INTERNET-DRAFT
Intended Status: Proposed Standard

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

Expires: July 3, 2023

January 4, 2023

BGP Dissemination of Flow Specification Rules for Tunneled Traffic draft-ietf-idr-flowspec-nvo3-17

Abstract

This draft specifies a Border Gateway Protocol (BGP) Network Layer Reachability Information (NLRI) encoding format for flow specifications (RFC 8955) that can match on a variety of tunneled traffic. In addition, flow specification components are specified for certain tunneling header fields.

Status of This Document

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://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.

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

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

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

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

Table of Contents

1. Introduction....4
1.1 Terminology....4

2. Tunneled Traffic Flow Specification NLRI....6
2.1 The SAFI Code Point....9
2.2 Tunnel Header Component Code Points....9
2.3 Specific Tunnel Types....11
2.3.1 VXLAN....11
2.3.2 VXLAN-GPE....12
2.3.3 NVGRE....13
2.3.4 L2TPv3....13
2.3.4.1 L2TPv3 Data Messages....13
2.3.4.2 L2TPv3 Control Messages....14
2.3.5 GRE....14
2.3.6 IP-in-IP....15
2.3.7 Geneve....15
2.4 Tunneled Traffic Actions....16

3. Order of Traffic Filtering Rules....17
4. Flow Spec Validation....18

5. Security Considerations....18
6. IANA Considerations....19

Normative References....20
Informative References....21

Acknowledgments....22
Authors' Addresses....22

1. Introduction

BGP Flow Specification (flowspec [RFC8955]) is an extension to BGP that supports the dissemination of traffic flow specification rules and actions to be taken on packets matching such specifications. It uses the BGP control plane to simplify the distribution of Access Control Lists (ACLs) and allows new filter rules to be injected to all BGP peers simultaneously without changing router configuration. A typical application of BGP flowspec is to automate the distribution of traffic filter lists to routers for Distributed Denial of Service (DDOS) mitigation.

BGP flowspec defines BGP Network Layer Reachability Information (NLRI) formats used to distribute traffic flow specification rules. AFI=1/SAFI=133 is for IPv4 unicast filtering. AFI=1/SAFI=134 is for IPv4 BGP/MPLS VPN filtering [RFC8955]. [RFC8956] and [FlowSpecL2] extend the flowspec rules for IPv6 and Layer 2 Ethernet packets respectively. None of these previously defined flow specifications are suitable for matching packets in cases of tunneling or encapsulation where there might be duplicates of a layer of header such as two IPv6 headers in IP-in-IP [RFC2003] or a nested header sequence such as the Layer 2 and 3 headers encapsulated in VXLAN [RFC7348].

In the cloud computing era, multi-tenancy has become a core requirement for data centers. It is increasingly common to see tunneled traffic with a field to distinguish tenants. An example is the Network Virtualization Over Layer 3 (NVO3 [RFC8014]) overlay technology that can satisfy multi-tenancy key requirements. VXLAN [RFC7348] and NVGRE [RFC7637] are two typical NVO3 encapsulations. Other encapsulations such as IP-in-IP or GRE may be encountered. Because these tunnel / overlay technologies involving an additional level of encapsulation, flow specification that can match on the inner header as well as the outer header and fields in any tunneling header are needed.

In summary, Flow Specifications should be able to include inner nested header information as well as fields specific to the type of tunneling in use such as virtual network / tenant ID. This draft specifies methods for accomplishing this using SAFI=77 and a new NLRI encoding. In addition, flow specification components are specified for certain tunneling header fields.

