IDR Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 27, 2022

S. Hares
Hickory Hill Consulting
D. Eastlake
Futurewei

July 26, 2021

BGP Flow Specification Version 2
draft-hares-idr-flowspec-v2-02

Abstract

BGP flow specification version 1 (RFC8955, RFC8956) describes the distribution of traffic filter policy (traffic filters and actions) which are distributed via BGP to BGP peers. Multiple applications utilize the BGP distributed traffic filter policy. These applications include: (1) mitigation of Denial of Service (DoS), (2) enabling of traffic filtering in BGP/MPLS VPNS, and(3)centralized traffic control for networks utilizing either SDN control of router firewall functions. During the deployment of BGP flow specification v1, the following issues were detected: 1) problems due to the lack of clear TLV encoding for rules for flow specifications, 2) desire to order filters rules, and 3) ordering of actions to provide deterministic actions. Version 2 of the BGP flow specification protocol addresses these features.

BGP Flow Specification v2 is encapsulated in a different NLRI which encapsulates previous flow specification information.

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 January 27, 2022.

Table of Contents

1. Introduction

BGP ([RFC4271]) flow specification (see [RFC8955] and [RFC8956]) describes the distribution of traffic filter policy (traffic filters and actions) which are distributed via BGP to BGP peers. The traffic filter policy is applied when packets are received on a router with the flow specification function turned on. Multiple applications utilize the BGP distributed traffic filter policy. These applications include: (1) mitigation of Denial of Service (DoS), (2) enabling of traffic filtering in BGP/MPLS VPNS, and (3)centralized traffic control for networks utilizing either SDN control of router firewall functions. During the deployment of BGP flow specification v1, the following issues were detected:

- problems such as non-extensibility due to the lack of clear TLV encoding,
- desire to order filtering rules, and
- desire to order actions to provide deterministic interactions of actions.

Version 2 of the BGP flow specification protocol addresses these features.

This document specifies six new BGP Flow Specification NLRI wit (3 AFIs (1, 2, and 6) with two SAFI (TBD1 and TBD2) that allow user-ordered list of traffic match filters, and user-ordered traffic match actions encoded in BGP Wide Communities. These NLRIs provide encoding for both the ordinary and the VPN cases for IPv4, IPv6, and layer 2 covering all the cases covered by [RFC8955], [RFC8956], [I-D.ietf-idr-flowspec-l2vpn], and [I-D.ietf-idr-flowspec-nvo3]. This document provides an overview in this section and the following other sections provide additional detail:

- section 2 - Definitions,
- section 3 - Rules for dissemination of Flow Specification v2,
- section 4 - Optional Security,
- section 5 - IANA considerations,
- section 6 - security considerations.

This section reviews the existing flow specification and provides a logical description of ordered flow specification.

1.1. Flow Specification v1 Review

If one considers the reception of the packet as an event, then BGP flow specification describes a set of Event-MatchCondition-Action (ECA) policies where the match-condition is defined in the BGP NLRI, and the action is defined either by the default condition (accept traffic) or actions defined in Extended BGP Community values [RFC4360].

The initial set of conditions [RFC8955] and [RFC8956] for this policy includes 13 types of match filters encoded the following: specific AFI/SAFIs for the IPv4 and IPv6 AFIs:

IPv4 traffic: AFI:1, SAFI:133;

IPv6 Traffic: AFI:2 SAFI:133

BGP/MPLS IPv4 VPN: AFI:1, SAFI: 134

BGP/MPLS IPv6 VPN: AFI:2, SAFI: 134

The 13 types of filters are the following:

