 5.

Besides the component types defined in [RFC5575bis] and [FlowSpecV6], this document specifies the following additional component types for L2 VPN Ethernet traffic filtering:

Type tbdA - Ethernet Type (EtherType)
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match two-octet field. op is encoded as specified in Section 4.2.3 of [RFC5575bis]. Values are encoded as 2-octet quantities. Ethernet II framing defines the two-octet Ethernet Type (EtherType) field in an Ethernet frame, preceded by destination and source MAC addresses, that identifies an upper layer protocol encapsulating the frame data.

Type tbdB - Source MAC
Encoding: <type (1 octet), MAC Address length (1 octet), MAC Address>

Defines the source MAC Address to match.

Type tbdC - Destination MAC
Encoding: <type (1 octet), MAC Address length (1 octet), MAC Address>

Defines the destination MAC Address to match.

Type tbdD - DSAP (Destination Service Access Point) in LLC
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match the 1-octet DSAP in the 802.2 LLC (Logical Link Control Header). Values are encoded as 1-octet quantities. op is encoded as specified in Section 4.2.3 of [RFC5575bis].

Type tbdE - SSAP (Source Service Access Point) in LLC
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match the 1-octet SSAP in the 802.2 LLC. Values are encoded as 1-octet quantities. op is encoded as specified in Section 4.2.3 of [RFC5575bis].

Type tbdF - Control field in LLC
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match 1-octet control field in the 802.2 LLC. Values are encoded as 1-octet quantities. op is encoded as specified in Section 4.2.3 of [RFC5575bis].

Type tbdG - SNAP
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match 5-octet SNAP (Sub-Network Access Protocol) field. Values are encoded as 5-octet quantities. op is encoded as specified in Section 4.2.3 of [RFC5575bis].

Type tbdH - VLAN ID
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match VLAN ID. Values are encoded as 2-octet quantities, where the four most significant bits are zero and the 12 least significant bits contain the VLAN value. op is encoded as specified in Section 4.2.3 of [RFC5575bis].

In the virtual local-area network (VLAN) stacking case, the VLAN ID is the outer VLAN ID.

Type tbdI - VLAN PCP
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match 3-bit VLAN PCP fields [802.1Q]. Values are encoded using a single octet, where the five most significant bits are zero and the three least significant bits contain the VLAN PCP value. op is encoded as specified in Section 4.2.3 of [RFC5575bis].

In the virtual local-area network (VLAN) stacking case, the VLAN PCP is outer VLAN PCP.

Type tbdJ - Inner VLAN ID
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match the inner VLAN ID using for virtual local-area network (VLAN) stacking or Q in Q use. Values are encoded as 2-octet quantities, where the four most significant bits are zero and the 12 least significant bits contain the VLAN value. op is encoded as specified in Section 4.2.3 of [RFC5575bis].

In single VLAN case, this component type MUST NOT be used. If it appears the match will fail.

Type tbdK - Inner VLAN PCP
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match 3-bit inner VLAN PCP fields [802.1Q] using for virtual local-area network (VLAN) stacking or Q in Q use. Values are encoded using a single octet, where the five most significant bits are zero and the three least significant bits contain the VLAN PCP value. op is encoded as specified in Section 4.2.3 of [RFC5575bis].

In single VLAN case, the component type MUST NOT be used. If it appears the match will fail.

Type tbdL - VLAN DEI
Encoding: <type (1 octet), length (1 octet), op (1 octet)>

This type tests the DEI bit in the VLAN tag. If op is zero, it matches if and only if the DEI bit is zero. If op is non-zero, it matches if and only if the DEI bit is one.

Type tbdM - Inner VLAN DEI
Encoding: <type (1 octet), length (1 octet), op (1 octet)>

This type tests the DEI bit in the inner VLAN tag. If op is zero, it matches if and only if the DEI bit is zero. If op is non-zero, it matches if and only if the DEI bit is one.

