Workgroup: IDR Working Group
Internet-Draft: draft-ietf-idr-flowspec-v2-01
Published: 21 October 2022
Intended Status: Standards Track
Expires: 24 April 2023
Authors: S. Hares
Hickory Hill Consulting
D. Eastlake
Futurewei Technologies
C. Yadlapalli
ATT
S. Maduscke
Verizon

BGP Flow Specification Version 2

Abstract

BGP flow specification version 1 (FSv1), defined in RFC 8955, RFC 8956, and RFC 9117 describes the distribution of traffic filter policy (traffic filters and actions) distributed via BGP. Multiple applications have used BGP FSv1 to distribute traffic filter policy. These applications include the following: mitigation of denial of service (DoS), enabling traffic filtering in BGP/MPLS VPNs, centralized traffic control of router firewall functions, and SFC traffic insertion.

During the deployment of BGP FSv1 a number of issues were detected due to lack of consistent TLV encoding for rules for flow specifications, lack of user ordering of filter rules and/or actions, and lack of clear definition of interaction with BGP peers not supporting FSv1. Version 2 of the BGP flow specification (FSv2) protocol addresses these features. In order to provide a clear demarcation between FSv1 and FSv2, a different NLRI encapsulates FSv2.

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 24 April 2023.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction
  1.1. Definitions and Acronyms
  1.2. RFC 2119 language
2. Flow Specification
  2.1. Flow Specification v1 (FSv1) Overview
  2.2. Flow Specification v2 (FSv2) Overview
3. FSv2 Filters and Actions
  3.1. IP header SubTLV (type=1)
    3.1.1. IP Destination Prefix (type = 1)
    3.1.2. IP Source Prefix (type = 2)
    3.1.3. IP Protocol (type = 3)
    3.1.4. Port (type = 4)
    3.1.5. Destination Port (type = 5)
    3.1.6. Source Port (type = 6)
    3.1.7. ICMP Type (type = 7)
    3.1.8. ICMP Code (type = 8)
    3.1.9. TCP Flags (type = 9)
    3.1.10. Packet length (type = 10 (0x0A))
    3.1.11. DSCP (Differentiaed Services Code Point)(type = 11 (0x0B))
    3.1.12. Fragment (type = 12 (0x0C))
    3.1.13. Flow Label(type = 13 (0xOD))
    3.1.14. TTL (type=14 (0x0E))
    3.1.15. Parts of SID (type = 15 (0xF))
    3.1.16. MPLS Label Match1 (type=16, 0x10)
    3.1.17. MPLS Label Match 2: Experimental bits match on top label (Type=17 (0x11))
  3.2. Encoding of FSV2 Actions (type=2)
    3.2.1. Action Chain operation (ACO) (1, 0x01)
    3.2.2. Traffic Actions per interface set (TAIS) (2, 0x02)
    3.2.3. Traffic rate limited by bytes (TRB) (6, 0x06)
    3.2.4. Traffic Action (TA)(7, 0x07)

3.2.5. Redirect to IPv4 (RDIPv4)(8,0x08)
3.2.6. Traffic marking (TM) (9, 0x09)
3.2.7. Traffic rate limited by packets (TRP) (12, 0xC)
3.2.8. Traffic redirect to IPv6 (RDIPv6) (13, 0xD)
3.2.9. Traffic insertion in SFC (TISFC)(14, 0xE)
3.2.10. Flow Specification Redirect to Indirection-ID (RDIID) (15, 0x0F)
3.2.11. MPLS Label Action (MPLSLA)(16, 0x10)
3.2.12. VLAN action (VLAN) (22, 0x16)
3.2.13. TPID action (TPID) (23, 0x17)
3.3. Extended Community vs. Action SubTLV formats
3.4. L2 Traffic Rules
3.5. SFC Traffic Rules
3.6. BGP/MPLS VPN IP Traffic Rules
3.7. BGP/MPLS VPN L2 Traffic Rules
3.8. Encoding of Actions passed in Wide Communities
4. Validation of FSv2 NLRI
4.1. Validation of FS NLRI (FSv1 or FSv2)
4.2. Validation of Flow Specification Actions
4.3. Error handling and Validation
5. Ordering for Flow Specification v2 (FSv2)
5.1. Ordering of FSv2 NLRI Filters
5.2. Ordering of the Actions
5.2.1. Action Chain Operation (ACO)
5.2.2. Summary of FSv2 ordering
6. Ordering of FS filters for BGP Peers support FSv1 and FSv2
7. Scalability and Aspirations for FSv2
8. Optional Security Additions
8.1. BGP FSv2 and BGPSEC
8.2. BGP FSv2 with ROA
9. IANA Considerations
9.1. Flow Specification V2 SAFIs
9.2. BGP Capability Code
9.3. Filter IP Component types
9.4. FSV2 NLRI TLV Types
9.5. Wide Community Assignments
10. Security Considerations
11. References
11.1. Normative References
11.2. Informative References
Authors' Addresses

1. Introduction

Modern IP routers have the capability to forward traffic and to classify, shape, rate limit, filter, or redirect packets based on administratively defined policies. These traffic policy mechanisms allow the operator to define match rules that operate on multiple fields within header of an IP data packet. The traffic policy allows

actions to be taken upon a match to be associated with each match rule. These rules can be more widely defined as "event-condition-action" (ECA) rules where the event is always the reception of a packet.

BGP ([RFC4271]) flow specification as defined by [RFC8955], [RFC8956], [RFC9117] specifies the distribution of traffic filter policy (traffic filters and actions) via BGP to a mesh of BGP peers (IBGP and EBGP peers). The traffic filter policy is applied when packets are received on a router with the flow specification function turned on. The flow specification protocol defined in [RFC8955], [RFC8956], and [RFC9117] will be called BGP flow specification version 1 (BGP FSv1) in this draft.

Some modern IP routers also include the abilities of firewalls which can match on a sequence of packet events based on administrative policy. These firewall capabilities allow for user ordering of match rules and user ordering of actions per match.

Multiple deployed applications currently use BGP FSv1 to distribute traffic filter policy. These applications include: 1) mitigation of Denial of Service (DoS), 2) traffic filtering in BGP/MPLS VPNS, and 3) centralized traffic control for networks utilizing SDN control of router firewall functions, 4) classifiers for insertion in an SFC, and 5) filters for SRv6 (segment routing v6).

During the deployment of BGP flow specification v1, the following issues were detected:

* lack of consistent TLV encoding prevented extension of encodings,

* inability to allow user defined order for filtering rules,

* inability to order actions to provide deterministic interactions or to allow users to define order for actions, and

* no clearly defined mechanisms for BGP peers which do not support flow specification v1.

Networks currently cope with some of these issues by limiting the type of traffic filter policy sent in BGP. Current Networks do not have a good workaround/solution for applications that receive but do not understand FSv1 policies.

This document defines version 2 of the BGP flow specification protocol to address these shortcomings in BGP FSv1. Version 2 of BGP flow specification will be denoted as BGP FSv2.

BGP FSv1 as defined in [RFC8955], [RFC8956], and [RFC9117] specified 2 SAFIs (133, 134) to be used with IPv4 AFI (AFI = 1) and IPv6 AFI (AFI=2).

This document specifies 2 new SAFIs (TBD1, TBD2) for FSv2 to be used with 5 AFIs (1, 2, 6, 25, and 31) to allow user-ordered lists of traffic match filters for user-ordered traffic match actions encoded in Communities (Wide or Extended).

FSv1 and FSv2 use different AFI/SAFIs to send flow specification filters. Since BGP route selection is performed per AFI/SAFI, this approach can be termed "ships in the night" based on AFI/SAFI.

FSv1 is a critical component of deployed applications. Therefore, this specification defines how FSv2 will interact with BGP peers that support either FSv2, FSv1, FSv2 and FSv1,or neither of them. It is expected that a transition to FSv2 will occur over time as new applications require FSv2 extensibility and user-defined ordering for rules and actions or network operators tire of the restrictions of FSv1 such as error handling issues and restricted topologies.

Section 2 contains the definition of Flow specification, a short review of FSv1 and an overview of FSv2. Section 3 contains the encoding rules for FSv2 and user-based encoding sent via BGP. Section 4 describes how to validate FSv2 NLRI. Section 5 discusses how to order FSV2 rules. Section 6 covers combining FSv2 user-ordered match rules and FSv1 rules. Section 6 also discusses how to combine user-ordered actions, FSv1 actions, and default actions. Sections 7-10 address an alternate security mechanism, considerations for IANA, security in deployments, and scalability aspirations.

1.1. Definitions and Acronyms

AFI - Address Family Identifier

AS - Autonomous System

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

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

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

DDOs - Distributed Denial of Service.

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.

FSv1 - Flow Specification version 1 [RFC8955] [RFC8956]

FSv2 - Flow Specification version 2 (this document)

NETCONF - The Network Configuration Protocol [RFC6241].

RESTCONF - The RESTCONF configuration Protocol [RFC8040]

RIB - Routing Information Base.

ROA - Route Origin Authentication [RFC6482]

RR - Route Reflector.

SAFI - Subsequent Address Family Identifier

1.2. 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] when, and only when, they appear in all capitals as shown here.

2. Flow Specification

A BGP Flow Specification is an n-tuple containing one or more match criteria that can be applied to IP traffic, traffic encapsulated in IP traffic or traffic associated with IP traffic. The following are examples of such traffic: IP packet or an IP packet inside a L2 packet (Ethernet), an MPLS packet, and SFC flow.

A given Flow Specification NLRI may be associated with a set of path attributes depending on the particular application, and attributes within that set may or may not include reachability information (e.g., NEXT_HOP). Extended Community or Wide Community attributes (well-known or AS-specific) MAY be used to encode a set of pre-determined actions.

A particular application is identified by a specific AFI/SAFI (Address Family Identifier/Subsequent Address Family Identifier) and corresponds to a distinct set of RIBs. Those RIBs should be treated independently of each other in order to assure noninterference between distinct applications.

BGP processing treats the NLRI as a key to entries in AFI/SAFI BGP databases. Entries that are placed in the Loc-RIB are then associated with a given set of semantics which are application dependent. Standard BGP mechanisms such as update filtering by NLRI or by attributes such as AS_PATH or large communities apply to the BGP Flow Specification defined NLRI-types.

Network operators can control the propagation of BGP routes by enabling or disabling the exchange of routes for a particular AFI/SAFI pair on a particular peering session. As such, the Flow Specification may be distributed to only a portion of the BGP infrastructure.

2.1. Flow Specification v1 (FSv1) Overview

The FSv1 NLRI defined in [RFC8955] and [RFC8956] include 13 match conditions encoded for the following AFI/SAFIs:

- *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

If one considers the reception of the packet as an event, then BGP FSv1 describes a set of Event-MatchCondition-Action (ECA) policies where:

- *event is the reception of a packet,
- *condition stands for "match conditions" defined in the BGP NLRI as an n-tuple of component filters, and
- *the action is either: the default condition (accept traffic), or a set of actions (1 or more) defined in Extended BGP Community values [RFC4360].