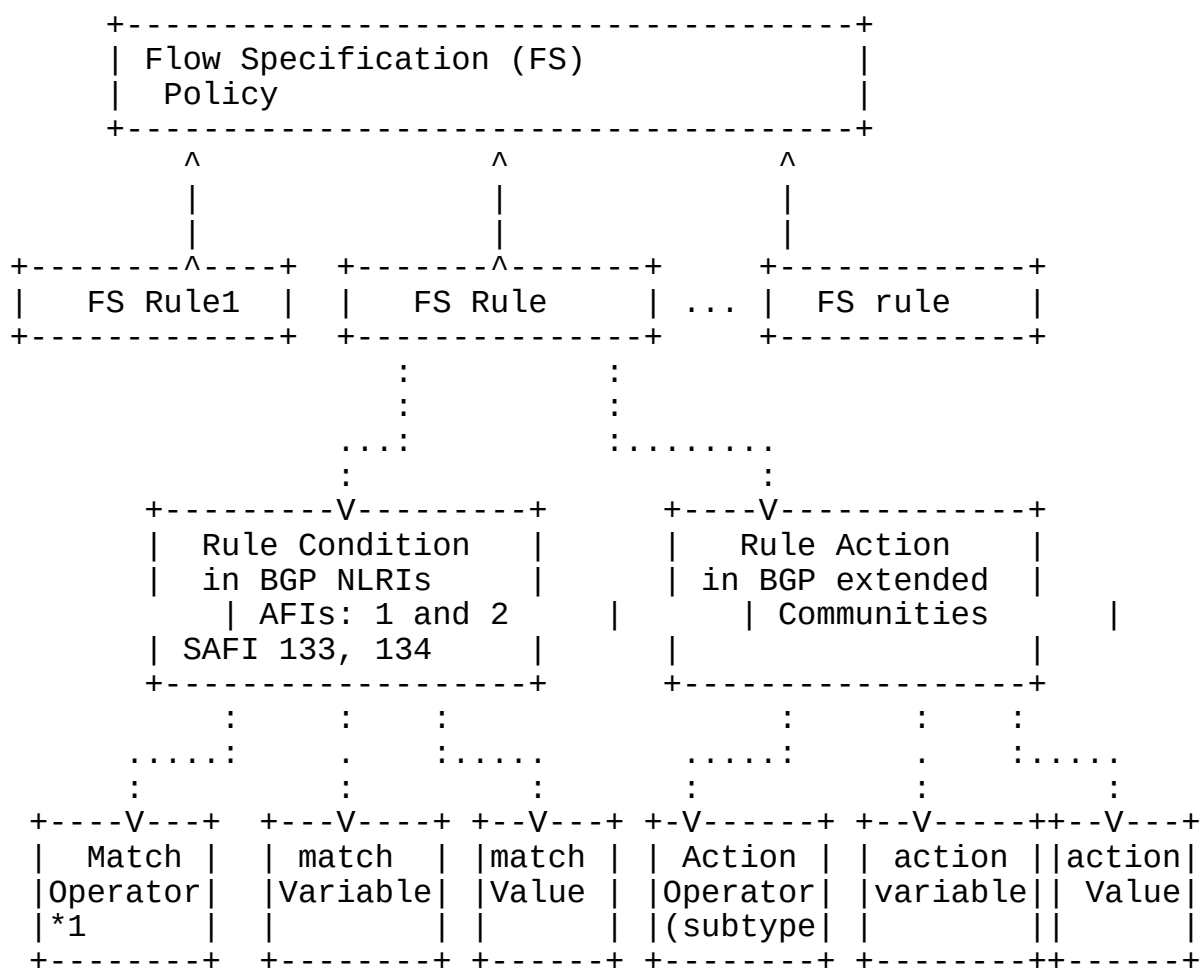
- Type 1: Destination Prefix
- Type 2: Source Prefix
- Type 3: IP Protocol (v4,[RFC8955]) or Upper Layer Protocol (v6, [RFC8956])
- Type 4: Port
- Type 5: Destination Port
- Type 6: Source Port
- Type 7: ICMPv4 Type (v4,[RFC8955]) or ICMPv6 Type (v6, [RFC8956])
- Type 8: ICMPv4 Code (v4,[RFC8955]) or ICMPv6 code(v6, [RFC8956])
- Type 9: TCP flags (v4,[RFC8955])
- Type 10: Packet length
- Type 11: DSCP marking

- Type 12: Fragment
- Type 13: Flow Label (v6, [RFC8956])

The actions specified [RFC8955] and [RFC8956] for exclusion on Extended Community (0xttss) are the following:

- Traffic rate limited by bytes (0x8006) [2 byte AS, 4 byte float]
- Traffic action (set by bitmask, bits 47 and 46 defined) (0x8007)
- rt-redirect IPv4 (0x8008) [2 byte AS, 4 octet value]
- rt-redirect IPv4 (0x8108) [4 byte IPv4 address, 2 octet value]
- rt-redirect IPv4 (0x8108) [4 byte AS, 2 octet value]
- traffic marking (0x8009) (DSCP value)
- Traffic rate limited by packets (0x800C) [2 byte AS, 4 byte float]
- rt-redirect IPv6 (0x820D) [2 byte AS, 4 octet value]
- rt-redirect IPv6 (0x810D) [4 byte IPv4 address, 2 octet value]
- rt-redirect IPv6 (0x820D) [4 byte AS, 2 octet value]

The flow specification filers and actions combine to make up flow specification rules associated with an NRLI. The Extended Communities for actions can be attached to a single rule or multiple rules. Figure 1 shows a diagram of the flow specification data structures.



*1 match operator may be complex.

Figure 1: BGP Flow Specification Policy

1.2. Ordering Flow Specification Data Proposed for v2

An minimal ordering specification of the rules is an order indicator per rule. The inclusion of names for each rule, match condition and action allows for logical indirection. The existing extended community which tags multiple NLRIs could be saved as an indirect reference by name. For Flow specification v1 actions, the Extended actions could be assigned default names. The actions could be linked to many NLRIs. Figure 2 below provides a logical diagram of the ordering of rules and the association of names per rule, rule match action, and rule action.

Since many policies also group data flow specifications under rule groups, many implementations may order set of rules under a

particular group policy. Network Management display of BGP filers may use the Rule Grouping mechanism to display the filters.

