Internet Engineering Task Force
Internet-Draft
Updates: 5575 (if approved)
Intended status: Standards Track
Expires: February 23, 2017

C. Loibl
Next Layer Communications
M. Bacher
T-Mobile Austria
August 22, 2016

Flowspec Clarification
draft-loibl-bacher-idr-flowspec-clarification-00

Abstract

This document clarifies multiple aspects of Flowspec (RFC 5575) to allow a consistent and robust implementation in an multi vendor / inter provider environment.

This document updates RFC 5575.

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 http://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 February 23, 2017.

Copyright Notice

Table of Contents

1. Introduction

This document clarifies multiple aspects of Flowspec (RFC 5575) to allow a consistent and robust implementation in a multi vendor / inter autonomous system environment. It describes only minimal changes in the behaviour defined by RFC 5575 [RFC5575] but clarifies many of the unclear sections of RFC 5575.

During a large interoperability lab session involving multiple routing equipment vendors (Alcatel, Cisco, Huawei, Juniper) we identified many incompatibilities in their Flowspec implementations. The identified incompatibilities range from valid Flowspec network layer reachability information (NLRI) updates being ignored by BGP speakers (and not forwarded according to the normal BGP update mechanisms) to the inability to parse valid Flowspec NLRIs causing BGP notifications sent to neighbors and sessions being closed, thus leading to partial or complete network outages. Most of the incompatibilities found during those tests were results of unclear or missing definitions in the flowspec RFC 5575.

2. Terminology

AS - Autonomous System

BGP - Border Gateway Protocol

DSCP - Diffserv Code Point

ICMP - Internet Control Messaging Protocol

NLRI - Network Layer Reachability Information

VPN - Virtual Private Network

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 RFC 2119 [RFC2119].

3. Clarification of the Comparison Operator

RFC 5575 section 4 under Type 3 defines the encoding of {operator, value} pairs where the operator bits 5-7 contain comparison flags (<, >, =). The following table obsoletes the last paragraph of RFC 5575 section 4 Type 3 ("The bits lt, gt, and eq can be combined to produce "less or equal", "greater or equal", and inequality values."). Combinations of the bits MUST be interpreted as follows:

<	>	=	Resulting operation
0	0	0	true (independent of the value)
0	0	1	== (equal)
0	1	0	> (greater than)
0	1	1	>= (greater than or equal)
1	0	0	< (less than)
1	0	1	<= (less than or equal)
1	1	0	!= (not equal value)
1	1	1	false (independent of the value)

Table 1: Comparison operation combinations

3.1. Changes to RFC5575

The behaviour of the bit-combinations is not explicitly defined. Especially the three combinations (0,0,0), (1,1,0), (1,1,1) are subject to misinterpretation.

4. Clarification of the Component Type Length

RFC 5575 section 4 under Type 3 defines the encoding of {operator, value} pairs. The operator component contains a 2 bit length field (bit 2 and 3) representing the length of the value field. This length field allows to encode a 1-4 byte long value. This {operator, value} pairs are used as a match criteria for the flow component type

3, 4, 5, 6, 7, 8, 10, 11. The allowed length for these components is as following:

Type	Length	Name
3	1-byte	IP protocol
4	1- or 2-byte	Port
5	1- or 2-byte	Destination port
6	1- or 2-byte	Source port
7	1-byte	ICMP type
8	1-byte	ICMP code
10	1- or 2-byte	Packet length
11	1-byte	DSCP

Table 2: Allowed value lengths

4.1. Changes to RFC5575

The allowed length of the value for a type 3 component (IP protocol) was not explicitly defined. This is inconsistent with all other types where the allowed length is explicitly specified.

5. (Re-)Validation of the Flow Specification NLRI

RFC 5575 section 6 defines a validation procedure for the flow specification NLRI. The outcome of the defined validation procedure is depending on the best-match unicast route for the destination prefix embedded in the flow specification.