The flow specification conditions and actions combine to make up FSv1 specification rules. Each FSv1 NLRI must have a type 1 component (destination prefix). Extended Communities with FSv1 actions can be attached to a single NLRI or multiple NLRIs in a BGP message

Within an AFI/SAFI pair, FSv1 rules are ordered based on the components in the packet (types 1-13) ordered from left-most to right-most and within the component types by value of the component. Rules are inserted in the rule list by component-based order where

an FSv1 rule with existing component type has higher precedence than one missing a specific component type,

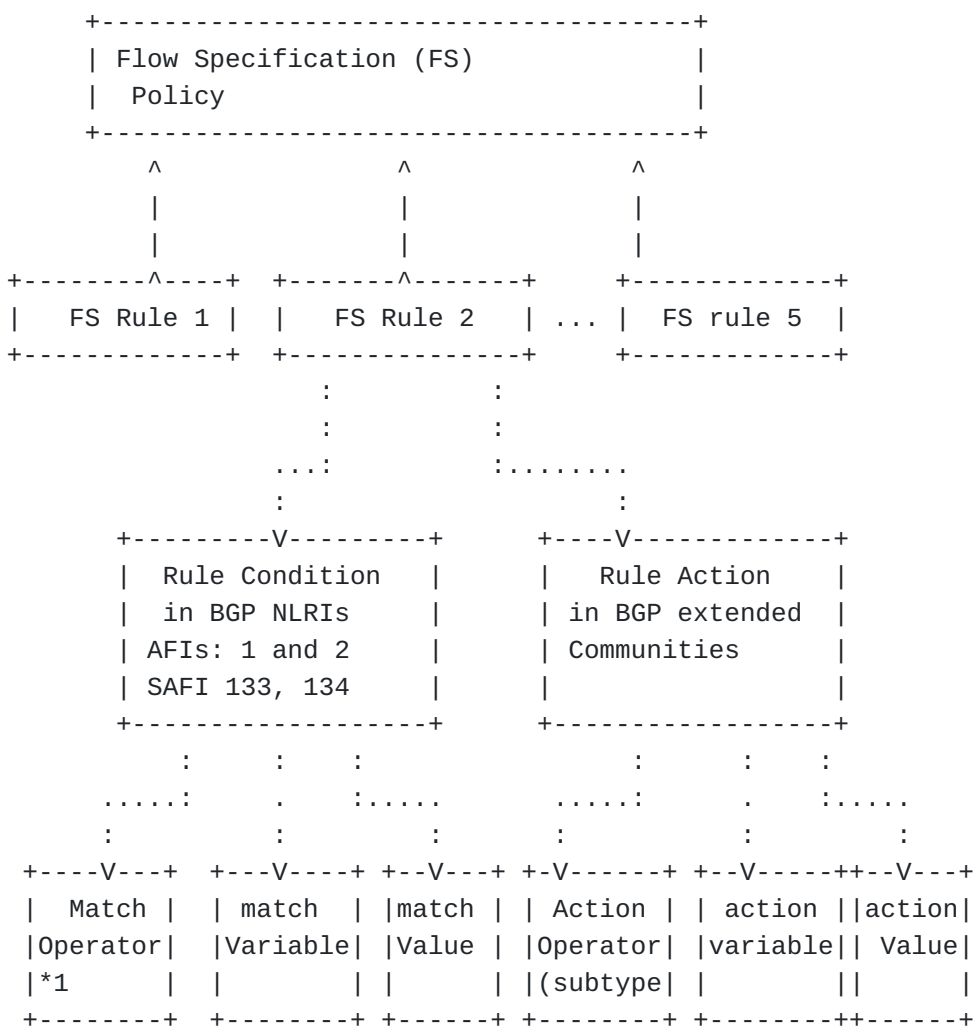
Since FSv1 specifications ([RFC8955], [RFC8956], and [RFC9117]) specify that the FSv1 NLRI MUST have a destination prefix (as component type 1) embedded in the flow specification, the FSv1 rules with destination components are ordered by IP Prefix comparison rules for IPv4 ([RFC8955]) and IPv6 ([RFC8956]). [RFC8955] specifies that more specific prefixes (aka longest match) have higher precedence than that of less specific prefixes and that for prefixes of the same length the lower IP number is selected (lowest IP value). [RFC8955] specifies that if the offsets within component 1 are the same, then the longest match and lowest IP comparison rules from [RFC8955] apply. If the offsets are different, then the lower offset has precedence.

These rules provide a set of FSv1 rules ordered by IP Destination Prefix by longest match and lowest IP address. [RFC8955] also states that the requirement for a destination prefix component "MAY be relaxed by explicit configuration" Since the rule insertions are based on comparing component types between two rules in order, this means the rules without destination prefixes are inserted after all rules which contain destination prefix component.

The actions specified in FSv1 are:

- *accept packet (default),
- *traffic flow limitation by bytes (0x6),
- *traffic-action (0x7),
- *redirect traffic (0x8),
- *mark traffic (0x9), and
- *traffic flow limitation by packets (12, 0xC)

Figure 1 shows a diagram of the FSv1 logical data structures with 5 rules. If FSv1 rules have destination prefix components (type=1) and FSv1 rule 5 does not have a destination prefix, then FSv1 rule 5 will be inserted in the policy after rules 1-4.



*1 match operator may be complex.

Figure 2-1: BGP Flow Specification v1 Policy

2.2. Flow Specification v2 (FSv2) Overview

Flow Specification v2 allows the user to order the flow specification rules and the actions associated with a rule. Each FSv2 rule may have one or more match conditions and one or more associated actions.

This FSv2 specification supports the components and actions for the following:

- *IPv4 (AFI=1, SAFI=TBD1),
- *IPv6 (AFI=2, SAFI=TBD2),
- *L2 (AFI=6, SAFI=TDB1),

*BGP/MPLS IPv4 VPN: (AFI=1, SAFI=TBD2),

*BGP/MPLS IPv6 VPN: (AFI=2, SAFI=TBD2),

*BGP/MPLS L2VPN (AFI=25, SAFI=TDB2),

*SFC: (AFI=31, SAFI=TBD1), and

*SFC VPN (AFI=31, SAFI=TBD2).

The FSv2 specification for tunnel traffic is outside the scope of this specification. The FSv1 specification for tunneled traffic is in [I-D.ietf-idr-flowspec-nvo3].

FSv2 operates in the ships-in-the night model with FSv1 so network operators can manipulate which the distribution of FSv2 and FSv1 using configuration parameters. Since the lack of deterministic ordering was an FSv1 problem, this specification provides rules and protocol features to keep filters in a deterministic order between FSv1 and FSv2.

The basic principles regarding ordering of flow specification filter rules are:

1) Rule-0 (zero) is defined to be 0/0 with the "permit-all" action.

2) FSv2 rules are ordered based on user-specified order.

-The user-specified order is carried in the FSv2 NLRI and a numerical lower value takes precedence over a numerically higher value. For rules received with the same order value, the FSv1 rules apply (order by component type and then by value of the components).

3) FSv2 rules are added starting with Rule 1 and FSv1 rules are added after FSv2 rules

-For example, BGP Peer A has FSv2 data base with 10 FSv2 rules (1-10). FSv1 user number is configured to start at 301 so 10 FSv1 rules are added at 301-310.

4) An FSv2 peer may receive BGP NLRI routes from a FSv1 peer or a BGP peer that does not support FSv1 or FSv2. The capabilities sent by a BGP peer indicate whether the AFI/SAFI can be received (FSv1 NLRI or FSv2 NLRI).

5) Associate a chain of actions to rules based on user-defined action number (1-n). (optional)

- If no actions are associated with a filter rule, the default is to drop traffic the filter rules match

- An action chain of 1-n actions can be associated with a set of filter rules can via Extended Communities or Wide Communities. Only Wide Communities can associate a user-defined order for the actions. Extended Community actions occur after actions with a user specified order (see section 5.2 for details).

Figure 2-2 provides a logical diagram of the FSv2 structure

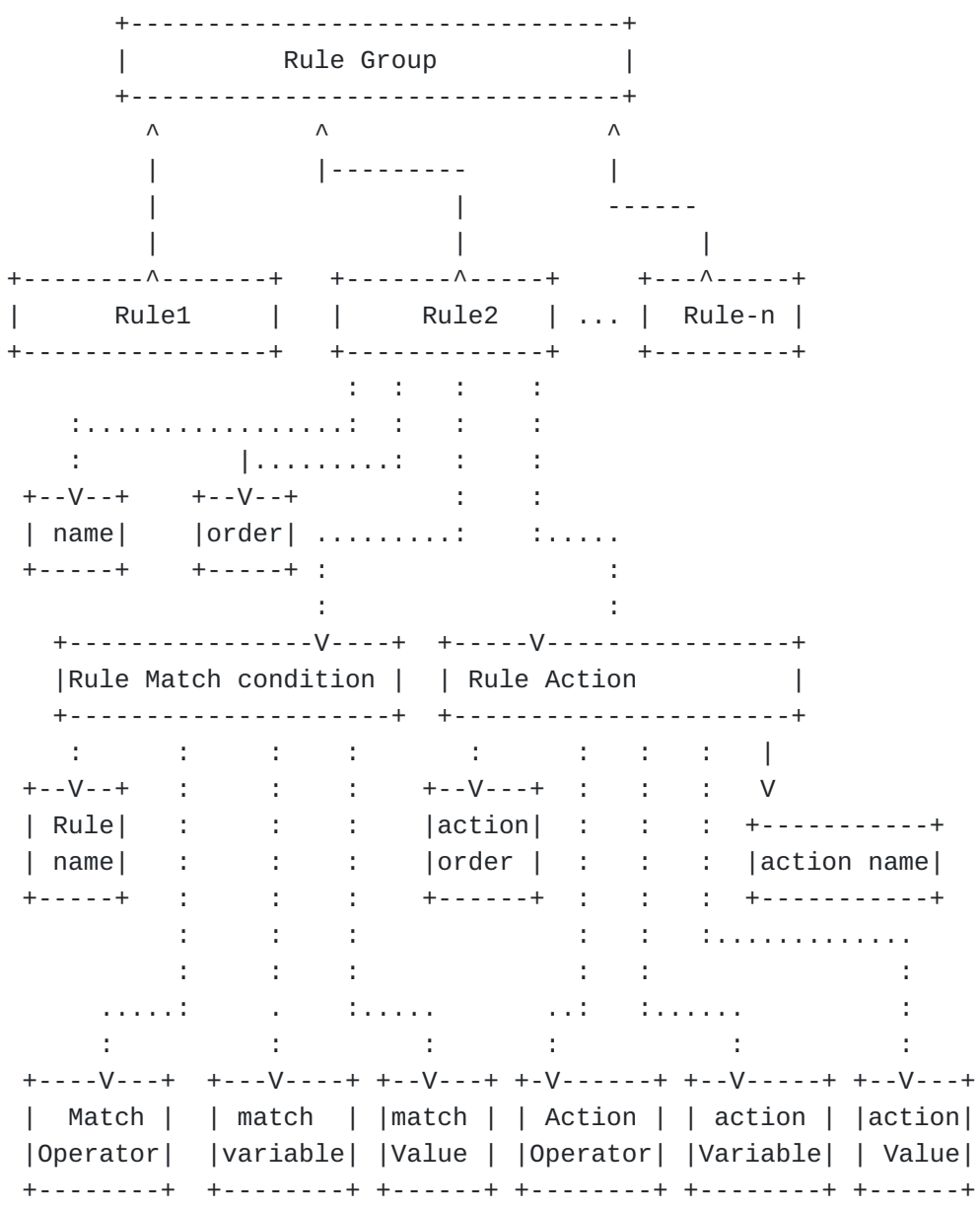


Figure 2-2: BGP FSv2 Data storage

3. FSv2 Filters and Actions

The BGP FSv2 uses an NRLI with the format for AFIs for IPv4 (AFI = 1), IPv6 (AFI = 2), L2 (AFI = 6), L2VPN (AFI=25), and SFC (AFI=31) with two following SAFIs to support transmission of the flow specification which supports user ordering of traffic filters and actions for IP traffic and IP VPN traffic.

This NLRI information is encoded using MP_REACH_NLRI and MP_UNREACH_NLRI attributes defined in [RFC4760]. When advertising FSv2 NLRI, the length of the Next-Hop Network Address MUST be set to

0. Upon reception, the Network Address in 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 FSv1, FSv2 (Code TBD3), or both.

The AFI/SAFI NLRI for BGP Flow Specification version 2 (FSv2) has the format:

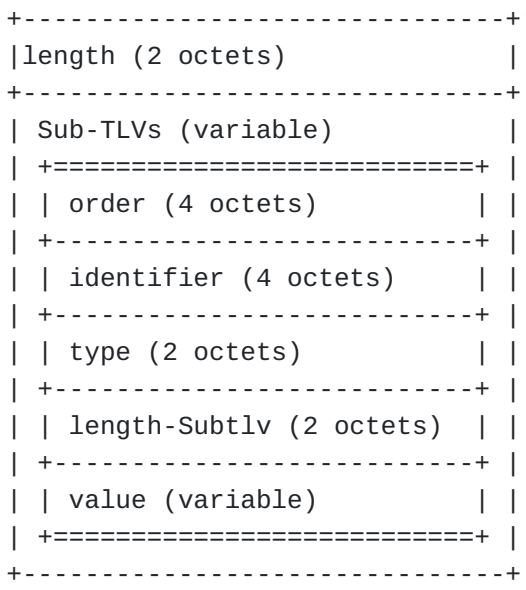


Figure 3-1: FSv2 format

where:

*length: length of field including all SubTLVs in octets.

-The combined lengths of any FSv2 NLRI in the MP_REACH_NLRI or MP_UNREACH_NLRI. The BGP NLRI length must be less than the packet size minus the other fields (BGP header, BGP Path Attributes, and NLRI).

*order: flow-specification global rule order number (4 octets).

*identifier: identifier for the rule (used for NM/Logging) (4 octets)

*type: contains a type for FSv2 TLV format of the NRLI (2 octets) which can be:

-0 - reserved,

-1 - IP Traffic Rules

-2- L2 traffic rules

-3- SFC Traffic rules

-4- SFC VPN Traffic rules

-5 - BGP/MPLS VPN IP Traffic Rules

-6 - BGP/MPLS VPN L2 Traffic Rules

*length-Subtlv: is the length of the value part of the Sub-TLV,

*value: value depends on the subTLV (see sections below).

3.1. IP header SubTLV (type=1)

The format of the IP header TLV value field is shown in figure 4. The AFI/SAFI field includes the AFI (2 octets), SAFI (1 octet). The AFI will be 1 (IPv4) or 2 (IPv6) and the SAFI will be TBD1 or, for the VPN case, TBD2. The IP header for the VPN case is specified in section 3.5.