3.1 Order of Traffic Filtering Rules

The original definition for the order of traffic filtering rules can be reused with new consideration for the MAC Address offset. As long

as the offsets are equal, the comparison is the same, retaining

longest-prefix-match semantics. If the offsets are not equal, the lowest offset has precedence, as this flow matches the most significant bit.

Pseudocode:

```
flow_rule_L2_cmp (a, b)
{
    comp1 = next_component(a);
    comp2 = next_component(b);
    while (comp1 || comp2) {
        // component_type returns infinity on end-of-list
        if (component_type(comp1) < component_type(comp2)) {
            return A_HAS_PRECEDENCE;
        }
        if (component_type(comp1) > component_type(comp2)) {
            return B_HAS_PRECEDENCE;
        }

        if (component_type(comp1) == MAC_DESTINATION || MAC_SOURCE) {
            common = MIN(MAC Address length (comp1),
                     MAC Address length (comp2));
            cmp = MAC Address compare(comp1, comp2, common);
            // not equal, lowest value has precedence
            // equal, longest match has precedence
        } else {
            common =
               MIN(component_length(comp1), component_length(comp2));
            cmp = memcmp(data(comp1), data(comp2), common);
            // not equal, lowest value has precedence
            // equal, longest string has precedence
        }
    }
    return EQUAL;
}
```

4. Ethernet Flow Specification Traffic Actions

The default action for a layer 2 traffic filtering flow specification is to accept traffic that matches that particular rule. The following extended community values per [RFC5575bis] can be used to specify particular actions in an L2 VPN network:

type	extended community	encoding
0x8006	traffic-rate	2-octet as#, 4-octet float
0x8007	traffic-action	bitmask
0x8008	redirect	6-octet Route Target
0x8009	traffic-marking	DSCP value

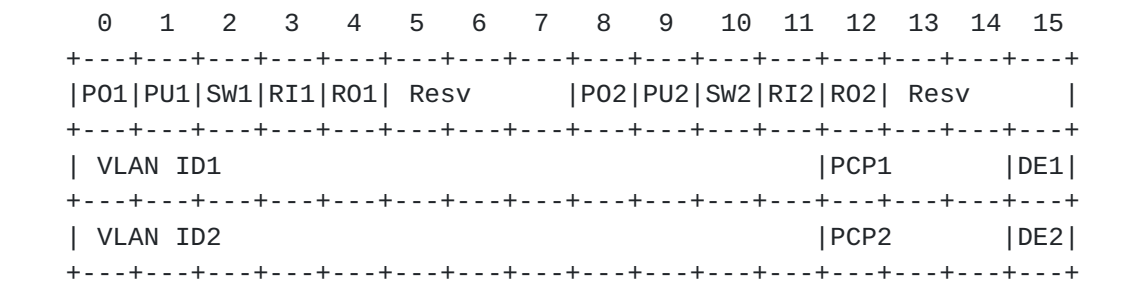
Redirect: The action should be redefined to allow the traffic to be redirected to a MAC or IP VRF routing instance that lists the specified route-target in its import policy.

Besides the above extended communities, this document also specifies the following BGP extended communities for Ethernet flows to extend [RFC5575bis]:

type	extended community	encoding
TBD1	VLAN-action	bitmask
TBD2	TPID-action	bitmask

4.1 VLAN-action

The VLAN-action extended community, as shown in the diagram below, consists of 6 octets that include t action Flags, two VLAN IDs, and the associated PCP and DEI values. The action Flags fields are further divided into two parts which correspond to the first action and the second action respectively. Bit 0 to bit 7 give the first action while bit 8 to bit 15 give the second action. The bits of PO, PU, SW, RI and RO in each part represent the action of Pop, Push, Swap, Rewrite inner VLAN and Rewrite outer VLAN respectively. Through this method, more complicated actions also can be represented in a single VLAN-action extended community, such as SwapPop, PushSwap, etc. For example, SwapPop action is the sequence of two actions, the first action is Swap and the second action is Pop.