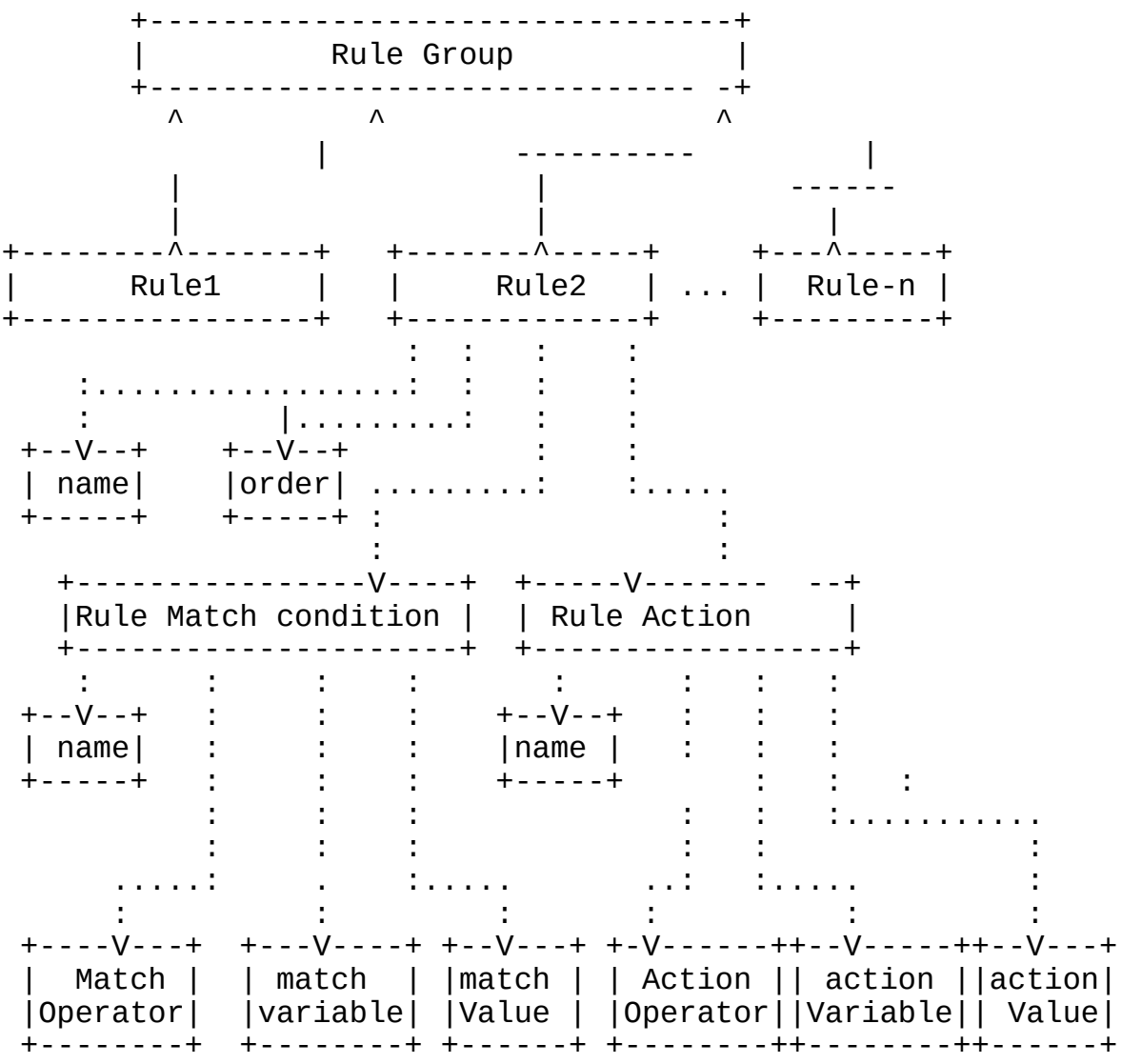


Figure 2: Order Flow Specification Data storage

2. Terminology

2.1. Definitions and Acronyms

BGPSEC - secure BGP [RFC8205] updated by [RFC8206]

BGP Session ephemeral state - state which does not survive the loss of BGP peer,

Ephemeral state - state which does not survive the reboot of a software module, or a hardware reboot. Ephemeral state can be ephemeral configuration state or operational state.

configuration state - state which persist across a reboot of software module within a routing system or a reboot of a hardware routing device.

NETCONF: The Network Configuration Protocol [RFC6241].

RESTCONF: The RESTCONF configuration Protocol [RFC8040]

ROA: Route Origin Authentication [RFC6482]

2.2. RFC 2119 language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174]

3. Dissemination of BGP Flow Specification v2 NLRI

The BGP Flow Specification version 2 (BGP-FS v2) uses an NRLI with the format for AFIs for IPv4 (AFI = 1), IPv6 (AFI = 2), and L2VPN (L2VPN = 6) with one of two following SAFIs (TBD1 for routes and TBD2 for VPN routes). This NLRI information is encoded using MP_REACH_NLRI and MP_UNREACH_NLRI attributes defined in [RFC4760]. When advertising Flow Specification, the length of the Next-Hop Network Address MUST be set to 0. The Network Address of the Next-Hop Field MUST be ignored.