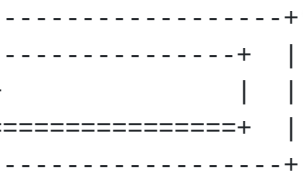


Figure 3-2 - IP Header TLV

Where: AFI is 1 (IPv4) or 2 (IPv6) and SAFI is TBD1.

Each SubTLV has the format:

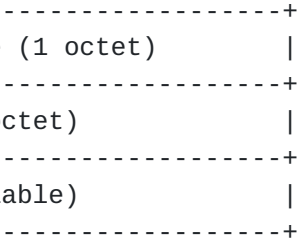


Figure 3-3 – IP header SubTLV format

Where:

SubTLV type: component values are defined in the "Flow Specification Component types" registry for IPv4 and IPv6 by [RFC8955], [RFC8956], and [I-D.ietf-idr-flowspec-srv6]

length: length of SubTLV (varies depending on SubTLV type).

value: dependent on the subTLV

-For descriptions of value portions for components 1-13 see [RFC8955] and [RFC8956]. For component 14 see [I-D.ietf-idr-flowspec-srv6].

Many of the components use the operators [numeric_op] and [bitmask_op] defined in [RFC8955]

The list of valid SubTLV types appears in Table 2.

Table 2 IP SubTLV Types for IP Filters

SubTLV-type	Definition
1	IP Destination prefix
2	IP Source prefix
3	IPv4 Protocol / IPv6 Upper Layer Protocol
4	Port
5	Destination Port
6	Source Port
7	ICMPv4 type / ICMPv6 type
8	ICMPv4 code / ICPv6 code
9	TCP Flags
10	Packet length
11	DSCP (Differentiated Services Code Point)
12	Fragment
13	Flow Label
14	TTL
15	Parts of SID
16	MPLS Match 1: Label in Label stack
17	MPLS Match 2: EXP bits in top Label

Ordering within the TLV in FSv2: The transmission of SubTLVs within a flow specification rule MUST be sent ascending order by SubTLV type. If the SubTLV types are the same, then the value fields are compared using mechanisms defined in [RFC8955] and [RFC8956] and MUST be in ascending order. NLRIs having TLVs which do not follow the above ordering rules MUST be considered as malformed by a BGP FSv2 propagator. This rule prevents any ambiguities that arise from the multiple copies of the same NLRI from multiple BGP FSv2

propagators. A BGP implementation SHOULD treat such malformed NLRIs as "Treat-as-withdraw" [RFC7606].

See [RFC8955], [RFC8956], and [I-D.ietf-idr-flowspec-srv6]. for specific details.

3.1.1. IP Destination Prefix (type = 1)

IPv4 Name: IP Destination Prefix (reference: [RFC8955])

IPv6 Name: IPv6 Destination Prefix (reference: [RFC8956])

IPv4 length: Prefix length in bits

IPv4 value: IPv4 Prefix (variable length)

IPv6 length: length of value

IPv6 value: [offset (1 octet)] [pattern (variable)] [padding(variable)]

If IPv6 length = 0 and offset = 0, then component matches every address. Otherwise, length must be offset "less than" length "less than" 129 or component is malformed.

3.1.2. IP Source Prefix (type = 2)

IPv4 Name: IP Source Prefix (reference: [RFC8955])

IPv6 Name: IPv6 Source Prefix (reference: [RFC8956])

IPv4 length: Prefix length in bits

IPv4 value: Source IPv4 Prefix (variable length)

IPv6 length: length of value

IPv6 value: [offset (1 octet)] [pattern (variable)] [padding(variable)]

If IPv6 length = 0 and offset = 0, then component matches every address. Otherwise, length must be offset < length < 129 or component is malformed.

3.1.3. IP Protocol (type = 3)

IPv4 Name: IP Protocol IP Source Prefix (reference: [RFC8955])

IPv6 Name: IPv6 Upper-Layer Protocol: (reference: [RFC8956])

IPv4 length: variable

IPv4 value: [numeric_op, value]+

IPv6 length: variable

IPv6 value: [numeric_op, value}+

where the value following each numeric_op is a single octet.

3.1.4. Port (type = 4)

IPv4/IPv6 Name: Port (reference: [RFC8955]), [RFC8956])

Filter defines: a set of port values to match either destination port or source port.

IPv4 length: variable

IPv4 value: [numeric_op, value]+

IPv6 length: variable

IPv6 value: [numeric_op, value]+

where the value following each numeric_op is a single octet.

Note-1: (from FSV1) In the presence of the port component (destination or source port), only a TCP (port 6) or UDP (port 17) packet can match the entire flow specification. If the packet is fragmented and this is not the first fragment, then the system may not be able to find the header. At this point, the FSv2 filter may fail to detect the correct flow. Similarly, if other IP options or the encapsulating security payload (ESP) is present, then the node may not be able to describe the transport header and the FSv2 filter may fail to detect the flow.

The restriction in note-1 comes from the inheritance of the FSv1 filter component for port. If better resolution is desired, a new FSv2 filter should be defined.

Note-2: FSv2 component only matches the first upper layer protocol value.

3.1.5. Destination Port (type = 5)

IPv4/IPv6 Name: Destination Port (reference: [RFC8955]), [RFC8956])

Filter defines: a list of match filters for destination port for TCP or UDP within a received packet

Length: variable

Component Value format: [numeric_op, value]+

3.1.6. Source Port (type = 6)

IPv4/IPv6 Name: Source Port (reference: [RFC8955]), [RFC8956])

Filter defines: a list of match filters for source port for TCP or UDP within a received packet

IPv4/IPv6 length: variable

IPv4/Ipv6 value: [numeric_op, value]+

3.1.7. ICMP Type (type = 7)

IPv4: ICMP Type (reference: [RFC8955])

Filter defines: Defines: a list of match criteria for ICMPv4 type

IPv6: ICMPv6 Type (reference: [RFC8956])

Filter defines: a list of match criteria for ICMPv6 type.

IPv4/IPv6 length: variable

IPv4/IPv6 value: [numeric_op, value]+

3.1.8. ICMP Code (type = 8)

IPv4: ICMP Type (reference: [RFC8955])

Filter defines: a list of match criteria for ICMPv4 code.

IPv6: ICMPv6 Type (reference: [RFC8956])

Filter defines: a list of match criteria for ICMPv6 code.

IPv4/IPv6 length: variable

IPv4/IPv6 value: [numeric_op, value]+

3.1.9. TCP Flags (type = 9)

IPv4/IPv6: TCP Flags Code (reference: [RFC8955])

Filter defines: a list of match criteria for TCP Control bits

IPv4/IPv6 length: variable

IPv4/IPv6 value: [bitmask_op, value]+

Note: a 2 octets bitmask match is always used for TCP-Flags

3.1.10. Packet length (type = 10 (0x0A))

IPv4/IPv6: Packet Length (reference: [RFC8955], [RFC8956])

Filter defines: a list of match criteria for length of packet (excluding L2 header but including IP header).

IPv4/IPv6 length: variable

IPv4/IPv6 value: [numeric_op, value]+

Note:[RFC8955] uses either 1 or 2 octet values.

3.1.11. DSCP (Differentiaed Services Code Point)(type = 11 (0x0B))

IPv4/IPv6: DSCP Code (reference: [RFC8955], [RFC8956])

Filter defines: a list of match criteria for DSCP code values to match the 6-bit DSCP field.

IPv4/IPv6 length: variable

IPv4/IPv6 value: [numeric_op, value]+

Note: This component uses the Numeric Operator (numeric_op) described in [RFC8955] in section 4.2.1.1. Type 11 component values MUST be encoded as single octet (numeric_op len=00).

The six least significant bits contain the DSCP value. All other bits SHOULD be treated as 0.

3.1.12. Fragment (type = 12 (0x0C))

IPv4/IPv6: Fragment (reference: [RFC8955], [RFC8956])

Filter defines: a list of match criteria for specific IP fragments.

Length: variable

Component Value format: [bitmask_op, value]+

Bitmask values are:

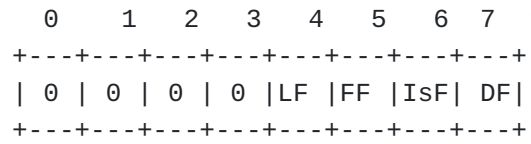


Figure 3-4

Where:

DF (don't fragment): match If IP header flags bit 1 (DF) is 1.

IsF(is a fragment other than first: match if IP header fragment offset is not 0.

FF (First Fragment): Match if [RFC0791] IP Header Fragment offset is zero and Flags Bit-2 (MF) is 1.

LF (last Fragment): Match if [RFC7091] IP header Fragment is not 0 And Flags bit-2 (MF) is 0

0: MUST be sent in NLRI encoding as 0, and MUST be ignored during reception.

3.1.13. Flow Label(type = 13 (0xOD))

IPv4/IPv6: Fragment (reference: [RFC8956])

Filter defines: a list of match criteria for 20-bit Flow Label in the IPv6 header field.

Length: variable

Component Value format: [numeric_op, value]+

3.1.14. TTL (type=14 (0x0E))

TTL: Defines matches for 8-bit TTL field in IP header

Encoding: <[numeric_op, value]+>

where: value is a 1 octet value for TTL.

ordering: by full value of number_op concatenated with value

conflict: none

reference: draft-bergeon-flowspec-ttl-match-00.txt

3.1.15. Parts of SID (type = 15 (0xF))

IPv6: Service Identifier Matches (reference: [I-D.ietf-idr-flowspec-srv6]

Filter defines: a list of match bit match criteria for some combinations of the LOC (location), FUNCT (function) and ARG (arguments) fields in the SID or or whole SID.

Length: variable

Component Value format: [type, LOC-Len, FUNCT-Len, ARG-Len, [op, value]+]

where:

- *type (1 octet): This indicates the new component type (TBD1, which is to be assigned by IANA).
- *LOC-Len (1 octet): This indicates the length in bits of LOC in SID.
- *FUNCT-Len (1 octet): This indicates the length in bits of FUNCT in SID.
- *ARG-Len (1 octet): This indicates the length in bits of ARG in SID.
- *[op, value]+: This contains a list of {operator, value} pairs that are used to match some parts of SID.

The total of three lengths (i.e., LOC length + FUNCT length + ARG length) MUST NOT be greater than 128. If it is greater than 128, an error occurs and it is treated as a withdrawal [RFC7606] and [RFC4760].

The operator (op) byte is encoded as:

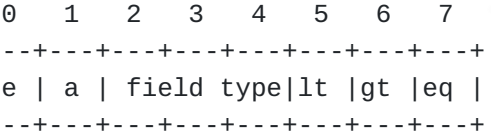


Figure 3-5

where:

where the behavior of each operator bit has clear similarity with that of [RFC8955]'s Numeric Operator field.

e (end-of-list bit): Set in the last {op, value} pair in the sequence.

a - AND bit: If unset, the previous term is logically ORed with the current one. If set, the operation is a logical AND. It should be unset in the first operator byte of a sequence. The AND operator has higher priority than OR for the purposes of evaluating logical expressions.

field type:

- -000: SID's LOC

- -001: SID's FUNCT

- -010: SID's ARG

- -011: SID's LOC:FUNCT (the concatenation of the LOC and FUNCTION fields)

- -100: SID's FUNCT:ARG (the concatenation of the FUNCTION and ARG fields)

- -101: SID's LOC:FUNCT:ARG (the concatenation of the FUNCTION and ARG fields)

Note: For an unknown field type, Error Handling is to "treat as withdrawal" [RFC7606] and [RFC4760].

lt: less than comparison between data' and value'.

gt: greater than comparison between data' and value'.

eq: equality between data' and value'.

The data' and value' used in lt, gt and eq are indicated by the field type in an operator and the value field following the operator.

The length of the value field depends on the field type and is the length of the SID parts being matched (see Table 3, Figure 3-6) in bytes, rounded up if that length is not a multiple of 8.