PO1: Pop action. If the PO1 flag is one, it indicates the outmost VLAN should be removed.

PU1: Push action. If PU1 is one, it indicates VLAN ID1 will be added, the associated PCP and DEI are PCP1 and DE1.

SW1: Swap action. If the SW1 flag is one, it indicates the outer VLAN and inner VLAN should be swapped.

PO2: Pop action. If the PO2 flag is one, it indicates the outmost VLAN should be removed.

PU2: Push action. If PU2 is one, it indicates VLAN ID2 will be added, the associated PCP and DEI are PCP2 and DE2.

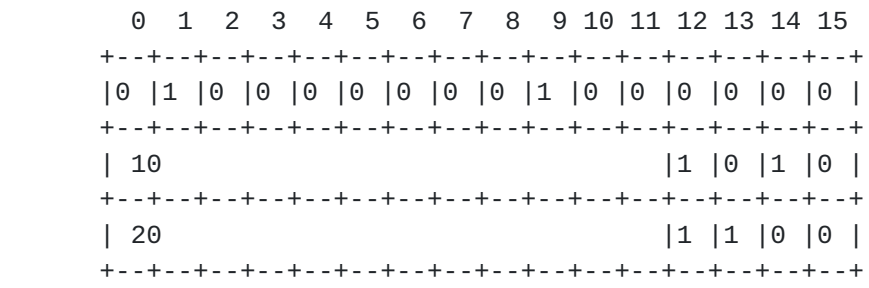
SW2: Swap action. If the SW2 flag is one, it indicates the outer VLAN and inner VLAN should be swapped.

RI1 and RI2: Rewrite inner VLAN action. If the RI flag is one, it indicates the inner VLAN should be replaced by a new VLAN where the new VLAN is VLAN ID1 and the associated PCP and DEO are PCP1 and DE1. If the VLAN ID1 is 0, the action is to only modify the PCP and DEI value of the inner VLAN.

RO1 and RO2: Rewrite outer VLAN action. If the RO flag is one, it indicates the outer VLAN should be replaced by a new VLAN where the new VLAN is VLAN ID and the associated PCP and DEI are PCP2 and DE2. If the VLAN ID2 is 0, the action is to only modify the PCP and DEI value of the outer VLAN.

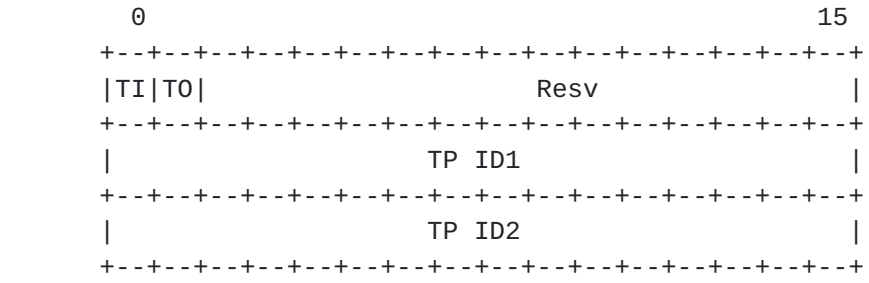
Resv, R1, and R2: Reserved for future use. MUST be sent as zero and ignored on receipt.

Giving an example below: if the action of PUSH Inner VLAN 10 with PCP value 5 DEI value 0 and Outer VLAN 20 with PCP value 6 DEI value 0 is needed, the format of the VLAN-action extended community is as follows:



4.2 TPID-action

The TPID-action extended community consists of 6 octets which includes the fields of action Flags, TPID1 and TPID2.



TI: Mapping inner TP ID action. If the TI flag is one, it indicates the inner TP ID should be replaced by a new TP ID, the new TP ID is TP ID1.

TO: Mapping outer TP ID action. If the TO flag is one, it indicates the outer TP ID should be replaced by a new TP ID, the new TP ID is TP ID2.