Implementations wishing to exchange flow specification rules MUST use BGP’s Capability Advertisement facility to exchange the Multiprotocol Extension Capability Code (Code 1) as defined in [RFC4760], and indicate a capability for flow specification v2 (Code TBD3).

3.1. Encoding of BGP-FS v2 Filters

The AFI/SAFI NLRI for BGP Flow Specification has the format:

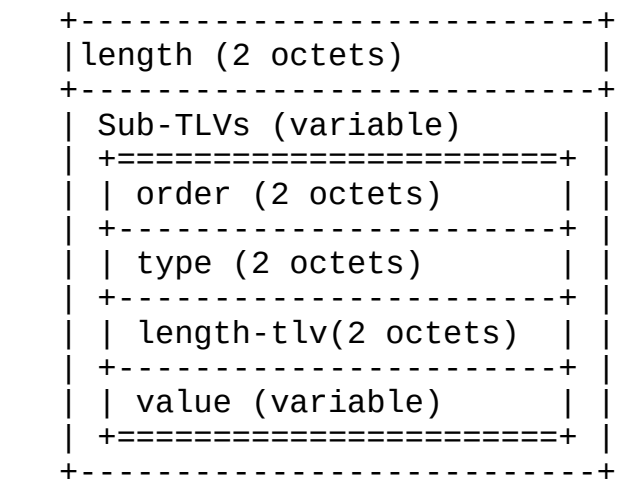


Figure 3 -Flow Specification v2 format

where:

- length - is the value of the initial length of field in overall bytes of the Sub-TLVs.
- order - is 2 octet field indicating the flow-specification global rule order.
- type - is one of the following types
  - identifier (value = 00),
  - match rule (value = 01) with default action of block traffic,
  - match rule (value = 02) with default action of permit traffic,
  - match rule (value = 03) with action TLVs,
  - match rule (value = 04) with Wide Communities Action TLVs,
  - match rule (value = 05) with tunnel matching from [I-D.ietf-idr-flowspec-l2vpn]
  - match rule (value = 06) with tunnel matching from [I-D.ietf-idr-flowspec-nvo3]
- length-tlv - is the length of the value part of the Sub-TLV,
- value is a series of sub-TLVs fields (TLV) depended on the type value defined in the sections below.

Filters are processed in the order specified by the user.

3.1.1. Encoding of Value field for Rule Identification (Value = 00)

The BGP flow specification V2 identifier sub-TLVs use the following format:

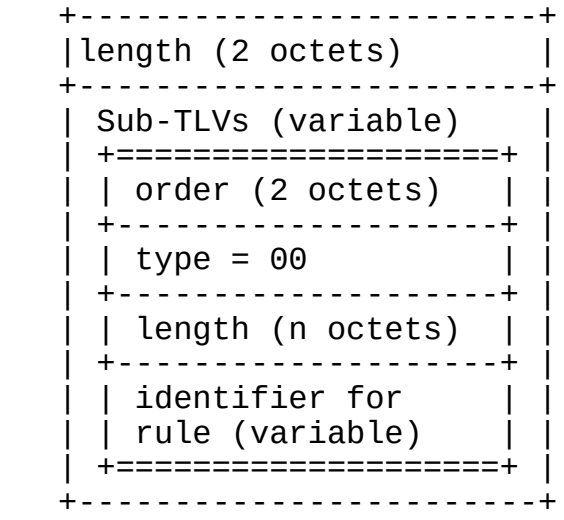


Figure 4 - NRLI revision

The identifier is a string of octets of variable length.

3.1.2. Encoding of Value field for default Action of Block traffic

The BGP flow specification V2 identifier sub-TLVs use the following format:

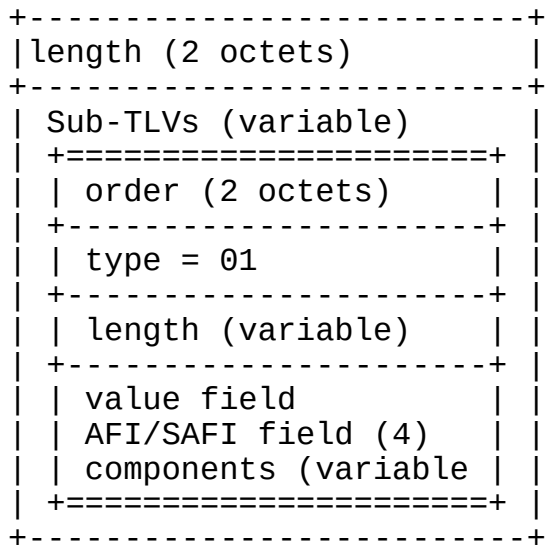


Figure 5 - Flow specification v2
with default Block traffic flow

Flow Specification v2 with a default Action of block traffic has AFI/ SAFI at the beginning of the enclosing MP_REACH_NLRI or MP_UNREACH_NLRI and the following sub-TLVs in the value field:

Component fields as defined in the following documents:

[RFC8955],

[RFC8956],

[I-D.ietf-idr-flowspec-l2vpn]

3.1.3. Encoding of Value field for default Action of Permit traffic

The BGP flow specification V2 identifier sub-TLVs use the following format:

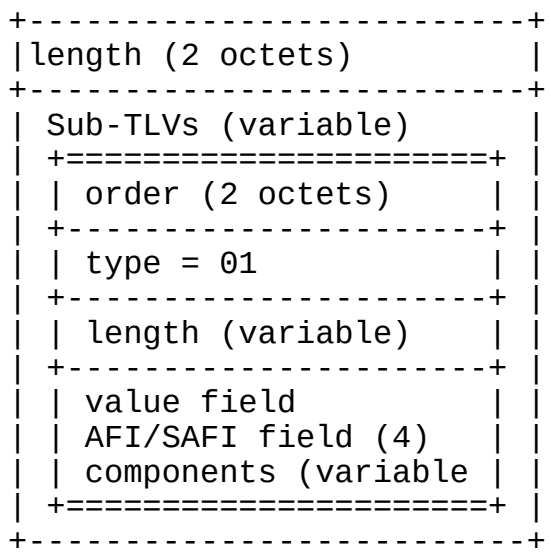


Figure 6 - Flow specification v2
with default permit traffic flow

Flow Specification v2 with Filters and Default action of block traffic has an AFI/SAFI at the beginning of the enclosing MP_REACH_NLRI or MP_UNREAC_NLRI and the following sub-TLVs in the value field:

a value of an AFI/SAFI field with 4 bytes [AFI 2 Bytes, SAFI 1 byte, 1 Byte reserved]

Component fields as defined in the following documents:

[RFC8955],

[RFC8956],

[I-D.ietf-idr-flowspec-l2vpn]

3.1.4. Encoding of Value field filters plus actions(Value = 03)

The BGP flow specification V2 identifier sub-TLVs use the following format:

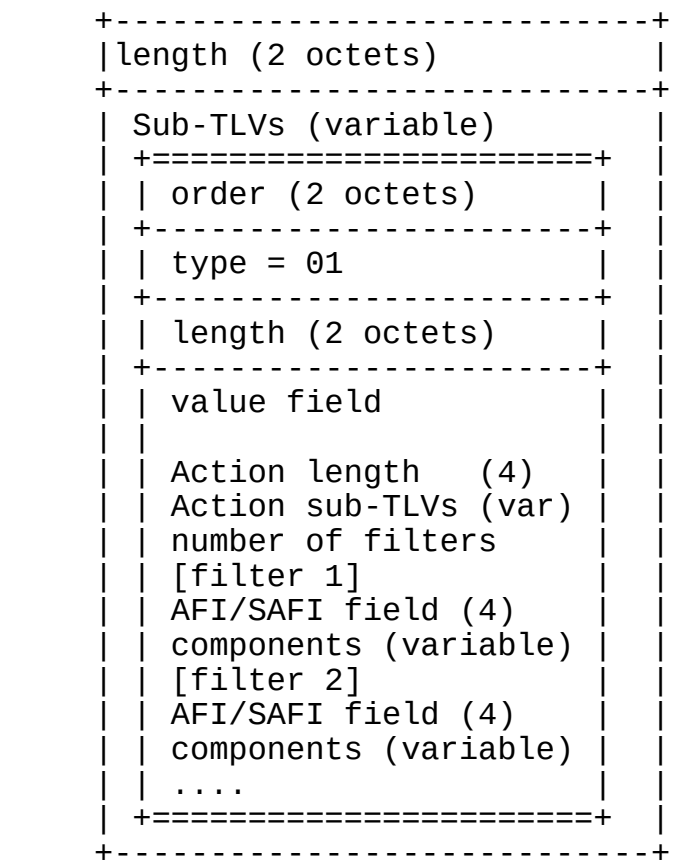


Figure 7 - Flow Specification with Actions encoded in NLRI

The Flow Specification v2 with action fields applies actions to the AFI/SAFI field. The format of the field is

Action length (4 bytes)

Action SubTLVs (variable) in format Type (2 bytes), length (2 bytes), and value (variable). The types are:

Extended community (01)

Wide Community (02)

[Type (2 bytes)][Extended-Community-type (2 bytes)][6 bytes]

Figure 8 - Extended Community action type encoding

The Extended community types are the following:

Type 1: Traffic rate limited by bytes (0x8006) [2 byte AS, 4 byte float]

Type 2: Traffic action (set by bitmask, bits 47 and 46 defined) (0x8007)

Type 3: rt-redirect IPv4 (0x8008) [2 byte AS, 4 octet value]