Table 3 - SID Parts fields

Field Type	Value
SID's LOC	value of LOC bits
SID's FUNCT	value of FUNCT bits
SID's ARG	value of ARG bits
SID's LOC:FUNCT	value of LOC:FUNCT bits
SID's FUNCT:ARG	value of FUNCT:ARG bits
SID's LOC:FUNCT:ARG	value of LOC:FUNCT:ARG bits

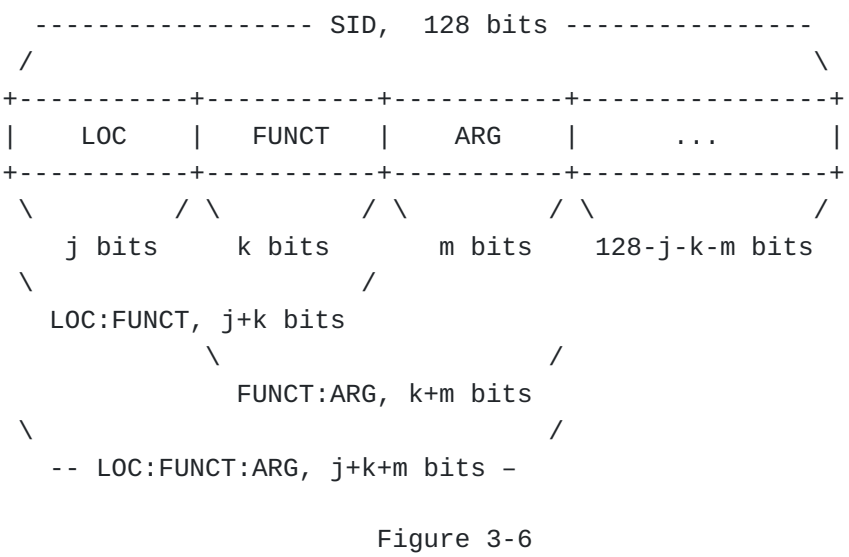


Figure 3-6

3.1.16. MPLS Label Match1 (type=16, 0x10)

Function: This match1 applies to MPLS Label field on the label stack.

reference: [I-D.ietf-idr-flowspec-mpls-match]

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

It contains a set of {operator, value} pairs that are used for the matching filter.

The operator byte is encoded as:

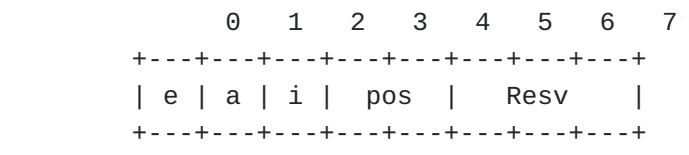


Figure 3-7

where:

e - end of list bit: Set in the last {op, value} pair in the list.

a - AND bit: If unset, the previous term is logically ORed with the current one. If set, the operation is a logical AND. It should be unset in the first operator byte of a sequence. The AND operator has higher priority than OR for the purposes of evaluating logical expressions.

i - before bit: If unset, apply matching filter before MPLS label data plane action; if set, apply matching filter after MPLS label data plane action.

pos - the label position indication bits: whose meaning for various values is shown below:

00:any position on the label stack - the presented label value is used to match any label on the label stack. When applying it, at least one label on the stack MUST match the value

01:top label indication- the presented label value MUST be used to match the top label on the label stack.

10: bottom label indication- the presented label value MUST match the bottom label on the label stack. When it is clear, the present label value can match to any label on the label stack

11: reserved value - - This value is reserved and MUST not be sent in the packet.

The value field is encoded as:

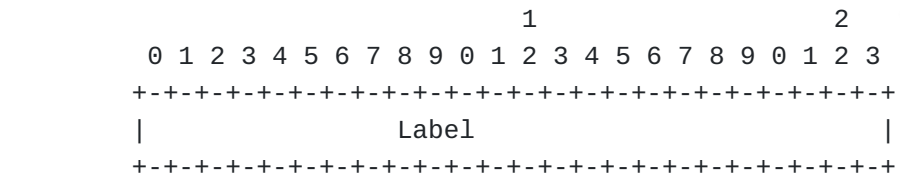


Figure 3-8

reference:

3.1.17. MPLS Label Match 2: Experimental bits match on top label (Type=17 (0x11))

Function: MPLS Match2 applies to MPLS Label experiment bits (EXP) on the top label in the label stack.

reference: [I-D.ietf-idr-flowspec-mpls-match]

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

- -[op,value] - Defines a list of {operation, value} pairs used to match 3-bit exp field on the top label of packets [RFC3032].

- -Values are encoded using a single byte, where the five most significant bits are zero and the three least significant bits contain the exp value.

3.2. Encoding of FSV2 Actions (type=2)

The FSv2 actions may be sent in an Extended Community or a Wide Community.

The Extended Community encodes the Flow Specification actions in the Extended Community format [RFC4360].

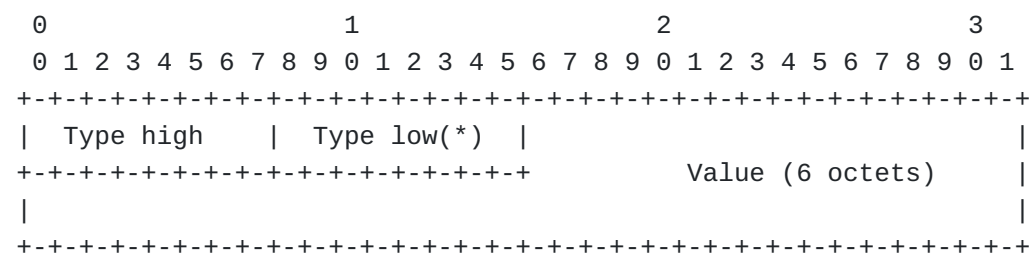


Figure 3-9

The Wide Community definition for FSv2 actions is as follows:

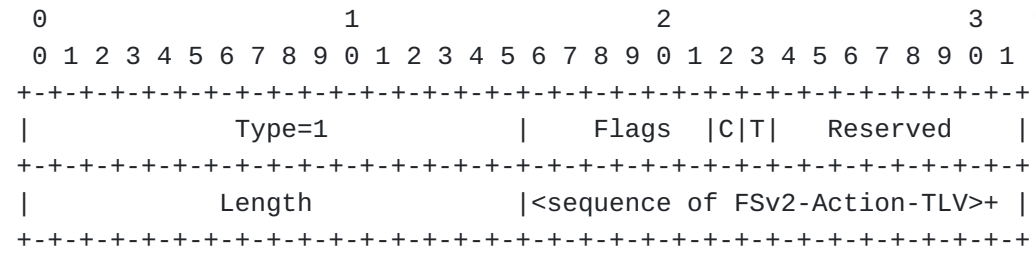


Figure 3-10

where FSv2-Action-TLV is defined as:

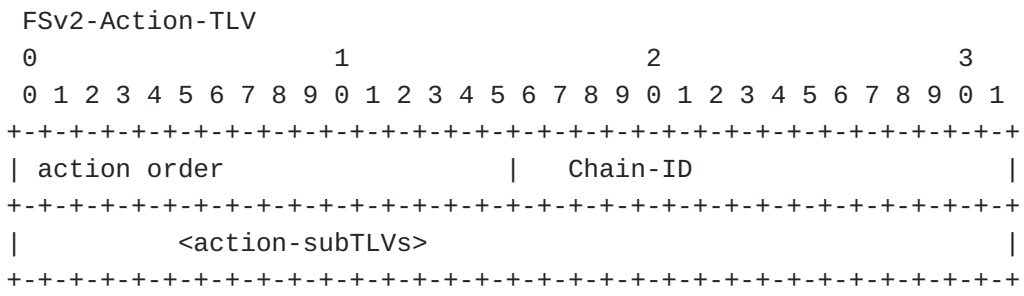


Figure 3-11

Where action-SubTLVs have the format:

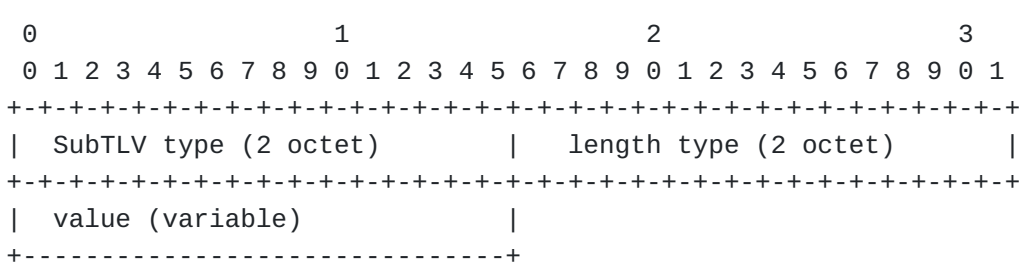


Figure 3-12

where:

action-order: is the user defined order of the action within the list

chain ID: is a 2-byte identifier for an action chain

length - is the length of the TLV

value - contains a sequence of action SubTLVs.

Each Action SubTLV has the format:

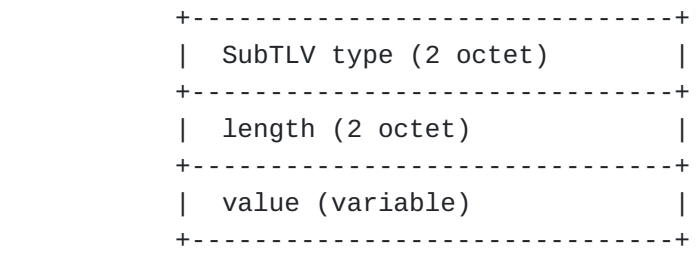


Figure 3-14

Where:

*SubTLV type: values are action type values shown in Table 4 below.

*length: is the length of the action SubTLV

*Value is specific to the SubTLV

Table 4 – FSv2 Action types

Action	Description
00	reserved
01	ACO: action chain operation
02	TAIS: traffic actions per interface group
06	TRB: traffic rate limited by bytes
07	TA: traffic action (terminal/sample)
08	RDIP: Redirect IPv4
09	TM: mark DSCP value
10	TBA (to be assigned)
11	TBA (to be assigned)
12	TRP: traffic rate limited by packets
13	RDIPv6: redirect to IPv6
14	TISFC: SFC Classifier Info (moved from OD to OE)
15	RDIID: redirect to Indirection-id (move from 0x00)
16	MPLSLA: MPLS label action
17-21	TBA (to be assigned)
22	VLAN: VLAN-Action (0x16)[draft-ietf-idr-flowspec-l2vpn-17]
23	TPID: TPID-Action (0x17)[draft-ietf-idr-flowspec-l2vpn-17]
24-254	TBA (to be assigned)
255	reserved

Figure 3-15

Ordering of actions within a rule:

The actions are first stored in user-defined order. If multiple actions exist for a single action order value, then the actions will be ordered by action type followed by value.

Action specifications must include descriptions of order comparison for the values within the action.

3.2.1. Action Chain operation (ACO) (1, 0x01)

SubTLV: 0x01

Length: variable

Value:

AC-failure-type - byte that determines the action on failure

AC-failure-value - variable depending on AC-failure-type.

Actions may succeed or fail and an Action chain must deal with it. The default value stored for an action chain that does not have this action chain is "stop on failure".

where:

AC-Failure types are:

-0x00 - default - stop on failure

-0x01 - continue on failure (best effort on actions)

-0x02 - conditional stop on failure - depending on AC-Failure-value

-0x03 - rollback - do all or nothing - depending in AC-Failure-value

AC-Failure values: TBD

Interactions with other actions: Interactions with all other Actions

Ordering within Action type: By AC-Failure type

3.2.2. Traffic Actions per interface set (TAIS) (2, 0x02)

SubTLV: 0x02

Length: 8 octets (6 in extended community)

Value field: [4-octet-AS] [GroupID 2-octet] [action 2-octet]

where:

Group-ID: identifier for group in 2 octets (14 lower bits)

-Note: Extended Community format will have 2 bits for action.

Action: determines inbound or outbound action where:

-Outbound(0x1): FSv2 rule MUST be applied in outbound Direction to interface set identified by Group-ID.

-Inbound (0x2): FSv2 rule MUST be applied in inbound Direction to interface set identified by Group-ID.

Value ordering: AS, then Group ID, then Action bytes.

Conflict: with any bi-direction action such as

1. traffic rate limited by bytes, or

2. traffic rate limited by packets.

Reference: [I-D.ietf-idr-flowspec-interfaceset]

3.2.3. Traffic rate limited by bytes (TRB) (6, 0x06)

SubTLV:0x06

Length: 8 octets

Value field:[4-octet-AS] [float (4 bytes)]

where:

[4-octet-AS]:4 byte AS number

-If FSv1 passes the lower 2 bytes of 4 byte AS number, use [TBD6] as higher 2 bytes to identify.

-Open issue : TBD6 can be either be chosen to match the common 2-byte to 4-byte or a unique value. Feedback is needed from WG and authors.

Float: maximum byte rate in IEEE 32-bit floating point [IEEE. 754.19895 format] in bytes per second.

-A value of 0 should result in all traffic for the particular flow to be discarded.

-On encoding the traffic-rate-packets MUST NOT be negative.

-On decoding, negative values MUST BE treated as zero (discard all traffic).

Value ordering: AS then float value

Action Conflict: traffic-rate-packets

reference: [RFC8955]

3.2.4. Traffic Action (TA)(7, 0x07)

SubTLV: 0x07

Length: 1

Value field: [1-octet action]

where the traffic action values are:

1 = Terminal flow specification action

2 = Sample - enables sampling and logging

3 = Terminal action + sample

Value ordering: By traffic action values

Conflicts/Interactions: duplication of packets also occurs in:

Redirect to IPv4 (action 0x08),

Redirect to IPv6 (action 0x0D (13)),

Redirect to SFC (action 0xOE (14))

Redirect to Indirection-ID (action 0xF (15)

3.2.5. Redirect to IPv4 (RDIPv4)(8,0x08)

SubTLV: 0x08

Length: 12 octets

Value field:

[4-byte-AS] [IPv4 address (4 octets] [ID (4 octets)] [Flag (1 octet)]

where:

4-octet-AS - is a 4-byte AS in a Route Target

IPv4 address - is an IP Address in RT

ID - the 4-octet value set by user

Flag is 1 octet value with the following definitions:

-0 - reserved

-1 - copy and redirect copy

Value ordering: 4-octet AS, then IP address, then ID (lowest to highest) with:

No AS specified uses AS value of zero.