1.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and

"OPTIONAL" in this document are to be interpreted as described in BCP

14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

The reader is assumed to be familiar with BGP terminology [RFC4271] [RFC4760]. The following terms and acronyms are used in this document with the meaning indicated:

ACL - Access Control List

DDoS - Distributed Denial of Service (Attack)

DSCP - Differentiated Services Code Point [RFC2474]

GRE - Generic Router Encapsulation [RFC2890]

L2TPv3 - Layer Two Tunneling Protocol - Version 3 [RFC3931]

NLRI - Network Layer Reachability Information [RFC4271] [RFC4760]

NVGRE - Network Virtualization Using Generic Routing Encapsulation [RFC7637]

NVO3 - Network Virtual Overlay Layer 3 [RFC8014]

PE - Provider Edge

VN - virtual network

VXLAN - Virtual eXtensible Local Area Network [RFC7348]

2. Tunneled Traffic Flow Specification NLRI

The Flowspec rules specified in [RFC8955], [RFC8956], and [FlowSpecL2] cannot match or filter tunneled traffic based on the tunnel type, tunnel header fields, or headers past the tunnel header. To enable flow specification of tunneled traffic, a new SAFI (77) and NLRI encoding are specified. This encoding, shown in Figure 1, enables flow specification of more than one layer of header when needed.