Type 4: rt-redirect IPv4 (0x8108) [4 byte IPv4 address, 2 octet value]

Type 5: rt-redirect IPv4 (0x8108) [4 byte AS, 2 octet value]

Type 6: traffic marking (0x8009) (DSCP value)

Type 7: Traffic rate limited by packets (0x800C) [2 byte AS, 4 byte float]

Type 8: rt-redirect IPv6 (0x820D) [2 byte AS, 4 octet value]

Type 9: rt-redirect IPv6 (0x810D) [4 byte IPv4 address, 2 octet value]

Type 10: rt-redirect IPv6 (0x820D) [4 byte AS, 2 octet value]

Component fields as defined in the following documents:

[RFC8955],

[RFC8956],

[I-D.ietf-idr-flowspec-l2vpn]

The BGP-FS version 2 actions are passed in a Wide Community [I-D.ietf-idr-wide-bgp-communities] atom with the following format.

3.1.5. Encoding of Value Fields filters passed in Wide Communities

The BGP Flow specification version 2 actions are passed in a Wide Community [I-D.ietf-idr-wide-bgp-communities] atom with the following format:

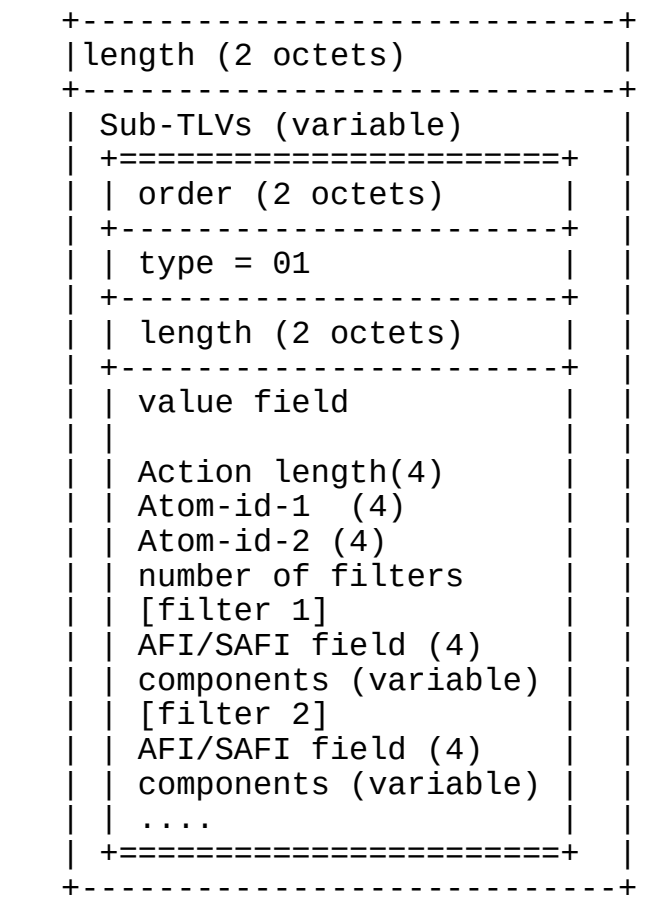


Figure 9 - Flow Specification with IDs for Wide Community Actions

The BGP Atom IDs in the Wide Community must contain:

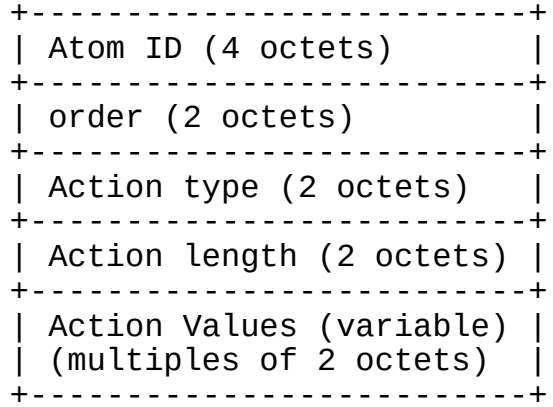


Figure 10
Wide Community Atom

where:

- Action type (2 octets) - is the type of action. These actions can be standardized (0x0001 - 0x3ffff), vendor specific (0x40000-0x7FFFF), or reserved (0x0, 0x80000-0xFFFFFFFF).

- Action length - length of actions including variable field,

- Action values - value of actions (variable) defined in individual definitions.

The BGP Flow Specification (BGP-FS) atom can be part of the Wide Community container (type 1) or the BGP Flow Specification Atom can be part of the BGP Flow Specification container (type 2) which will have:

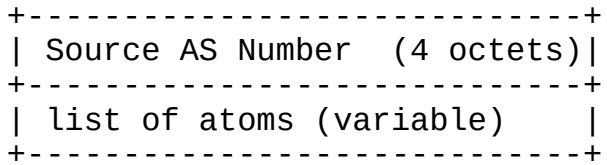


Figure 11: Atom format

3.1.6. Encoding of Value Fields filters for Tunnels (Value = 05)

AFter the initial order, type, and length the values for matching tunneled packets are the format show in Figure 12. The Tunnel Type field is a value from the IANA BGP Tunnel Encapsulation Attribute Tunnel Types Registry. If it is desired to match the packeet headers after the tunnel header, the Inner AFI field specifies the AFI for that match which is ANDed with the Outer Flowspec match. An absent Inner FlowSpec is consider to always match. The Inner flow specification for tunnel filter can also include tunnel header field components from [I-D.ietf-idr-flowspec-nvo3].