No IP specified uses IP value of zero.

No ID specified uses ID value of zero.

Conflicts/Interactions: Any redirection or traffic sampling found in:

Traffic Action (action 0x07),

Redirect to IPv6 (action 0x0D (13)),

Redirect to SFC (action 0xOE (14))

Redirect to Indirection-ID (action 0xF (15)

reference: [RFC8955], draft-ietf-idr-flowspec-ip-02.txt

3.2.6. Traffic marking (TM) (9, 0x09)

SubTLV: 0x09

Length: 1 octet

Value: DSCP field with the 2 left most bits zero

The DSCP field format is:

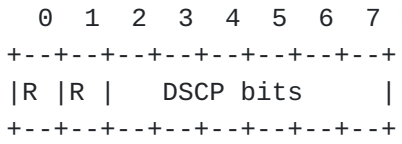


Figure 3-16

where:

R - reserved bits (set to zero to send, ignored upon reception and set to zero.

DSCP - 6 bits of DSCP values

Ordering within Value: Based on DSCP value

Conflicts: none

reference: [RFC8955]

3.2.7. Traffic rate limited by packets (TRP) (12, 0xC)

SubTLV:12 (0xC)

Length: 8

Value field: [4-octet-AS] [float (4 octet)]

Where:

4-octet AS - is the AS setting this value

Float - specifies maximum rate in IEEE 32-bit format [IEEE. 754.185] in packets per second.

- -A traffic rate of zero should result in all packets being discard.
- -On encoding the traffic-rate-packets MUST NOT be negative.
- -On decoding, negative values MUST BE treated as zero (discard all traffic).

Ordering within Value: Based on DSCP value

Conflicts: Traffic rate limited by bytes (0x06)

reference: [RFC8955]

3.2.8. Traffic redirect to IPv6 (RDIPv6) (13, 0xD)

SubTLV: 13 (0xD)

Length: 24 octets

Value field: [4-octet-as] [IPv6-address (16 octets)] [local administrator (2 octets] [Flag (1 octets)]

where:

4-octet-AS - is the AS requesting action in 4-byte AS format,

IPv6-address - is the redirection IPv6 address

Local administrator - 2 bytes assigned by network administrator.

lag (1 octet) with the following definitions:

-0 - reserved

-1 - copy and redirect copy

Ordering within Value: AS, then IPv6, the flag (low to high)

Conflicts/Interactions: Any redirection or traffic sampling found in:

Traffic Action (action 0x07) ,

Redirect to IPv4 (action 0x08 (8)),

Redirect to SFC (action 0xOE (14))

Redirect to Indirection-ID (action 0xF (15)

3.2.9. Traffic insertion in SFC (TISFC)(14, 0xE)

SubTLV:14 (0xE)

Note: replace IANA 0xD FSv1 with FSv2 OxE.

Length: 6 octets

Value field: [SPI (3 octets)][SI (1 octet)][SFT (2 octet)]

where:

SPI - is the service path identifier

SI - is the service index

SFT - is the service function type.

Value ordering: SPI, then SI, then SFT (lowest to highest)

Conflicts/Interactions: Any redirection or traffic sampling found in:

Traffic Action (action 0x07) ,

Redirect to IPv4 (action 0x08 (8)),

Redirect to IPv6 (action 0xOD (13))

Redirect to Indirection-ID (action 0xF (15)

Reference: [RFC9015]

3.2.10. Flow Specification Redirect to Indirection-ID (RDIID) (15, 0x0F)

SubTLV: 15 (0x0F)

note: current value is 0x00 for FSv1

Length: 6 octets

Value field:

[Flags (1 octet)] [ID-Type (1 octet)][Generalized-ID (4 octets)]

Figure 3-17

where:

Flags: are defined as:

-[S-ID]: sequence number for indirection IDs (3 bits).

oValue of zero means sequence is not set and all other S-ID values should be ignored

-[C] - copy packets matching this ID

ID-Type: type of indirection ID with following values:

-0 - localized ID

-1 - Node with SID/index in MPLS SR

-2 - Node with SID/label in MPLS SR

-3 - Node with Binding Segment ID with SID/Index

-4 - Node with Binding Segment ID with SID/Label

-5 - Tunnel ID

Generalized-ID (G-ID): indirection value

Value Ordering: first indirection ID, then Generalized ID

Action Value ordering: ID-Type by value (lowest to highest)

Conflicts/Interactions: Any redirection or traffic sampling found in:

Traffic Action (action 0x07),

Redirect to IPv4 (action 0x08 (8)),

Redirect to IPv6 (action 0x0D, (13)

Redirect to SFC (action 0xOE (14))

reference: [I-D.ietf-idr-flowspec-path-redirect]

3.2.11. MPLS Label Action (MPLSLA)(16, 0x10)

Function: MPLS Label actions

SubTLV: 16 (0x10)

Length: 6 octets

Value:

[action (1 octet)

[order (1 octet)

[Label Stack Entry (4 octets)]

where Action:

Action	Function
0	Push the MPLS tag
1	Pop the outermost MPLS tag in the packet
2	Swap the MPLS tag with the outermost MPLS tag in the packet
3~15	Reserved

Figure 3-18

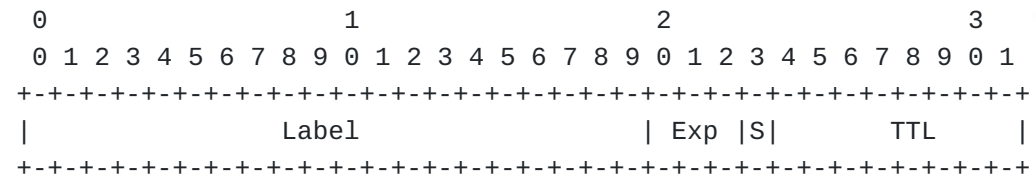


Figure 3-19 - Label Stack Entry

Action Value ordering: ID-Type, then value (lowest to highest)

Value Ordering: order, action, label, Exp

Conflicts/Interactions: Any redirection for IP before MPLS

Traffic Action (action 0x07),

Redirect to IPv4 (action 0x08 (8)),

Redirect to IPv6 (action 0x0D, (13)

Redirect to SFC (action 0xOE (14))

reference: [I-D.ietf-idr-bgp-flowspec-label]

3.2.12. VLAN action (VLAN) (22, 0x16)

Function: Rewrite inner or outer VLAN header

SubTLV: 22 (0x16)

Length: 6 octets

Value:

[Rewrite-actions (2 octets)]

[vlan-PCP-DE-1 (2 octets)]

[vlan-PCP-DE-2 [2 octets)]

where:

Rewrite-actions - is as follows:

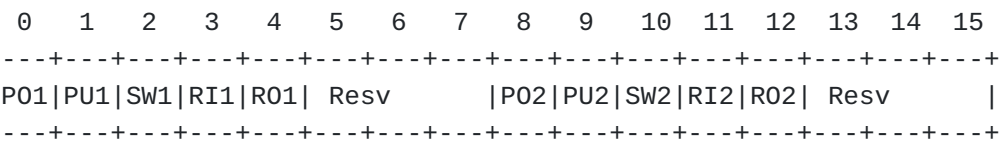


Figure 3-20

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

PU1: Push action. If PU1 is one, it indicates VLAN ID1 will be added, the associated Priority Code Point (PCP and Drop Eligibility Indicator (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 outermost 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 RIx flag is one where "x" is "1" or "2"), it indicates the inner VLAN should be replaced by a new VLAN where the new VLAN is VLAN IDx and the associated PCP and DEI are PCPx and DEx. If the VLAN IDx 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 ROx flag is one (where "x" is "1" or "2"), it indicates the outer VLAN should be replaced by a new VLAN where the new VLAN is VLAN IDx and the associated PCP and DEI are PCPx and DEx. If the VLAN IDx is 0, the action is to only modify the PCP and DEI value of the outer VLAN.

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

Value ordering: rewrite-actions, VLAN1, VLAN2, PCP-DE1, PCP-DE2

Conflicts: TIPD Action

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

3.2.13. TPID action (TPID) (23, 0x17)

Function: Replace Inner or outer TP

SubTLV: 23 (0x17)

Length: 6 octets

Value:

[Rewrite-actions (2 octets)]

[TP-ID-1 (2 octets)]

[TP-ID-2 (2 octets)]

Where: rewrite-actions are bitmask (2 octets) with 2 actions as follows:

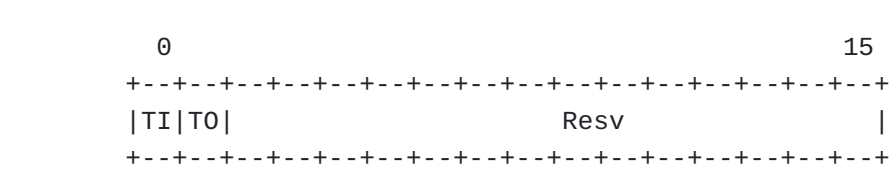


Figure 3-21

TI: Mapping inner Tag Protocol (TP) ID (typically a VLAN) 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

Value Ordering: rewrite-actions, TP-ID-1, TP-ID-2

Conflicts: VLAN action

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

3.3. Extended Community vs. Action SubTLV formats

The SubTLV format is used for the Wide communities and for the action subTLVs in the NLRI.

Sub-TLV type	Action Name	Action SubTLV format	Extended Community format
1	ACO	type: 1 (0x01) length:variable	not applicable (n/a)
2	TAIS	type: 2 (0x02) length:8 [4-octet-as] [group-3-octet] [flags-1-octet]	type: 0x0702 or 0x4702 length: 6 [4-octet-AS] [flags-group] (2 octets)
3-5	reserved		

Sub-TLV type	Action Name	Action SubTLV format	Extended Community format
6	TRB	type:6 (0x06) length:8 [4-byte-AS] [float (4 octets)]	type:8006 length: 6 octets [2-byte-AS] [float (4 octets)]
7	TA	type:7 length:1 flags: (1 octet)	type:8007 length:6 octets flags (6 octets)
8	RDIPv4	type:8 length: 12 [4-byte-AS] [IPv4-address]	type:8008 length: 6 octets [AS-2-octets] [IPv4 address] type:8108 length: 6 octets [IPv4 address] [ID-2 octets] type:8208 length: 6 octets [AS-4-octets] [ID-2-octets]
9	TM	type:9 length:1 DSCP: 1 octet	type:8009 length: 6 octets DSCP: 1 octet
10		type:10 (0X0A)	TBA

11		type:11 (0x0B)	TBA
12	TRP	type:12 (0x0C) length: 8 octets [4-byte-AS] [float-4-octet]	type: 0x800C length: 6 octets [2-byte-AS] [float-4-octet]
13	RDIPv6	type:13 (0x0D) length:22 [4-byte-AS] [IPv6-address (16)] [local-admin (2)]	type:0x000C length: 18 octets [IPv6-address (16)] [local-admin (2)]

Sub-TLV type	Action Name	Action SubTLV format	Extended Community format
14	TISFC	type:14 (0x0E) length:6 SPI (3 octets) SI (1 octet) SFT (2 octets)	type: 0xD (FSv1) type: 0xE (FSv2) length:6 SPI (3 octets) SI (1 octet) SFT (2 octets)
15	RDIID	type:15 (0x0F) length: 6 flags (1) ID-type (1) G-ID (4 octets)	type: 0900 (FSv1) length 6 flags (1) ID type (1) G-ID (4-octets)
16	MPLSLA	type:16 (0x10)	
16-21	TBA	-	
22	VLAN	type:22 (0x16) length:6 [rewrite-action(2)] [vlan-pcp-de-1 (2)] [vlan-pcp-de-2 (2)]	Type: (TBD) length:6 [rewrite-actions (2)] [vlan-pcp-de-1 (2)] [vlan-pcp-de-2 (2)]
23	TPID	type:23 (0x17) length:6 [rewrite-action(2)] [TP-ID-1 (2)] [TP-ID-2 (2)]	Type: (TBD) length:6 [rewrite-actions (2)] [TP-ID-1 (2)] [TP-ID-2 (2)]

3.4. L2 Traffic Rules

The format of the L2 header TLV value field is shown in Figure 3-22. The AFI/SAFI field includes the AFI (2 octets), SAFI (1 octet).