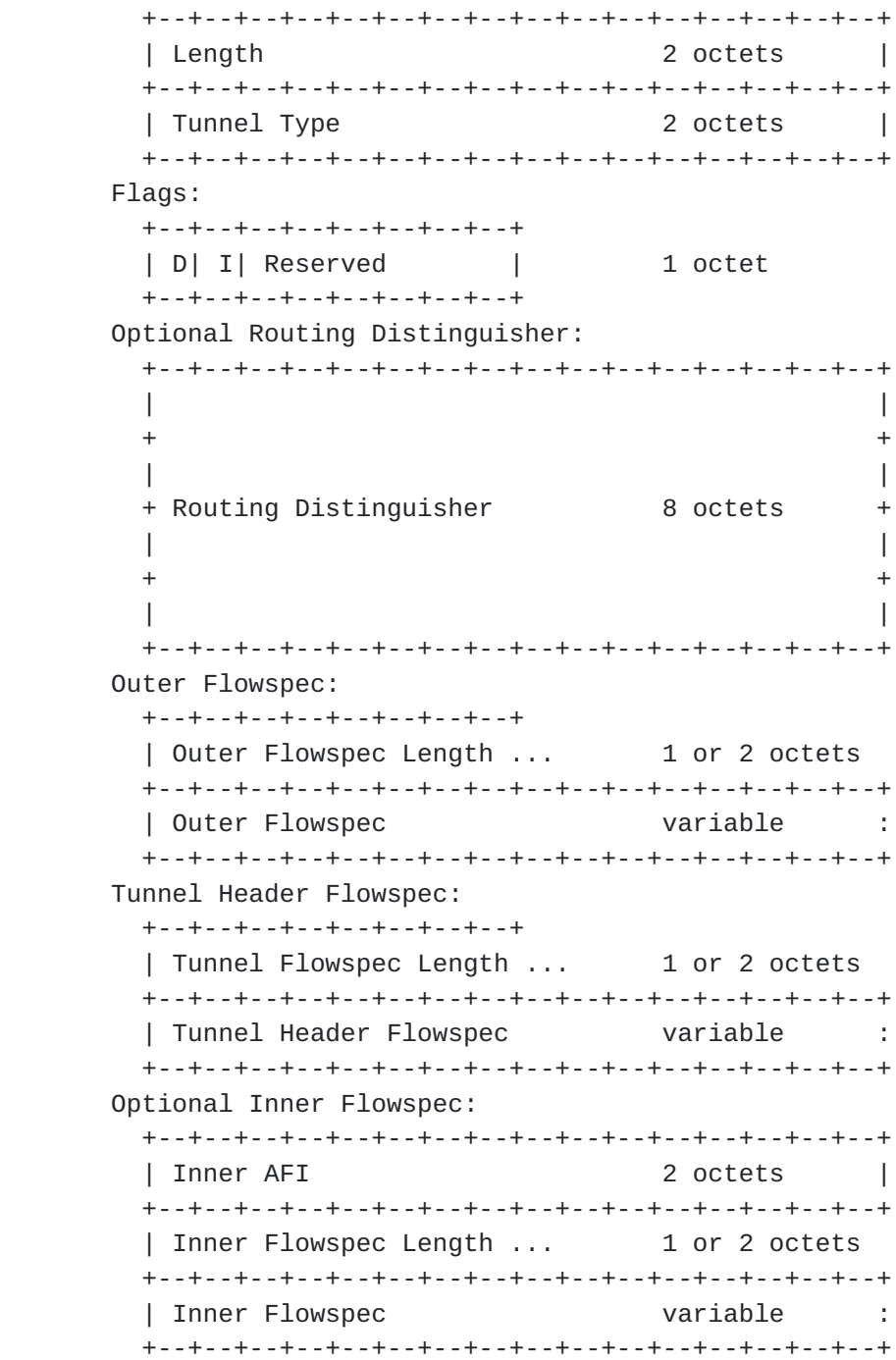


Figure 1. Tunneled Traffic Flowspec NLRI

Length - The NLRI Length encoded as an unsigned integer including the Tunnel Type.

Tunnel Type - The type of tunnel using a value from the IANA BGP Tunnel Encapsulation Attribute Tunnel Types registry.

Flags: D bit - Indicates the presence of the Routing Distinguisher
(see below).

Flags: I bit - Indicates the presence of the Inner AFI and the Inner Flowspec (see below).

Flags: Reserved - Six bits that MUST be sent as zero and ignored on receipt.

Routing Distinguisher - If the outer Layer 3 address belongs to a BGP/MPLS VPN, the routing distinguisher is included to indicate traffic filtering within that VPN. Because NVO3 outer layer addresses normally belong to a public network, a Route Distinguisher field is normally not needed for NVO3.

Outer Flowspec / Length - The flow specification for the outer header. The length is encoded as provided in Section 4.1 of [RFC8955]. The AFI for the Outer Flowspec is the AFI at the beginning of the BGP multiprotocol MP_REACH_NLRI or MP_UNREACH_NLRI containing the tunneled traffic flow specification NLRI.

Tunnel Header Flowspec / Length - The flow specification for the tunneling header. The length is encoded as provided in Section 4.1 of [RFC8955]. This specifies matching criterion on tunnel header fields as well as, implicitly, on the tunnel type which is indicated by the Tunnel Type field above. For some types of tunneling, such as IP-in-IP, there may be no tunnel header fields. For other types of tunneling, there may be several tunnel header fields on which matching can be specified with this flowspec. If a Tunnel Type has no tunnel header fields or it is not desired to filter on header fields, the Tunnel Flowspec length field is present but has value zero.

Inner AFI - Depending on the Tunnel Type, there may be an Inner AFI that indicate the type of inner flow specification. The "Inner SAFI" is implicitly 133 for flowspec.

Inner Flowspec / Length - Depending on the Tunnel Type, there may be an inner flowspec for the header level encapsulated within the outer header. The length is encoded as provided in Section 4.1 of [RFC8955].

A Tunneled Traffic Flowspec matches if the Outer Flowspec, Tunnel Type, and Tunnel Header Flowspec match and, in addition, each of the following optional items matches if it is present:

- Inner Flowspec, and
- Routing Distinguisher.

An omitted (as can be done for the Inner Flowspec) or null flowspec is considered to always match.

2.1 The SAFI Code Point

Use of the tunneled traffic flow specification NLRI format is indicated by SAFI=77. This is used in conjunction with the AFI for the outer header, that is AFI=1 for IPv4, AFI=2 for IPv6, and AFI=6 for Layer 2.

2.2 Tunnel Header Component Code Points

For most cases of tunneled traffic, there are tunnel header fields that can be tested by components that appear in the Tunnel Header Flowspec field. The types for these components are specified in a Tunnel Header Flowspec component registry (see Section 6) and the initial entries in this registry are specified below.