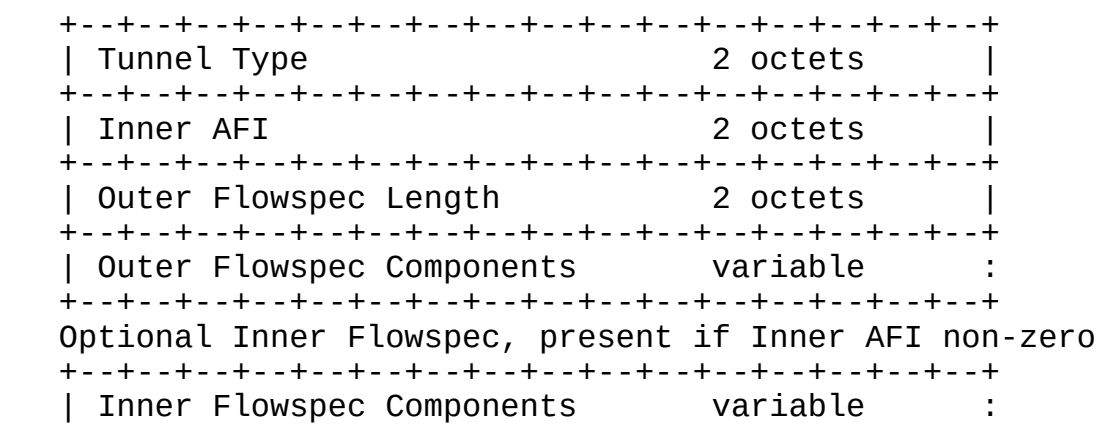


Figure 12

4. Optional Security Additions

This section discusses the optional BGP Security additions for BGP-FS v2 relating to BGPSEC [RFC8205] and ROA.

4.1. BGP FS v2 and BGPSEC

Flow specification v1 ([RFC8955] and [RFC8956]) do not BGP Flow specifications to be passed BGPSEC [RFC8205] BGP Flow Specification v2 can be passed in BGPSEC, but it is not required.

4.2. BGP FS v2 with with ROA

BGP Flow Specification v2 can utilize ROAs in the validation. If BGP-FS v2 is used with BGPSEC and ROA, the first thing is to validate the route within BGPSEC and second to utilize BGP ROA to validate the route origin.

The BGP-FS peers using both ROA and BGP-FS validation determine that a BGP Flow specification is valid if and only if one of the following cases:

- If the BGP Flow Specification NLRI has a IPv4 or IPv6 address in destination address match filter and the following is true:
  - A BGP ROA has been received to validate the originator, and
  - the route is the best-match unicast route for the destination prefix embedded in the match filter; or

- If a BGP ROA has not been received that matches the IPv4 or IPv6 destination address in the destination filter, the match filter must abide by the [RFC8955] and [RFC8956] validation rules of:
  - The originator match of the flow specification matches the originator of the best-match unicast route for the destination prefix filter embedded in the flow specification", and
  - No more specific unicast routes exist when compared with the flow destination prefix that have been received from a different neighboring AS than the best-match unicast route, which has been determined in step A.

The best match is defined to be the longest-match NLRI with the highest preference.

4.3. Revise Flow Specification Security for centralized Server

The distribution of Flow Specifications from a centralized server supports mitigation of DoS attacks. [I-D.ietf-idr-bgp-flowspec-oid] suggests the following redefined procedure for validation for this case:

A route is valid if the following conditions holds true:

- The originator of the flow specification matches the originator of the best-match unicast route for the destination prefix embedded in the flow specification.
- The AS_PATH and AS4_PATH attribute of the flow specification are empty (on originating AS)
- The AS_PATH and AS4_PATH attribute of the flow specification does not contain AS_SET and AS_SEQUENCE segments (on originating AS with AS Confederation)

This reduced validation mechanism can be used for BGP-FS v2 within a single domain.