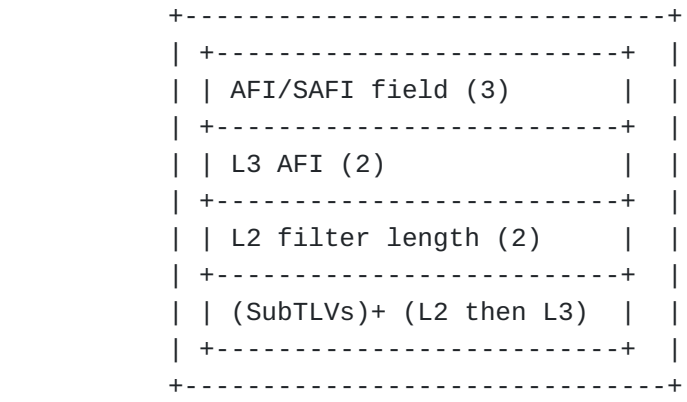


Figure 3-22 -L2 Header TLV value

Where:

AFI/SAFI field has AFI is 6 (IEEE 802) and SAFI is TBD1.

L3 AFI is zero if the filter test only L2 fields, otherwise it is or 2 depending on whether the filter L3 tests after the L2 header are for IPv4 or IPv6.

L2 filter length is the length of the L2 SubTLVs in bytes. These are followed by the L3 SubTLVs is the L3 AFI field is non-zero.

Each L2 SubTLV has the format shown in Figure 3-23. (The L3 SubTLVs are as defined in Section 4.1.)

Each SubTLV has the format:

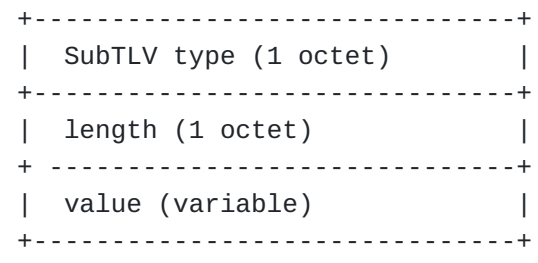


Figure 3-23

SubTLV type: A component type value defined in the "L2 Flow Specification Component Types" registry for L2 by [draft-ietf-idr-flowspec-l2vpn).

Where the SubTLVs have the following component types:

Component Types Table

Component type	Description
1	EtherType
2	Source MAC
3	Destination MAC
4	DSAP (destination service access point)
5	SSAP (source service access point)
6	control field in LLC
7	SNAP
8	VLAN ID
9	VPAN PCP
10	Inner VLAN ID
11	Inner VLAN PCP
12	VLAN DEI
13	VLAN DEI
14	Source MAC special bits
15	Destination MAC special bits

Table 4 – L2 VPN components

See [I-D.ietf-idr-flowspec-l2vpn] for the details on the format and value fields for each component.

Value ordering: Ordering of L2 FSv2 rules will be by user-defined order of the rule. For FSv2 filters within the same rule, the ordering will be by component number and then by value within the component. See [I-D.ietf-idr-flowspec-l2vpn] for the ordering of the values within the component.

L2 VPN filtering using SAFI TBD2 is specified in section 3.6.

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

3.5. SFC Traffic Rules

The FSv2 filters allow for filtering of the SFC NLRI family of routes. The traffic NLRIs filtered are from SFC AFI/SAFI (AFI = 31, SAFI=9).

The FSv2 filters provide this filtering with SFC AFI (AFI=31) and SAFI for FSv2 filters (SAFI = TB1).

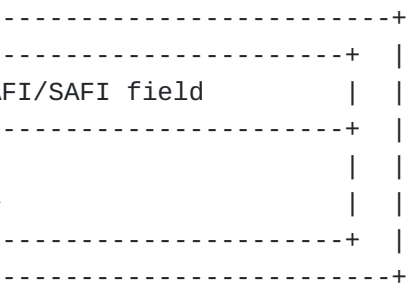


Figure 3-24

Each SubTLV has the format:

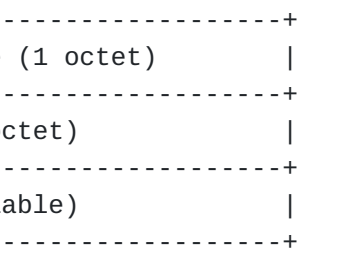


Figure 3-25 – Tunneled SubTLV format

The components listed are:

1 = SFIR RD Type (types 1, 2, 3)
2 = SFIR RD Value
3 = SFIR Pool ID
4 = SFIR MPLS context/label
5 = SFPR SPI
6 = SPF attribute fields

Table 6 – SFC Filter types

Ordering is by: User-defined rule order, component number, and then value within component.

reference: [RFC9015], [TBD]

3.6. BGP/MPLS VPN IP Traffic Rules

The format of the match filter for BGP/MPLS VPN IP traffic is very similar to the format for non-VPN IP traffic as defined in Section 3.1 except that the SAFI is TBD2 and the initial NLRI header has an 8-byte Route Distinguisher added to it as shown in Figure 3-26. The SubTLV format and filter components formats remain the same.

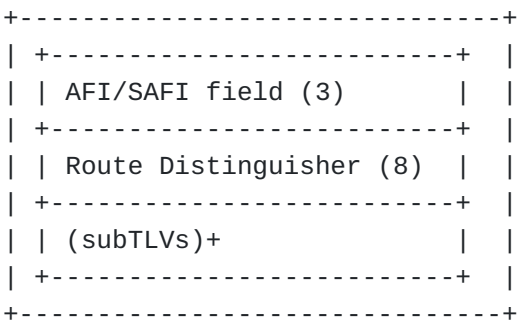


Figure 3-26: VPN IP Filter Header

3.7. BGP/MPLS VPN L2 Traffic Rules

The format of the match filter for BGP/MPLS VPN IP traffic is very similar to the format for non-VPN L2 traffic as defined in Section 3.4 except that the SAFI is TBD2 and the initial NLRI header has an 8-byte Route Distinguisher added to it right after the AFI/SAFI as shown in Figure 3-27 The SubTLV format and filter components formats remain the same.

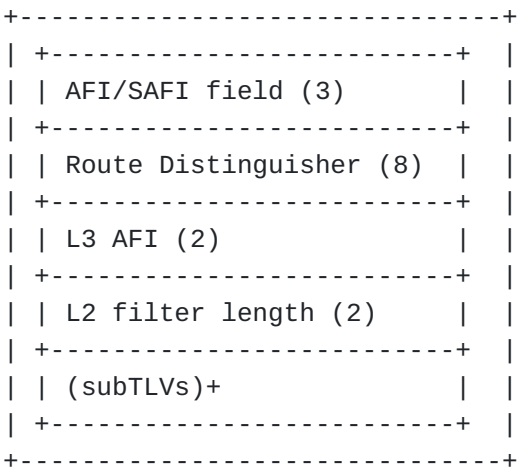


Figure 3-27: VPN L2 Filter Header

3.8. Encoding of Actions passed in Wide Communities

The BGP FSv2 actions are passed in a Wide Community attribute with a BGP Wide Community container (type 01) [I-D.ietf-idr-wide-bgp-communities] with community of FSv2 Actions (TBD4) and Wide Community attributes of Target TLV, Exclude TLVs, and Parameter TLVs. The Parameter MUST contains an FSv2 Atom which contains a sequence of Action TLVs.

BGP Wide Community Container
with FSv2 actions

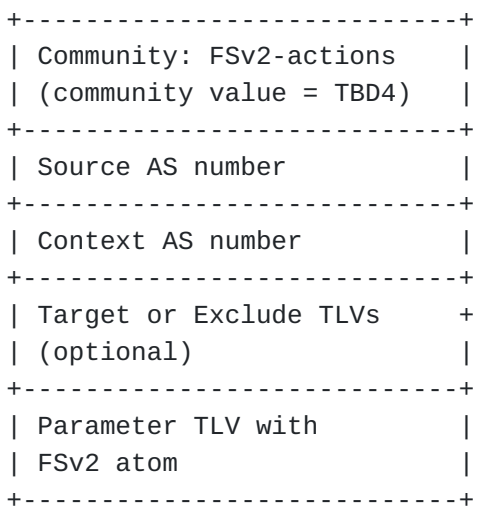


figure 3-28 - BGP

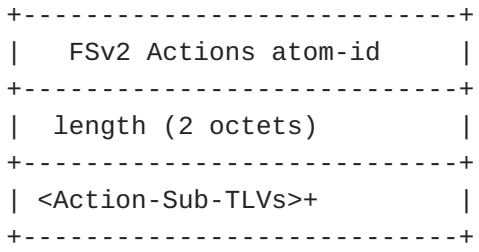


Figure 3-29 - Flow Specification
with IDs for Wide Community Actions

where:

Atom-id: (TBD5)

length: variable depending on SubTLVS

Action Sub-TLVs as defined above

4. Validation of FSv2 NLRI

The validation of FSv2 NLRI adheres to the combination of rules for general BGP FSv1 NLRI found in [RFC8955], [RFC8956], [RFC9117], and the specific additions made for SFC NLRI [RFC9015], and L2VPN NLRI [I-D.ietf-idr-flowspec-l2vpn].

To provide clarity, the full validation process for flow specification routes (FSv1 or FSv2) is described in this section rather than simply referring to the relevant portions of these RFCs. Validation only occurs after BGP UPDATE message reception and the

FSv2 NLRI and the path attributes relating to FSv2 (Extended community and Wide Community) have been determined to be well-formed. Any MALFORMED FSv2 NRLI is handled as a "TREAT as WITHDRAW" [RFC7606].

4.1. Validation of FS NLRI (FSv1 or FSv2)

Flow specifications received from a BGP peer that are accepted in the respective Adj-RIB-In are used as input to the route selection process. Although the forwarding attributes of the two routes for tbe same prefix may be the same, BGP is still required to perform its path selection algorithm in order to select the correct set of attributes to advertise.

The first step of the BGP Route selection procedure (section 9.1.2 of [RFC4271] is to exclude from the selection procedure routes that are considered unfeasible. In the context of IP routing information, this is used to validate that the NEXT_HOP Attribute of a given route is resolvable.

The concept can be extended in the case of the Flow Specification NLRI to allow other validation procedures.

The FSv2 validation process validates the FSv2 NLRI with following unicast routes received over the same AFI (1 or 2) but different SAFIs:

- Flow specification routes (FSv1 or FSv2) received over SAFI=133 will be validated against SAFI=1,

- Flow Specification routes (FSv1 or FSv2) received over SAFI=134 will be validated against SAFI=128, and

- Flow Specification routes (FSv1 or FSv2) [AFI =1, 2] received over SAFI=77 will be validated using only the Outer Flow Spec against SAFI = 133.

The FSv2 validates L2 FSv2 NLRI with the following L2 routes received over the same AFI (25), but a different SAFI:

- Flow specification routes (FSv1 or FSv2)received over SAFI=135 are validated against SAFI=128.

In the absence of explicit configuration, a Flow specification NLRI (FSv1 or FSv2) MUST be validated such that it is considered feasible if and only if all of the conditions are true:

a) A destination prefix component is embedded in the Flow Specification,

b) One of the following conditions holds true:

-1. The originator of the Flow Specification matches the originator of the best-math unicast route for the destination prefix embedded in the flow specification (this is the unicast route with the longest possible prefix length covering the destination prefix embedded in the flow specification).

-2. The AS_PATH attribute of the flow specification is empty or contains only an AS_CONFED_SEQUENCE segment [RFC5065].

o2a.This condition should be enabled by default.

o2b.This condition may be disabled by explicit configuration on a BGP Speaker,

o2c.As an extension to this rule, a given non-empty AS_PATH (besides AS_CONFED_SEQUENCE segments) MAY be permitted by policy].

c) There are no "more-specific" unicast routes when compared with the flow destination prefix that have been received from a different neighbor AS than the best-match unicast route, which has been determined in rule b.

However, part of rule a may be relaxed by explicit configuration, permitting Flow Specifications that include no destination prefix component. If such is the case, rules b and c are moot and MUST be disregarded.

By "originator" of a BGP route, we mean either the address of the originator in the ORIGINATOR_ID Attribute [RFC4456] or the source address of the BGP peer, if this path attribute is not present.

A BGP implementation MUST enforce that the AS in the left-most position of the AS_PATH attribute of a Flow Specification Route (FSv1 or FSv2) received via the Exterior Border Gateway Protocol (eBGP) matches the AS in the left-most position of the AS_PATH attribute of the best-match unicast route for the destination prefix embedded in the Flow Specification (FSv1 or FSv2) NLRI.

The best-match unicast route may change over time independently of the Flow Specification NLRI (FSv1 or FSv2). Therefore, a revalidation of the Flow Specification MUST be performed whenever unicast routes change. Revalidation is defined as retesting rules a to c as described above.

4.2. Validation of Flow Specification Actions

Flow Specifications may be mapped to actions using Extended Communities or a Wide Communities. The FSv2 actions in Extended Communities and Wide communities can be associated with large number of NLRIs.