All Tunnel Header field components defined below and all such components added in the future have a TLV structure as follows:

- one octet of type followed by
- one octet giving the length of the value part as an unsigned integer number of octets followed by
- the specific matching operations/values as determined by the type.

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

Defines a list of {operation, value} pairs used to match the 24-bit VN ID (Virtual Network Identifier) that is used as the tenant identification in some tunneling headers. For VXLAN and Geneve encapsulation, the VN ID field is the VNI. For NVGRE encapsulation, the VN ID is the VSID. op is encoded as specified in Section 4.2.3 of [RFC8955]. Values are encoded as a 1, 2, or 4 octet quantity. If value is 24-bits, it is left-justified in the first 3 octets of the value and the last value octet MUST be sent as zero and ignored on receipt.

Type 2 - Flow ID
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match 8-bit Flow ID fields which are currently only useful for NVGRE encapsulation. op is encoded as specified in Section 4.2.3 of [RFC8955]. Values are encoded as a 1-octet quantity.

Type 3 - Session
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match a 32-bit Session field. This field is called Key in GRE [RFC2890] encapsulation and Session ID in L2TPv3 encapsulation. op is encoded as specified in Section 4.2.3 of [RFC8955]. Values are encoded as a 1, 2, or 4 octet quantity; if 1 or 2 octets are provided, these are right justified and padded on the left with zeros.

Type 4 - Cookie
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match a variable length Cookie field. This is only useful in L2TPv3 encapsulation. op is encoded as specified in Section 4.2.3 of [RFC8955]. Values are encoded as a 1, 2, 4, or 8 octet quantity. If the Cookie does not fit exactly into the value length, it is left justified and padded with following octets that MUST be sent as zero and ignored on receipt.

Type 5 - Tunnel Header Flags
Encoding: <type (1 octet), length (1 octet), [op, bitmask]+>

Defines a list of {operation, bitmask} pairs used to match against the tunnel header flags field. op is encoded as in Section 4.2.9 of [RFC8955]. bitmask is encoded as 1 octet for VXLAN-GPE and Geneve and as 2 octets for L2TPv3 control messages. When matching on L2TPv3 control message flags, the 3-bit Version subfield is treated as if it was zero.

Type 6 - L2TP Control Version
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match against the L2TP Control Message Version. op is encoded as in Section 4.2.3 of [RFC8955]. Value is encoded as 1 octet.

Type 7 - L2TPv3 Control Connection ID
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match against the L2TPv3 Control Connection ID. op is encoded as in Section 4.2.3 of [RFC8955]. Value is encoded as 4 octets.

Type 8 - L2TPv3 Ns
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match against the L2TPv3 control message Ns field. op is encoded as in Section 4.2.3 of [RFC8955]. Value is encoded as 2 octets.

Type 9 - L2TPv3 Nr
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match against the L2TPv3 control message Nr field. op is encoded as in Section 4.2.3 of [RFC8955]. Values are encoded as 2 octets.

Type 10 - Protocol Type
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match against the GRE and Geneve Protocol Type fields. op is encoded as in Section 4.2.3 of [RFC8955]. Values are encoded as 2 octets.

Type 11 - GRE Sequence
Encoding: <type (1 octet), length (1 octet), [op, value]+>

Defines a list of {operation, value} pairs used to match against the GRE Sequence field. op is encoded as in Section 4.2.3 of [RFC8955]. Values are encoded as a 1, 2, or 4 octet quantity; if 1 or 2 octets are provided, these are right justified and padded on the left with zeros.

2.3 Specific Tunnel Types

The following subsections describe how to handle flow specification for several specific tunnel types.

2.3.1 VXLAN

The headers on a VXLAN [RFC7348] data packet are an outer Ethernet header, an outer IP header, a UDP header, the VXLAN header, and an inner Ethernet header. This inner Ethernet header is frequently, but not always, followed by an inner IP header. If the tunnel type is VXLAN, the I flag MUST be set in the Tunneled Traffic Flow Specification.

