BGP Flow Spec for DDoS mitigation

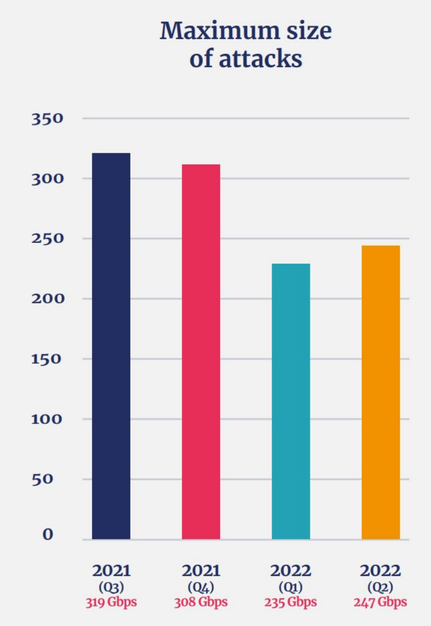
Hello

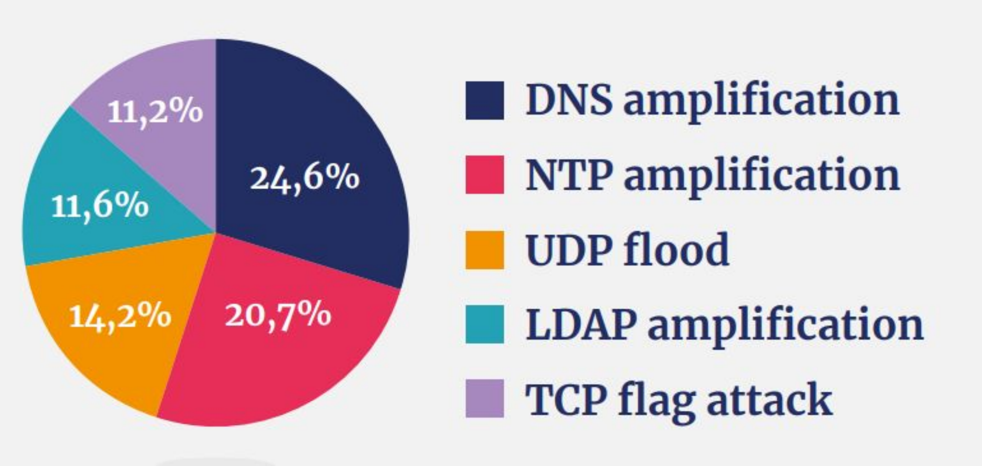
I’m Pavel Odintsov, DDoS mitigation enthusiast, the author of FastNetMon: https://fastnetmon.com and CTO of FastNetMon LTD.

Ways to contact me:

- linkedin.com/in/podintsov
- github.com/pavel-odintsov
- twitter.com/odintsov_pavel
- IRC, Libera Chat, pavel_odintsov
- pavel@fastnetmon.com