The ordering of precedence for these actions in the case when the user-defined order is the same follows the precedence of the FSv2 NLRI action TLV values (lowest to highest). User-defined order is the same when the order value for action is the same. All Extended Community actions MUST be translated to the user-defined order data format for internal comparison. By default, all Extended Community actions SHOULD be translated to a single value.

Actions may conflict, duplicate, or complement other actions. An example of conflict is the packet rate limiting by byte and by packet. An example of a duplicate is the request to copy or sample a packet under one of the redirect functions (RDIPv4, RDIPv6, RDIID,) Each FSv2 actions in this document defines the potential conflicts or duplications. Specifications for new FSv2 actions outside of this specification MUST specify interactions or conflicts with any FSv2 actions (that appear in this specification or subsequent specifications).

Well-formed syntactically correct actions should be linked to a filtering rule in the order the actions should be taken. If one action in the ordered list fails, the default procedure is for the action process for this rule to stop and flag the error via system management. By explicit configuration, the action processing may continue after errors.

Implementations MAY wish to log the actions taken by FS actions (FSv1 or FSv2).

4.3. Error handling and Validation

The following two error handling rules must be followed by all BGP speakers which support FSv2:

* FSv2 NLRI having TLVs which do not have the correct lengths or syntax must be considered MALFORMED.

* FSv2 NLRIs having TLVs which do not follow the above ordering rules described in section 4.1 MUST be considered as malformed by a BGP FSv2 propagator.

The above two rules prevent any ambiguity that arises from the multiple copies of the same NLRI from multiple BGP FSv2 propagators.

A BGP implementation SHOULD treat such malformed NLRIs as 'Treat-as-withdraw' [RFC7606]

An implementation for a BGP speaker supporting both FSv1 and FSv2 MUST support the error handling for both FSv1 and FSv2.

5. Ordering for Flow Specification v2 (FSv2)

Flow Specification v2 allows the user to order flow specification rules and the actions associated with a rule. Each FSv2 rule has one or more match conditions and one or more actions associated with that match condition.

This section describes how to order FSv2 filters received from a peer prior to transmission to another peer. The same ordering should be used for the ordering of forwarding filtering installed based on only FSv2 filters.

Section 7.0 describes how a BGP peer that supports FSv1 and FSv2 should order the flow specification filters during the installation of these flow specification filters into FIBs or firewall engines in routers.

The BGP distribution of FSv1 NLRI and FSv2 NLRI and their associated path attributes for actions (Wide Communities and Extended Communities) is "ships-in-the-night" forwarding of different AFI/SAFI information. This recommended ordering provides for deterministic ordering of filters sent by the BGP distribution.

5.1. Ordering of FSv2 NLRI Filters

The basic principles regarding ordering of rules are simple:

1) Rule-0 (zero) is defined to be 0/0 with the "permit-all" action

   -BGP peers which do not support flow specification permit traffic for routes received. Rule-0 is defined to be "permit-all" for 0/0 which is the normal case for filtering for routes received by BGP.

   -By configuration option, the "permit-all" may be set to "deny-all" if traffic rules on routers used as BGP must have a "route" AND a firewall filter to allow traffic flow.

2) FSv2 rules are ordered based on the user-defined order numbers specified in the FSv2 NLRI (rules 1-n).

3) If multiple FSv2 NLRI have the same user-defined order, then the filters are ordered by type of FSv2 NRLI filters (see Table

1, section 4) with lowest numerical number have the best precedence.

- For the same user-defined order and the same value for the FSv2 filters type, then the filters are ordered by FSv2 the component type for that FSv2 filter type (see Tables 3-6) with the lowest number having the best precedence.

- For the same user-defined order, the same value of FSv2 Filter Type, and the same value for the component type, then the filters are ordered by value within the component type. Each component type defines value ordering.

- For component types inherited from the FSv1 component types, there are the following two types of comparisons:

  - FSv1 component value comparison for the IP prefix values, compares the length of the two prefixes. If the length is different, the longer prefix has precedence. If the length is the same, the lower IP number has precedence.

  - For all other FSv1 component types, unless specified, the component data is compared using the memcmp() function defined by [ISO_IEC_9899]. For strings with the same length, the lowest string memcmp() value has precedence. For strings of different lengths, the common prefix is compared. If the common string prefix is not equal, then the string with the lowest string prefix has higher precedence. If the common prefix is equal, the longest string is considered to have higher precedence

Notes:

* Since the user can define rules that re-order these value comparisons, this order is arbitrary and set to provide a deterministic default.

5.2. Ordering of the Actions

The FSv2 specification allows for actions to be associated by:

a) a Wide Community path attribute, or

b) an Extended Community path attribute.

Actions may be ordered by user-defined action order number from 1-n (where n is 2**16-2 and the value 2**16-1 is reserved.

Byy default, extended community actions are associated with default order number 32768 [0x8000] or a specific configured value for the FSv2 domain.

Action user-order number zero is defined to have an Action type of "Set Action Chain operation" (ACO) (value 0x01) that defines the default action chain process. For details on "set action chain operation" see section 3.2.1 or section 5.2.1 below.

If the user-defined action number for two actions are the same, then the actions are ordered by FSv2 action types (see Table 3 for a list of action types). If the user-defined action number and the FSv2 action types are the same, then the order must be defined by the FSv2 action.

5.2.1. Action Chain Operation (ACO)

The "Action Chain Operation" (ACO) changes the way the actions after the current action in an action chain are handled after a failure. If no action chain operations are set, then the default action of "stop upon failure" (value 0x00) will be used for the chain.

5.2.1.1. Example 1 - Default ACO

Use Case 1: Rate limit to 600 packets per second

Description: The provider will support 600 packets per second All Packets sampled for reporting purposes and packet streams over 600 packets per second will be dropped.

Suppose BGP Peer A has a

* a Wide Community action with user-defined order 10 with Traffic Sampling

* a Wide Community action with user-defined order 11 from AS 2020 that limits packet-based rate limit of 600 packets per second.

* an Extended Community from AS 2020 that does limits packet-based rate limit of 50 packets per second.

The FSv2 data base would store the following action chain:

* at user-defined action order 10

  - A user action of type 7 (traffic action) with values of Sampling and logging.

*at user-defined action order 11

-a user action type of 12 (packet-based rate limit) with values of AS 2020 and float value for 600 packets per second (pps)

*at user-defined action order 32768 (0x8000) with type 12 and values of A user action of type 12 with values of AS 2020 and float value of 50 packets/second.

Normal action:

The match on the traffic would cause a sample of the traffic (probably with packet rate saved in logging) followed by a rate limit to 600 pps. The Extended community action would further limit the rate to 50 packets per second.

When does the action chain stop?

The default process for the action chain is to stop on failure. If there is no failure, then all three actions would occur. This is probably not what the user wants.

If there is failure at action 10 (sample and log), then there would be no rate limiting per packet (actions 11 and action 32768).

If there is failure at action 11 (rate limit to packet 600), then there would be no rate limiting per packet (action 32768).

The different options for Action chain ordering (ACO) have been worked on with NETCONF/RESTCONF configuration and actions.

5.2.1.2. Example 2: Redirect traffic over limit to processing via SFC

Use case 2: Redirect traffic over limit to processing via SFC.

Description: The normal function is for traffic over the limit to be forwarded for offline processing and reporting to a customer.

Suppose we have the following 4 actions defined for a match:

*Sent Redirect to indirection ID (0x01) with user-defined match 2 attached in wide community,

*Traffic rate limit by bytes (0x07) with user-defined match 1 attached in wide community,

*Traffic sample (0x07) sent in extended community, and

*SF classifier Info (0x0E) sent in extended community.

These 4 filters rate limit a potential DDoS attack by: a) redirect the packet to indirection ID (for slower speed processing), sample to local hardware, and forward the attack traffic via a SFC to a data collection box.

The FSv2 action list for the match would look like this

Action 0: Operation of action chain (0x01) (stop upon failure)

Action 1: Traffic Rate limit by byte (0x07)

Action 2: Redirect to Redirection ID (0x0F)

Action 32768 (0x8000) Traffic Action (0x07) Sample

Action 32768 (0x8000) SFC Classifier: (0xE)

If the redirect to a redirection ID fails, then Traffic Sample and sending the data to an SFC classifier for forwarding via SFC will not happen. The traffic is limited, but not redirect away from the network and a sample sent to DDOS processing via a SFC classifier.

Suppose the following 5 actions were defined for a FSV2 filter:

*Set Action Chain Operation (ACO) (0x01) to continue on failure (ox01) at user-order 2 attached in wide community,

*redirect to indirection ID (0x0F) at user-order 2 attached in wide community,

*traffic rate limit by bytes (0x07)with user-order 1 attached in wide community,

*Traffic sample (0x07) attached via extended community, and

*SFC classifier Info (0x0E) attached in extended community.

The FSv2 action list for the match would look like this:

Action 00: Operation of action chain (0x01) (stop upon failure)

Action 01:Traffic Rate limit by byte (0x07)

Action 02:Set Action Chain Operation (ACO) (0x01) (continue on failure)

Action 02: Redirect to Redirection ID (0F)

Action 32768 (0x8000): Traffic Action (0x07) Sample

Action 32768 (0x8000): SFC classifier (0x0E) forward via SFC [to DDoS classifier]

If the redirect to a redirection ID fails, the action chain will continue on to sample the data and enact SFC classifier actions.

5.2.2. Summary of FSv2 ordering

Operators should use user-defined ordering to clearly specify the actions desired upon a match. The FSv2 actions default ordering is specified to provide deterministic order for actions which have the same user-defined order and same type.

FS Action (lowest value to highest)	Value Order (lowest to highest)
0x01: ACO: Action chain operation	Failure flag
0x02: TAIS:Traffic actions per Interface group	AS, then Group-ID, then Action ID
0x03-0x05 to be assigned	TBD
0x06: TRB: Traffic rate limit by bytes	AS, then float value
0x07: TA: Traffic Action	traffic action value
0x08: RDIP: Redirect to IP	AS, then IP Address, then ID
0x09: TM: Traffic Marking	DSCP value (lowest to highest)
0x0A: AL2: Associated L2 Info.	TBD
0x0B: AET: Associated E-tree Info.	TBD
0x0C: TRP: Traffic Rate limit by bytes	AS, then float value
0x0D: RDIPv6: Traffic Redirect to IPv6	AS, IPv6 value, then local Admin
0x0E: TISFC: Traffic insertion to SFC	SPI, then SI, the SFT
0xOF: Redirect to Indirection-ID	ID-type, then Generalized-ID
0x10: MPLSLA: MPLS Label stack	order, action, label, Exp
0x16 – VLAN action	rewrite-actions, VALN1, VLAN2, PCP-DE1, PCP-DE2
0x17 – TPID action	rewrite actions, TP-ID-1, TP-ID-2

Figure 6-1

6. Ordering of FS filters for BGP Peers support FSv1 and FSv2

FSv2 allows the user to order flow specification rules and the actions associated with a rule. Each FSv2 rule has one or more match conditions and one or more actions associated with each rule.

FSv1 and FSv2 filters are sent as different AFI/SAFI pairs so FSv1 and FSv2 operate as ships-in-the-night. Some BGP peers in an AS may support both FSv1 and FSv2. Other BGP peers may support FSv1 or FSv2. Some BGP will not support FSv1 or FSV2. A coherent flow specification technology must have consistent best practices for ordering the FSv1 and FSv2 filter rules.

One simple rule captures the best practice: Order the FSv1 filters after the FSv2 filter by placing the FSv1 filters after the FSv2 filters.

To operationally make this work, all flow specification filters should be included the same data base with the FSv1 filters being assigned a user- defined order beyond the normal size of FSv2 user-ordered values. A few examples, may help to illustrate this best practice.

Example 1: User ordered numbering - Suppose you might have 1,000 rules for the FSv2 filters. Assign all the FSv1 user defined rules to 1,001 (or better yet 2,000). The FSv1 rules will be ordered by the components and component values.

Example 2: Storage of actions - All FSv1 actions are defined ordered actions in FSv2. Translate your FSv1 actions into FSv2 ordered actions for storing in a common FSv1-FSv2 flow specification data base.

Example 3: Mixed Flow Specification Support -

Suppose an FSv2 peer (BGP Peer A) has the capability to send either FSv1 or FSv2. BGP Peer A peers with BGP Peers B, C, D and E.

BGP Peer B can only send FSv1 routes (NLRI + Extended Community). BGP Peer C can send FSv2 routes (NLRI + path attributes (wide community or extended community or none)). BGP Peer D cannot send any FS routes. BGP E can send FSv2 and FSv1 routes

BGP Peer A sends FSv1 routes in its databases to BGP B. Since the FSv2 NLRI cannot be sent to the FSv1 peer, only the FSv1 NLRI is sent. BGP Peer A sends to BGP C the FSv2 routes in its database (configured or received).

BGP peer A would not send the FSv1 NLRI or FSv2 NLRI to BGP Peer D. The BGP Peer D does not support for these NLRI.