Resv: Reserved for future use. MUST be sent as zero and ignored on receipt.

5. Flow Spec Validation

Flow-specs received over AFI=25/SAFI=134 are validated against routing reachability received over AFI=25/SAFI=128 as modified to conform to [FlowSpecOID].

6. IANA Considerations

IANA is requested to change the description for SAFI 134 [RFC5575bis] to read as follows and to change the reference for it to [this document]:

134 VPN dissemination of flow specification rules

IANA is requested to allocate 11 new values in the Flow-Spec Component Type registry as follows:

type	Reference	description
tbdA	[this document]	Ethernet Type
tbdB	[this document]	Source MAC
tbdC	[this document]	Destination MAC
tbdD	[this document]	DSAP in LLC
tbdE	[this document]	SSAP in LLC
tbdF	[this document]	Control field in LLC
tbdG	[this document]	SNAP
tbdH	[this document]	VLAN ID
tbdI	[this document]	VLAN PCP
tbdJ	[this document]	Inner VLAN ID
tbdK	[this document]	Inner VLAN PCP
tbdL	[this document]	VLAN DEI
tbdM	[this document]	Inner VLAN DEI

IANA is requested to assign two values from the "BGP Extended Communities Type - extended, transitive" registry [suggested value provided in square brackets]:

Type value	Name	Reference
TBD1[0x080A]	Flow spec VLAN action	[this document]
TBD2[0x080B]	Flow spec TPID action	[this document]

7. Security Considerations

For General BGP Flow-spec Security Considerations, see [RFC5575bis].

VLAN tagging identifies Layer 2 communities which are commonly expected to be isolated except when higher layer connection is provided, such as Layer 3 routing. The ability of the Flow-spec VLAN action to change the VLAN ID in a frame thus may compromise security.

8. Acknowledgements

The authors wish to acknowledge the important contributions of the following:

Hannes Gredler, Xiaohu Xu, Zhenbin Li, Lucy Yong, and Feng Dong.

9. Contributors

Qiandeng Liang
Huawei Technologies
101 Software Avenue, Yuhuatai District
Nanjing 210012
China

Email: liangqiandeng@huawei.com

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>.

[RFC4761] Kompella, K., Ed. and Y. Rekhter, Ed., "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, DOI 10.17487/RFC4761, January 2007, <https://www.rfc-editor.org/info/rfc4761>.

[RFC4762] Lasserre, M., Ed. and V. Kompella, Ed., "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, DOI 10.17487/RFC4762, January 2007, <https://www.rfc-editor.org/info/rfc4762>.

[RFC6074] Rosen, E., Davie, B., Radoaca, V., and W. Luo, "Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)", RFC 6074, DOI 10.17487/RFC6074, January 2011, <https://www.rfc-editor.org/info/rfc6074>.

[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>.

[FlowSpecOID] Uttaro, J., Alcaide, J., Filsfils, C.. Smith, D., Mohapatra, P., draft-ietf-idr-bgp-flowspec-oid, work in progress.

[FlowSpecV6] McPherson, D., Raszuk, R., Pithawala, B., akarch@cisco.com, a., and S. Hares, "Dissemination of Flow Specification Rules for IPv6", draft-ietf-idr-flow-spec-v6-09 (work in progress), November 2017.

[RFC5575bis] Hares, S., Loibl, C., Raszuk, R., McPherson, D., Bacher, M., "Dissemination of Flow Specification Rules", draft-ietf-idr-rfc5575bis-17, Work in progress, June 2019.

Informative References

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

Authors' Addresses

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China

Email: haoweiguo@huawei.com

Donald E. Eastlake, 3rd
Futurewei Technologies
2386 Panoramic Circle
Apopka, FL 32703
USA

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

James Uttaro
AT&T

Email: uttaro@att.com

Stephane Litkowski
Cisco Systems, Inc.

Email: slitkows.ietf@gmail.com

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

Email: zhuangshunwan@huawei.com