If the outer Ethernet header is not being matched, the version (IPv4 or IPv6) of the outer IP header is indicated by the AFI at the beginning of the multiprotocol MP_REACH_NLRI or MP_UNREACH_NLRI containing the Tunneled Traffic Flow Specification NLRI. The outer Flowspec is used to filter the outer headers including, if desired, the UDP header.

If the outer Ethernet header is being matched, then the initial AFI

is 6 [FlowSpecL2] and the Outer Flowspec can match the outer Ethernet header, specify the IP version of the outer IP header, and match that IP header including, if desired, the UDP header.

The Tunnel Header Flowspec can be used to filter on the VXLAN header VN ID (VNI).

The Inner Flowspec can be used on the Inner Ethernet header [FlowSpecL2] and any following IP header. If the inner AFI is 6, then the inner Flowspec provides filtering of the Layer 2 header, indicates whether filtering on a following IPv4 or IPv6 header is desired, and if it is desired provides the Flowspec components for that filtering. If the Inner AFI is 1 or 2, the Inner Ethernet header is not matched and to match the Flowspec the Inner Ethernet header must be followed by an IPv4 or IPv6 header, respectively, and the inner Flowspec is used to filter that inner IP header.

The inner MAC/IP address is associated with the VN ID. In the NVO3 terminating into a VPN scenario, if multiple access VN IDs map to one VPN instance, one shared VN ID can be carried in the flowspec rule to enforce the rule on the entire VPN instance and the shared VN ID and VPN correspondence should be configured on each VPN PE beforehand. In this case, the function of the Layer 3 VN ID is the same as a Route Distinguisher: it acts as the identification of the VPN instance.

2.3.2 VXLAN-GPE

VXLAN-GPE [GPE] is similar to VXLAN. The VXLAN-GPE header is the same size as the VXLAN header but has been extended from the VXLAN header by specifying a number of bits that are reserved in the VXLAN header. In particular, a number of additional flag bits are specified and a Next Protocol field is added that is valid if the P flag bit is set in the VXLAN-GPE header. These flags bits can be tested using the Tunnel Header Flags flowspec component defined above. VXLAN and VXLAN-GPE are distinguished by the port number in the UDP header that precedes the VXLAN or VXLAN-GPE headers.

If the VXLAN-GPE header P flag is zero, then that header is followed by the same sequence as for VXLAN and the same flowspec choices apply (see Section 2.3.1).

If the VXLAN-GPE header P flag is one and that header's next protocol field is 1, then the VXLAN-GPE header is followed by an IPv4 header (there is no Inner Ethernet header). The Inner Flowspec matches only if the Inner AFI is 1 and the Inner Flowspec matches.

If the VXLAN-GPE header P flag is one and that header's next protocol

field is 2, then the VXLAN-GPE header is followed by an IPv6 header

(there is no Inner Ethernet header). The Inner Flowspec match only if the Inner AFI is 2 and the Inner Flowspec matches.

2.3.3 NVGRE

NVGRE [RFC7637] is similar to VXLAN except that the UDP header and VXLAN header immediately after the outer IP header are replaced by a GRE (Generic Router Encapsulation) header. The GRE header as used in NVGRE has no Checksum or Reserved1 field as shown in [RFC2890] but there are Virtual Subnet ID and Flow ID fields in place of what is labeled in [RFC2890] as the Key field. Processing and restrictions for NVGRE are as in Section 2.3.1 eliminating references to a UDP header and replacing references to the VXLAN header and its VN ID with references to the GRE header and its VN ID (VSID) and Flow ID.

2.3.4 L2TPv3

The headers on an L2TPv3 [RFC3931] packets are an outer Ethernet header, an outer IP header, the L2TPv3 header, an inner Ethernet header, and possibly an inner IP header if indicated by the inner Ethernet header EtherType. The Outer Flowspec operates on the outer headers that precede the L2TPv3 Session Header. The version of IP in the outer IP header is specified by either the outer AFI at the beginning of the MP_REACH_NLRI or MP_UNREACH_NLRI or, if that AFI is 6 (L2), optionally specified by the inner AFI within that L2 flowspec.