5. IANA Considerations

This section complies with [RFC7153]

5.1. Flow Specificatoin V2 SAFIs

IANA is required to assign two SAFI Values from the registry at https://www.iana.org/assignments/safi-namespace from the Standard Action Range as follows:

Value	Description	Reference
TBD1	BGP-FS V2	[This document]
TBD2	BGP-FS V2 VPN	[this document]

5.2. BGP Capability Code

IANA is requested to assign a Capability Code from the registry at https://www.iana.org/assignments/capability-codes/ from the IETF Review range as follows:

Value	Description	Reference	Controller
TBD3	Flow Specification V2	[this document]	IETF

5.3. New Registries

IANA is requested to create the following new registries:

Registry for BGP-FS v2 action types with the

0x00 - reserved

0x01 - 0x3FFFF - standards action

0x40000- 0x7FFFF - vendor specific filters

0x80000 -0xFFFFFFFF - reserved

0x80000 -0xFFFFFFFF - reserved

Registry for BGP-FS v2 action types with the following ranges:

0x0 - reserved

0x01 - 0x3ffff - standards action

0x40000 - 0x7ffff - vendor actions

0x80000 - 0xFFFFFFF - reserved

6. Security Considerations

The use of ROA improves on [RFC8955] to check the route origination is valid can improve the validation sequence for a multiple-AS environment. The use of BGPSEC [RFC8205] to secure the packet can

increase security of BGP flow specification information sent in the packet.

The use of the reduced validation within an AS [I-D.ietf-idr-bgp-flowspec-oid] can provide adequate validation for distribution of flow specification within an single autonomous system for prevention of DDOS.

Distribution of flow filters may provide insight into traffic being sent within an AS, but this information should be composite information that does not reveal the traffic patterns of individuals.

7. References

7.1. Normative References

[I-D.ietf-idr-bgp-flowspec-oid]
Uttaro, J., Alcaide, J., Filsfils, C., Smith, D., and P. Mohapatra, "Revised Validation Procedure for BGP Flow Specifications", draft-ietf-idr-bgp-flowspec-oid-15 (work in progress), June 2021.

[I-D.ietf-idr-flowspec-l2vpn]
Hao, W., Eastlake, D. E., Litkowski, S., and S. Zhuang, "BGP Dissemination of L2 Flow Specification Rules", draft-ietf-idr-flowspec-l2vpn-17 (work in progress), May 2021.

[I-D.ietf-idr-flowspec-nvo3]
Eastlake, D., Weiguo, H., Zhuang, S., Li, Z., and R. Gu, "BGP Dissemination of Flow Specification Rules for Tunneled Traffic", draft-ietf-idr-flowspec-nvo3-13 (work in progress), February 2021.

[I-D.ietf-idr-wide-bgp-communities]
Raszuk, R., Haas, J., Lange, A., Decraene, B., Amante, S., and P. Jakma, "BGP Community Container Attribute", draft-ietf-idr-wide-bgp-communities-05 (work in progress), July 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>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <https://www.rfc-editor.org/info/rfc4360>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <https://www.rfc-editor.org/info/rfc4760>.

[RFC6482] Lepinski, M., Kent, S., and D. Kong, "A Profile for Route Origin Authorizations (ROAs)", RFC 6482, DOI 10.17487/RFC6482, February 2012, <https://www.rfc-editor.org/info/rfc6482>.

[RFC7153] Rosen, E. and Y. Rekhter, "IANA Registries for BGP Extended Communities", RFC 7153, DOI 10.17487/RFC7153, March 2014, <https://www.rfc-editor.org/info/rfc7153>.

[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>.

[RFC8955] Loibl, C., Hares, S., Raszuk, R., McPherson, D., and M. Bacher, "Dissemination of Flow Specification Rules", RFC 8955, DOI 10.17487/RFC8955, December 2020, <https://www.rfc-editor.org/info/rfc8955>.

[RFC8956] Loibl, C., Ed., Raszuk, R., Ed., and S. Hares, Ed., "Dissemination of Flow Specification Rules for IPv6", RFC 8956, DOI 10.17487/RFC8956, December 2020, <https://www.rfc-editor.org/info/rfc8956>.

7.2. Informative References

[RFC6241] Enns, R., Ed., Bjorklund, M., Ed., Schoenwaelder, J., Ed., and A. Bierman, Ed., "Network Configuration Protocol (NETCONF)", RFC 6241, DOI 10.17487/RFC6241, June 2011, <https://www.rfc-editor.org/info/rfc6241>.

[RFC8040] Bierman, A., Bjorklund, M., and K. Watsen, "RESTCONF Protocol", RFC 8040, DOI 10.17487/RFC8040, January 2017, <https://www.rfc-editor.org/info/rfc8040>.

[RFC8205] Lepinski, M., Ed. and K. Sriram, Ed., "BGPsec Protocol Specification", RFC 8205, DOI 10.17487/RFC8205, September 2017, <https://www.rfc-editor.org/info/rfc8205>.

[RFC8206] George, W. and S. Murphy, "BGPsec Considerations for Autonomous System (AS) Migration", RFC 8206, DOI 10.17487/RFC8206, September 2017, <https://www.rfc-editor.org/info/rfc8206>.

Authors' Addresses

Susan Hares
Hickory Hill Consulting
7453 Hickory Hill
Saline, MI 48176
USA

Email: shares@ndzh.com

Donald Eastlake
Futurewei
FL
USA

Email: d3e3e3@gmail.com