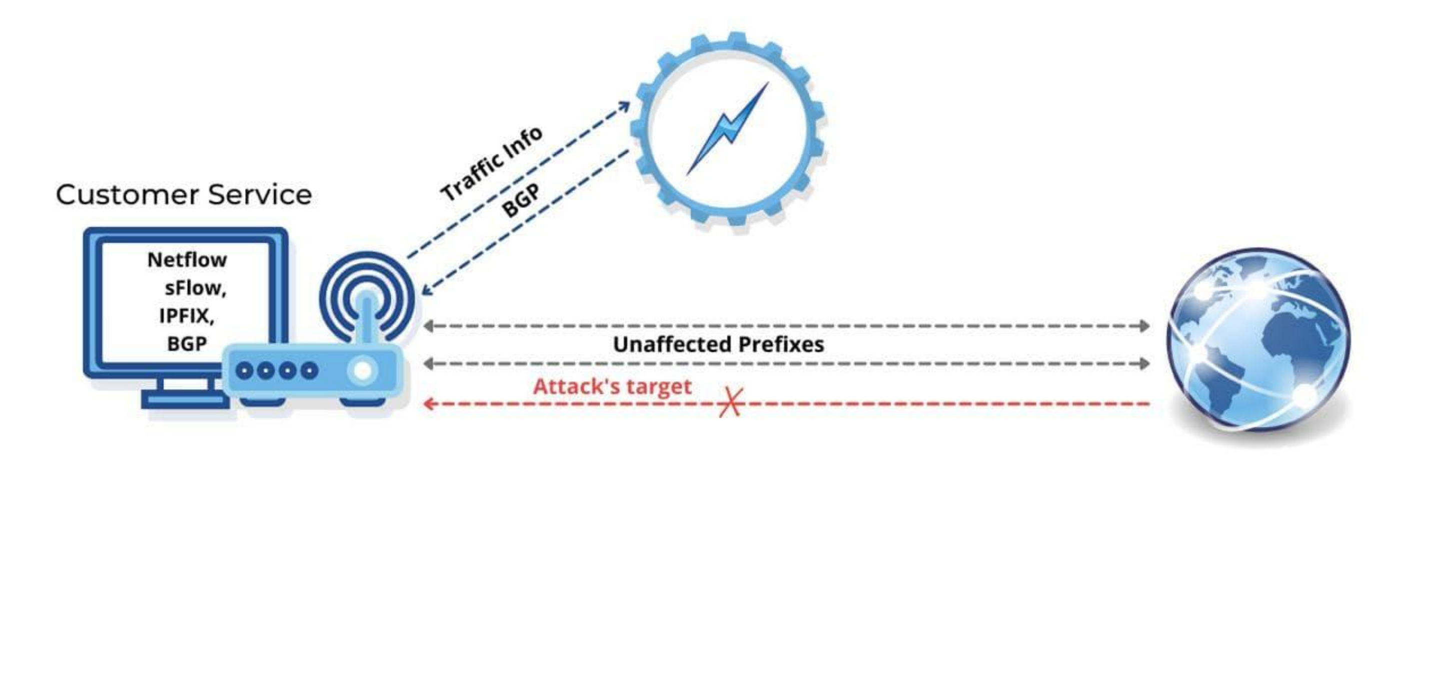
Current DDoS Weather





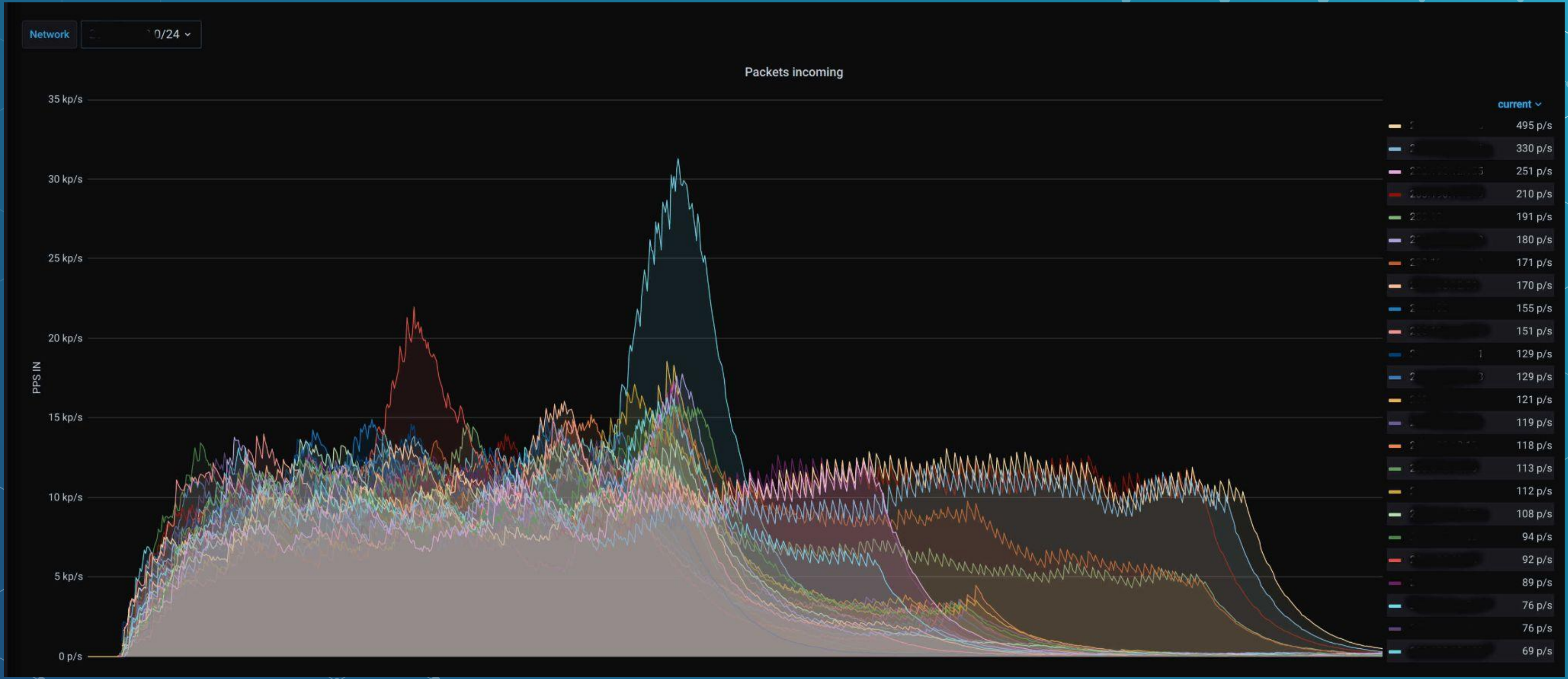
Data provided by The Dutch National Scrubbing Center (NaWas), Q2 2022