L2TPv3 data messages and control messages both start with a Session ID and are distinguished by whether the Session ID is non-zero or zero, respectively. Data message filtering is further specified in Section 2.3.4.1 and control message filtering is further specified in Section 2.3.4.2.

2.3.4.1 L2TPv3 Data Messages

For data messages, the L2TPv3 Session Header consists of a 32-bit non-zero Session ID followed by a variable length Cookie (maximum length 8 octets). A Tunnel Header flowspec is assumed to apply to data messages unless the first component requires a zero Session ID.

The Session ID and Cookie can be filtered on by using the Session and Cookie flowspec components in the Tunnel Header Flowspec. To filter on Cookie or even be able to bypass Cookie and parse the remainder of

the L2TPv3 packet, the node implementing tunneled traffic flowspec

needs to know the length and/or value of the Cookie fields of interest. This is negotiated at L2TPv3 session establishment and it is out of scope for this document how the node would learn this information. Of course, if flowspec is being used for DDOS mitigation and the Cookie has a fixed length and/or value in the DDOS traffic, this could be learned by inspecting that traffic.

If the I flag bit is zero, then no filtering is done on data beyond the L2TPv3 header. If the I flag is one, indicating the presence of an Inner Flowspec, and the node implementing flowspec does not know the length of the L2TPv3 header Cookie, the match fails. If that node does know the length of that Cookie, the Inner Flowspec if matched against the headers at the beginning of that data using the Inner AFI. If that Inner AFI is 1 or 2, then an inner IP header is required and filtering can be done on that IPv4 or IPv6 header respectively. If the Inner AFI is 6, filtering is done on the inner Ethernet header and, if an IPv4 or IPv6 inner AFI is specified within the inner L2 flowspec, done on the following IP header [FlowSpecL2].

2.3.4.2 L2TPv3 Control Messages

Control messages are distinguished by starting with a zero value 32-bit Session ID. L2TPv3 control message flowspecs MUST start with a Session component that requires Session to be zero. For L2TPv3 control messages, there is no Cookie but there are L2TPv3 flags, a 3-bit Version field, a 32-bit Control Connection ID, and 16-bit Ns and Nr sequence numbers. These can be tested using the Tunnel Header Flags, L2TP Control Version, L2TPv3 Control Connection ID, L2TPv3 Ns, and L2TPv3 Nr flowspec components in the Tunnel Header Flowspec.

2.3.5 GRE

Generic Router Encapsulation (GRE [RFC2890]) is another type of encapsulation. The Outer Flowspec operates on the outer headers that precede the GRE header. The version of IP is specified by the outer AFI at the beginning of the MP_REACH_NLRI or MP_UNREACH_NLRI.

The Tunnel Header Flags component can be used to match the first two octets of the GRE header. The Protocol Type component can be used to match the corresponding GRE header field. The Session and GRE Sequence components can be used to match on the GRE Key and GRE Sequence fields if those fields are present respectively. If either of those fields is not present, a component to match on that field fails.

If the I flag bit is zero, no filtering is done on data after the GRE

header. If the I flag bit is one in the tunnel flowspec, then there is an inner AFI and inner flowspec and the Protocol Type field of the GRE header must correspond to the Inner AFI as follows for the tunnel Flowspec to match. Otherwise, the match fails.

GRE Protocol Type	Inner AFI
0x0800 (IPv4)	1
0x86DD (IPv6)	2
0x6558	6

With the I flag a one and the Inner AFI and GRE Protocol Type fields correspond, the Inner Flowspec is used to filter the inner IP headers (Inner AFI=1 or 2) or the inner Ethernet header and optionally a following IP header (Inner AFI=6).