BGP Peer A sends the NLRI for both FSv1 and FSv2 to BGP Peer E.

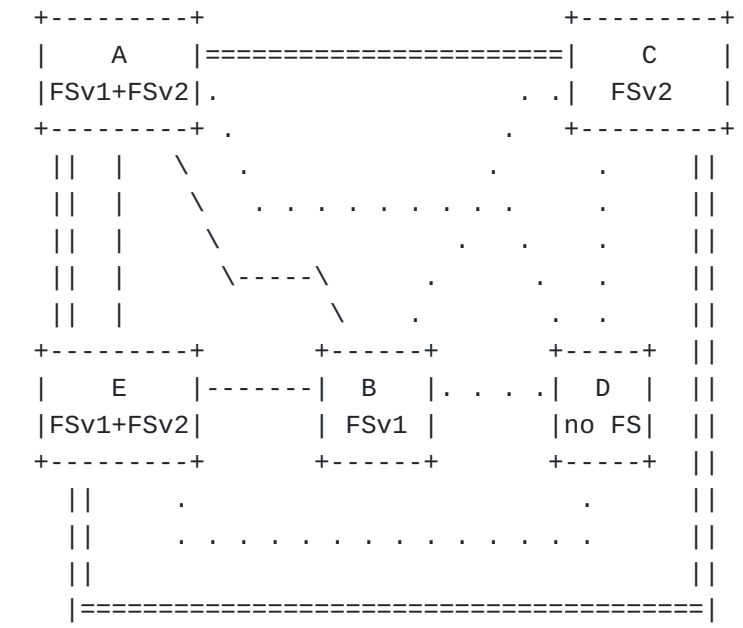


Figure 6-2: FSv1 and FVs2 Peering

7. Scalability and Aspirations for FSv2

Operational issues drive the deployment of BGP flow specification as a quick and scalable way to distribute filters. The early operations accepted the fact validation of the distribution of filter needed to be done outside of the BGP distribution mechanism. Other mechanisms (NETCONF/RESTCONF or PCEP) have reply-request protocols.

These features within BGP have not changed. BGP still does not have an action-reply feature.

NETCONF/RESTCONF latest enhancements provide action/response features which scale. The combination of a quick distribution of filters via BGP and a long-term action in NETCONF/RESTCONF that ask for reporting of the installation of FSv2 filters may provide the best scalability.

The combination of NETCONF/RESTCONF network management protocols and BGP focuses each protocol on the strengths of scalability.

FSv2 will be deployed in webs of BGP peers which have some BGP peers passing FSv1, some BGP peers passing FSv2, some BGP peers passing FSv1 and FSv2, and some BGP peers not passing any routes.

The TLV encoding and deterministic behaviors of FSv2 will not deprecate the need for careful design of the distribution of flow specification filters in this mixed environment. The needs of networks for flow specification are different depending on the network topology and the deployment technology for BGP peers sending flow specification.

Suppose we have a centralized RR connected to DDoS processing sending out flow specification to a second tier of RR who distribute the information to targeted nodes. This type of distribution has one set of needs for FSv2 and the transition from FSv1 to FSv2

Suppose we have Data Center with a 3-tier backbone trying to distribute DDoS or other filters from the spine to combinational nodes, to the leaf BGP nodes. The BGP peers may use RR or normal BGP distribution. This deployment has another set of needs for FSv2 and the transition from FSv1 to FSV2.

Suppose we have a corporate network with a few AS sending DDoS filters using basic BGP from a variety of sites. Perhaps the corporate network will be satisfied with FSv1 for a long time.

These examples are given to indicate that BGP FSv2, like so many BGP protocols, needs to be carefully tuned to aid the mitigation services within the network. This protocol suite starts the migration toward better tools using FSv2, but it does not end it. With FSv2 TLVs and deterministic actions, new operational mechanisms can start to be understood and utilized.

This FSv2 specification is merely the start of a revolution of work - not the end.

8. Optional Security Additions

This section discusses the optional BGP Security additions for BGP-FS v2 relating to BGPSEC [RFC8205] and ROA [RFC6482].

8.1. BGP FSv2 and BGPSEC

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

FSv1 and FSv2 may be sent via BGPSEC.

8.2. BGP FSv2 with ROA

BGP FSv2 can utilize ROAs in the validation. If BGP FSv2 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 as follows:

-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.

9. IANA Considerations

This section complies with [RFC7153].

9.1. Flow Specification V2 SAFIs

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

Value	Description	Reference
TBD1	BGP FSv2	[this document]
TBD2	BGP FSv2 VPN	[this document]

9.2. 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

9.3. Filter IP Component types

IANA is requested to indicate [this draft] as a reference on the following assignments in the Flow Specification Component Types Registry:

Value	Description	Reference
1	Destination filter	[RFC8955][RFC8956][this document]
2	Source Prefix	[RFC8955][RFC8956][this document]
3	IP Protocol	[RFC8955][RFC8956][this document]
4	Port	[RFC8955][RFC8956][this document]
5	Destination Port	[RFC8955][RFC8956][this document]
6	Source Port	[RFC8955][RFC8956][this document]
7	ICMP Type [v4 or v6]	[RFC8955][RFC8956][this document]
8	ICMP Code [v4 or v6]	[RFC8955][RFC8956][this document]
9	TCP Flags [v4]	[RFC8955][RFC8956][this document]
10	Packet Length	[RFC8955][RFC8956][this document]
11	DSCP marking	[RFC8955][RFC8956][this document]
12	Fragment	[RFC8955][RFC8956][this document]
13	Flow Label	[RFC8956][this document]
14	TTL	[this document]
15	Partial SID	[draft-ietf-idr-flowspec-srv6] [this document]
16	MPLS Label Match 1	[this document] [draft-ietf-idr-flowspec-mpls-match]
17	MPLS Label Match 2	[this document] [draft-ietf-idr-flowspec-mpls-match]

9.4. FSV2 NLRI TLV Types

IANA is requested to create the following two new registries on a new "Flow Specification v2 TLV Types" web page.

Name: BGP FSv2 TLV types
Reference: [this document]
Registration Procedures: 0x01-0x3FFF Standards Action.

Type	Use	Reference
0x00	Reserved	[this document]
0x01	IP traffic rules	[this document]
0x02	FSv2 Actions	[this document]
0x03	L2 traffic rules	[this document]
0x04	tunnel traffic rules	[this document]
0x05	SFC AFI filter rules	[this document]
0x06	BGP/MPLS VPN IP traffic rules	[this document]
0x07	BGP/MPLS VPN L2 traffic rules	[this document]
0x08-0x3FFF	Unassigned	[this document]
0x4000-0x7FFF	Vendor specific	[this document]
0x8000-0xFFFF	Reserved	[this document]

Name: BGP FSv2 Action types
Reference: [this document]
Registration Procedure: 0x01-0x3FFF Standards Action.

Type	Use	Reference
0x00	Reserved	[this document]
0x01	ACO: Action Chain Operation	[this document]
0x02	TAIS: Traffic actions per interface group	[this document]
0x03	Unassigned	[this document]
0x04	Unassigned	[this document]
0x05	Unassigned	[this document]
0x06	TRB: traffic rate limited by bytes	[this document]
0x07	TA: Traffic action (terminal/sample)	[this document]
0x08	RDIPv4: redirect IPv4	[this document]
0x09	TM: traffic marking (DSCP)	[this document]
0x0A	AL2: associate L2 Information	[this document]
0x0B	AET: associate E-Tree information	[this document]
0x0C	TRP: traffic rate limited by packets	[this document]
0x0D	RDIPv6: Redirect to IPv6	[this document]
0x0E	TISFC: Traffic insertion to SFC	[this document]
0x0F	RDIID: Redirect to indirection-iD	[this document]
0x10	MPLS Label Action	[this document]
0x11	unassigned	[this document]
0x12	unassigned	[this document]
0x13	unassigned	[this document]
0x14	unassigned	[this document]
0x15	unassigned	[this document]
0x16	VLAN action	[this document]
0x17	TIPD action	[this document]
0x18-0x3ff	Unassigned	[this document]
0x4000-0x7fff	Vendor assigned	[this document]
0x8000-0xFFFF	Reserved	[this document]

9.5. Wide Community Assignments

IANA is requested to assign values in the BGP Community Container Atom Type Registry

Name	Type value
FSv2 action atom	TBD5

IANA is requested to assign values from the Registered Type 1 BGP Wide Community Types:

Name	type Value
FSv2 Actions	TBD4

10. Security Considerations

The use of ROA improves on [RFC8955] by checking to see of the route origination. This check 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 [RFC9117] can provide adequate validation for distribution of flow specification within a 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.

11. References

11.1. Normative References

[I-D.ietf-idr-bgp-flowspec-label] Liang, Q., Hares, S., You, J., Raszuk, R., and D. Ma, "Carrying Label Information for BGP FlowSpec", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-flowspec-label-02, 20 October 2022, <https://datatracker.ietf.org/api/v1/doc/document/draft-ietf-idr-bgp-flowspec-label/>.

[I-D.ietf-idr-flowspec-interfaceset] Litkowski, S., Simpson, A., Patel, K., Haas, J., and L. Yong, "Applying BGP flowspec rules on a specific interface set", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-interfaceset-05, 18 November 2019, <https://www.ietf.org/archive/id/draft-ietf-idr-flowspec-interfaceset-05.txt>.

[I-D.ietf-idr-flowspec-l2vpn]
Weiguo, H., Eastlake, D. E., Litkowski, S., and S. Zhuang, "BGP Dissemination of L2 Flow Specification Rules", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-l2vpn-20, 9 October 2022, <https://www.ietf.org/archive/id/draft-ietf-idr-flowspec-l2vpn-20.txt>.

[I-D.ietf-idr-flowspec-mpls-match] Yong, L., Hares, S., Liang, Q., and J. You, "BGP Flow Specification Filter for MPLS Label", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-mpls-match-02, 20 October 2022, <https://datatracker.ietf.org/api/v1/doc/document/draft-ietf-idr-flowspec-mpls-match/>.

[I-D.ietf-idr-flowspec-nvo3] Eastlake, D., Hao, W., Zhuang, S., Li, Z., and R. Gu, "BGP Dissemination of Flow Specification Rules for Tunneled Traffic", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-nvo3-16, 10 July 2022, <https://www.ietf.org/archive/id/draft-ietf-idr-flowspec-nvo3-16.txt>.

[I-D.ietf-idr-flowspec-path-redirect] Velde, G. V. D., Patel, K., and Z. Li, "Flowspec Indirection-id Redirect", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-path-redirect-11, 26 May 2020, <https://www.ietf.org/archive/id/draft-ietf-idr-flowspec-path-redirect-11.txt>.

[I-D.ietf-idr-flowspec-srv6]
Li, Z., Li, L., Chen, H., Loibl, C., Mishra, G. S., Fan, Y., Zhu, Y., Liu, L., and X. Liu, "BGP Flow Specification for SRv6", Work in Progress, Internet-Draft, draft-ietf-idr-flowspec-srv6-02, 3 October 2022, <https://www.ietf.org/archive/id/draft-ietf-idr-flowspec-srv6-02.txt>.

[I-D.ietf-idr-wide-bgp-communities]
Raszuk, R., Haas, J., Lange, A., Decraene, B., Amante, S., and P. Jakma, "BGP Community Container Attribute", Work in Progress, Internet-Draft, draft-ietf-idr-wide-bgp-communities-08, 11 July 2022, <https://www.ietf.org/archive/id/draft-ietf-idr-wide-bgp-communities-08.txt>.

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, DOI 10.17487/RFC0791, September 1981, <https://www.rfc-editor.org/info/rfc791>.

[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>.

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, DOI 10.17487/RFC3032, January 2001, <https://www.rfc-editor.org/info/rfc3032>.

[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>.

[RFC5065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 5065, DOI 10.17487/RFC5065, August 2007, <https://www.rfc-editor.org/info/rfc5065>.

[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>.

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <https://www.rfc-editor.org/info/rfc7606>.

[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>.

[RFC9015] Farrel, A., Drake, J., Rosen, E., Uttaro, J., and L. Jalil, "BGP Control Plane for the Network Service Header in Service Function Chaining", RFC 9015, DOI 10.17487/RFC9015, June 2021, <https://www.rfc-editor.org/info/rfc9015>.

[RFC9117] Uttaro, J., Alcaide, J., Filsfils, C., Smith, D., and P. Mohapatra, "Revised Validation Procedure for BGP Flow Specifications", RFC 9117, DOI 10.17487/RFC9117, August 2021, <https://www.rfc-editor.org/info/rfc9117>.

11.2. 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
United States of America

Phone: +1-734-604-0332
Email: shares@ndzh.com

Donald Eastlake
Futurewei Technologies

2386 Panoramic Circle
Apopka, FL 32703
United States of America

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

Chaitanya Yadlapalli
ATT
United States of America

Email: cy098d@att.com

Sven Maduschke
Verizon
Germany

Email: sven.maduschke@de.verizon.com