Since the best-match unicast route may change over the time independently of the flow specification NLRI, revalidation of the flow specification NLRI MUST be performed whenever unicast routes change. Thus changes in the best-match unicast route can effect the validation-state of a flow specification NLRI.

5.1. Changes to RFC5575

Explicit definition of the requirement of revalidation. Revalidation of flow specification NLRI is not explicitly described in RFC5575, however it is compared with the "validation" of the reachability of the NEXT_HOP in the context of IP routing information. A unicast route becomes unfeasible if the NEXT_HOP for that particular route becomes unreachable.

6. Transitivity of Traffic Filtering Actions

RFC 5575 section 7 defines a minimum set of filtering actions. The predefined traffic filtering actions are standardized as BGP extended community values [RFC4360]. All predefined filtering action communities SHALL be treated as transitive BGP extended communities.

6.1. Changes to RFC5575

The transitivity of the traffic filtering action extended community was only defined for the "traffic-rate" action (defined as non-transitive). Since all filtering communities were assigned from an transitive pool by IANA, for consistency with RFC4360 section 2 this document explicitly redefines the "traffic-rate" action as transitive. It also defines the transitivity of all other traffic filtering actions as transitive (this definition is missing in RFC5575). This redefinition also reflects the behaviour of many of the current implementations.

7. Clarification of Flowspec NLRI Parsing and Validation

A flow specification NLRI is syntactically correct if it is encoded according to RFC 5575 section 4. The semantic of the NLRI is opaque to BGP. As a result of this statement the following behaviour is expected:

Flow specification NLRI propagation through the network is following the BGP propagation mechanisms independent of the semantic of the particular NLRI itself. The inability of the particular implementation to actually make use of a given flow specification SHOULD NOT affect the BGP NLRI propagation. Nor SHOULD a semantical incorrect NLRI affect the propagation of the NLRI (the NLRI, even if semantically incorrect should be propergated according to BGP propagation mechanisms).

Invalid NLRI semantics SHOULD NOT trigger BGP failures (ie BGP notifications).

An example of a syntactically correct, but semantically incorrect NLRI match criteria may be the following:

```
IP Protocol == 1, Port == 2
```

Since IP protocol 1 (ICMP) packets do not contain a port information the NLRI is incorrect from the semantical perspective and may not be applied to the forwarding plane in the network. However it is still syntactically correct and thus subject to BGP propagation.

7.1. Changes to RFC5575

None. See RFC5575 section 3 last 2 paragraphs.

8. Acknowledgements

The authors would like to thank Alexander Mayrhofer and Nicolas Fevrier for their comments and support.

9. IANA Considerations

This document has no IANA actions.

10. Security Considerations

The required filtering action for a specific NLRI may vary throughout the network. Extensive modification and filtering of filter actions is needed in an inter AS setting. Therefore implementations SHALL provide a policy framework to allow modification (add, modify, delete) of the filtering actions.

Especially in an inter-AS-setting unverified filtering actions like "redirect" (0x8008) or "traffic-marking" (0x8009) may potentially be harmful ("redirect" may allow any Flowspec-peer to redirect any traffic into arbitrary VPNs; "traffic-marking" allows any malicious Flowspec-peer to assign different forwarding classes to arbitrary traffic).

Inbound and outbound Flowspec route filters may also be necessary in order to match and filter specific filtering actions and flow component types from being accepted or sent by the local BGP daemon. The implementations SHALL therefore also provide a policy framework which provides the described functionality.

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <http://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, <http://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, <http://www.rfc-editor.org/info/rfc4760>.

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

11.2. Informative References

[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, <http://www.rfc-editor.org/info/rfc2474>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <http://www.rfc-editor.org/info/rfc4364>.

Authors' Addresses

Christoph Loibl
Next Layer Communications
Mariahilfer Guertel 37/7
Vienna 1150
AT

Phone: +43 664 1176414
Email: cl@tix.at

Martin Bacher
T-Mobile Austria
Rennweg 97-99
Vienna 1030
AT

Phone: +43 676 8200 5143
Email: martin.bacher@t-mobile.at