2.3.6 IP-in-IP

IP-in-IP encapsulation [RFC2003] is indicated when an outer IP header indicates an inner IP IPv4 or IPv6 header by the value of the outer IP header's Protocol (IPv4) or Next Protocol (IPv6) field.

The IP version of the outer IP header (IPv4 or IPv6) matched is indicated by an AFI of 1 or 2 at the beginning of the MP_REACH_NLRI or MP_UNREACH_NLRI while if that AFI is 6, it indicates a match on the outer Ethernet header and, optionally, the following IP Header [FlowSpecL2]. The IP version of the inner IP header is indicated by the Inner AFI and the Inner Flowspec applies to the inner IP header.

There is no tunnel header so there are no fields that can be matched by the Tunnel Header Flowspec in the case of IP-in-IP.

2.3.7 Geneve

The headers on a Geneve [RFC8926] encapsulated packet are an outer Ethernet header, an outer IP header, a UDP header, the Geneve header, and subsequent headers depending on the Geneve header Protocol Type field.

If the outer Ethernet header is not being matched, the version (IPv4 or IPv6) of the outer IP header is indicated by the AFI at the beginning of the multiprotocol MP_REACH_NLRI or MP_UNREACH_NLRI containing the Tunneled Traffic Flow Specification NLRI. The outer Flowspec is used to filter the outer headers including, if desired, the UDP header.

If the outer Ethernet header is being matched, then the initial AFI is 6 [FlowSpecL2] and the Outer Flowspec can match the outer Ethernet header, specify the IP version of the outer IP header, and match that IP header including, if desired, the UDP header.

The Tunnel Header Flowspec can be used to filter on the Protocol Type field and/or the VNI field in the Geneve header. The flags octet of the Geneve header, the second octet of that header, can be filtered using the Tunnel Header Flags component.

If an Inner Flowspec is present, it is used to match the header(s) after the Geneve header. The Protocol Type field in the Geneve header must correspond to the Inner AFI as shown in the table in Section 2.3.5 above or the match fails. If the Inner AFI and GRE Protocol Type fields correspond, the Inner Flowspec is used to filter the inner IP headers (Inner AFI=1 or 2) or the inner Ethernet header and optionally a following IP header (Inner AFI=6).

2.4 Tunneled Traffic Actions

The traffic filtering actions previously specified in [RFC8955] and [FlowSpecL2] are used for tunneled traffic. For Traffic Marking in NVO3, only the DSCP in the outer header can be modified.

3. Order of Traffic Filtering Rules

The following rules determine which flowspec takes precedence where one or more are applicable and at least one of the applicable flowspecs is a tunneled traffic flowspec:

- In comparing an applicable tunneled traffic flow specification with an applicable non-tunneled flow specification, the tunneled specification has precedence.

- If comparing tunneled traffic flow specifications, if all are applicable, the tunnel types will be the same. Any that have a Routing Distinguisher will take precedence over those without a Routing Distinguisher. Of those with a Routing Distinguisher, all applicable flowspecs will have the same Routing Distinguisher.

- At this point in the process, all remaining contenders for the highest precedence will either not have a Routing Distinguisher or have equal Routing Distinguishers. If more than one contender remain, those with an L2 Outer Flowspec take precedence over those with an L3 Outer Flowspec. If the Outer Flowspec AFI is the same, their order of precedence is determined by comparing the Outer Flowspecs as described in [RFC8955] and [RFC8956] for AFI for 1 or 2 respectively or [FlowSpecL2] for AFI=6.

- If the Outer Flowspecs are equal, then the Tunnel Header Flowspecs are compared using the usual sequential component comparison process [RFC8955].

- If the Tunnel Header Flowspecs are equal then compare the "I" flag. Those with an Inner Flowspec take precedence over those without an Inner Flowspec. If you get to this stage in the ordering process, those without an Inner Flowspec are equal. For those with an Inner Flowspec, check the Inner AFI. An L2 Inner AFI (AFI=6) takes precedence over an L3 Inner AFI.

- If the Inner AFIs are equal, precedence is determined by comparing the Inner Flowspecs as described in [FlowSpecL2] for L2 or [RFC8955] for L3.

4. Flow Spec Validation

Flowspecs received over AFI=1/SAFI=77 or AFI=2/SAFI=77 are validated, using only the Outer Flowspec, against routing reachability received over AFI=1/SAFI=133 and AFI=2/SAFI=133 respectively, as modified by [RFC9117].

5. Security Considerations

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

For general Flowspec security considerations, see [RFC8955].

6. IANA Considerations

IANA has assigned the following SAFI:

Value	Description	Reference
77	Tunneled Traffic Flowspec	[This document]

IANA is requested to create a Tunnel Header Flow Spec Component Type registry on the Flow Spec Component Types registries web page as follows:

Name: Tunnel Flow Spec Component Types
Reference: [this document]
Registration Procedures:

0	Reserved
1-127	Specification Required
128-254	First Come First Served
255	Reserved

Initial contents:

Type	Name	Reference
0	reserved	[this document]
1	VN ID	[this document]
2	Flow ID	[this document]
3	Session	[this document]
4	Cookie	[this document]
5	Tunnel Header Flags	[this document]
6	L2TP Control Version	[this document]
7	L2TPv3 Control Connection ID	[this document]
8	L2TPv3 Ns	[this document]
9	L2TPv3 Nr	[this document]
10	Protocol Type	[this document]
11	GRE Sequence	[this document]
12-254	unassigned	[this document]
255	reserved	[this document]

Normative References

[RFC2003] - Perkins, C., "IP Encapsulation within IP", RFC 2003, DOI 10.17487/RFC2003, October 1996, <https://www.rfc-editor.org/info/rfc2003>.

[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>.

[RFC2474] - Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, DOI 10.17487/RFC2474, December 1998, <https://www.rfc-editor.org/info/rfc2474>.

[RFC2890] - Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, DOI 10.17487/RFC2890, September 2000, <https://www.rfc-editor.org/info/rfc2890>.

[RFC3931] - Lau, J., Ed., Townsley, M., Ed., and I. Goyret, Ed., "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, DOI 10.17487/RFC3931, March 2005, <https://www.rfc-editor.org/info/rfc3931>.

[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>.

[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>.

[RFC7348] - Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <https://www.rfc-editor.org/info/rfc7348>.

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

[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>.

[RFC8926] - Gross, J., Ed., Ganga, I., Ed., and T. Sridhar, Ed., "Geneve: Generic Network Virtualization Encapsulation", RFC 8926, DOI 10.17487/RFC8926, November 2020, <https://www.rfc-editor.org/info/rfc8926>.

[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>.

[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>.

[FlowSpecL2] - W. Hao, et al, "Dissemination of Flow Specification Rules for L2 VPN", draft-ietf-idr-flowspec-l2vpn, work in progress.

Informative References

[RFC8014] - Black, D., Hudson, J., Kreeger, L., Lasserre, M., and T. Narten, "An Architecture for Data-Center Network Virtualization over Layer 3 (NVO3)", RFC 8014, DOI 10.17487/RFC8014, December 2016, <https://www.rfc-editor.org/info/rfc8014>.

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

Acknowledgments

The authors wish to acknowledge the important contributions of the following listed in alphabetic order:

Jeff Haas, Susan Hares, Yizhou Li, Qiandeng Liang, Greg Mirsky, Nan Wu, Robert Raszuk, and Lucy Yong

Authors' Addresses

Donald Eastlake
Futurewei Technologies
2386 Panoramic Circle
Apopka, FL 32703 USA

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

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012 China

Email: haoweiguo@huawei.com

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

Email: zhuangshunwan@huawei.com

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

Email: lizhenbin@huawei.com

Rong Gu
China Mobile

Email: gurong_cmcc@outlook.com

Copyright, Disclaimer, and Additional IPR Provisions

Copyright (c) 2023 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 (http://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.