BGP Blackhole / RTBH



What is the problem?

Carpet Bombing Attack



What is BGP Flow Spec / RFC5575

- Protocol to configure distributed firewall
- BGP NLRI (Network Layer Reachability Information)
- RFC 5575 standard was published in 2009

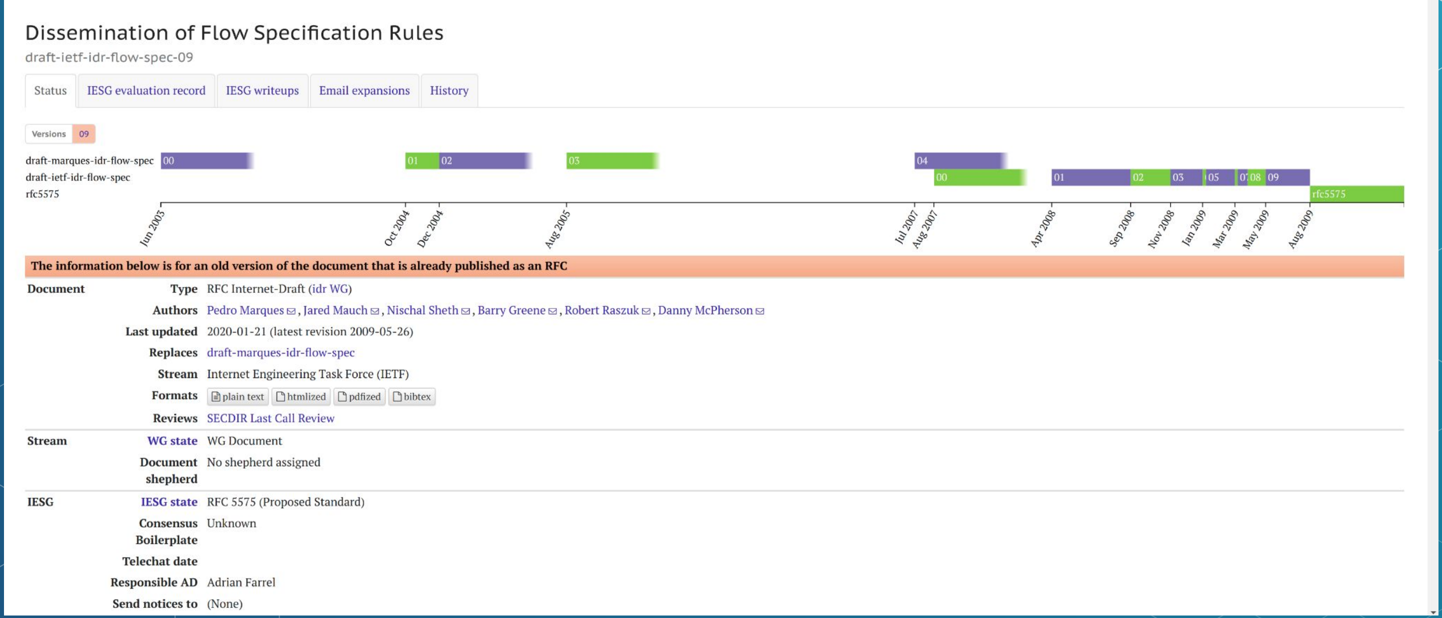
BGP Flow Spec filtering capabilities

- Source prefix (IPv4 or IPv6)
- Destination prefix (IPv4 or IPv6)
- IP Protocol number
- List or range of source ports for TCP and UDP
- List or range of destination ports for TCP and UDP
- ICMP code
- TCP flags
- Packet length
- Fragmentation flags (do not fragment, is fragment, first or last fragment)
- DSCP

BGP Flow Spec filtering actions

- Drop
- Rate limit
- Accept
- Mark (DSCP)
- Redirect to VRF
- Redirect to nexthop (draft)

Workgroup spent 6 years on RFC 5575



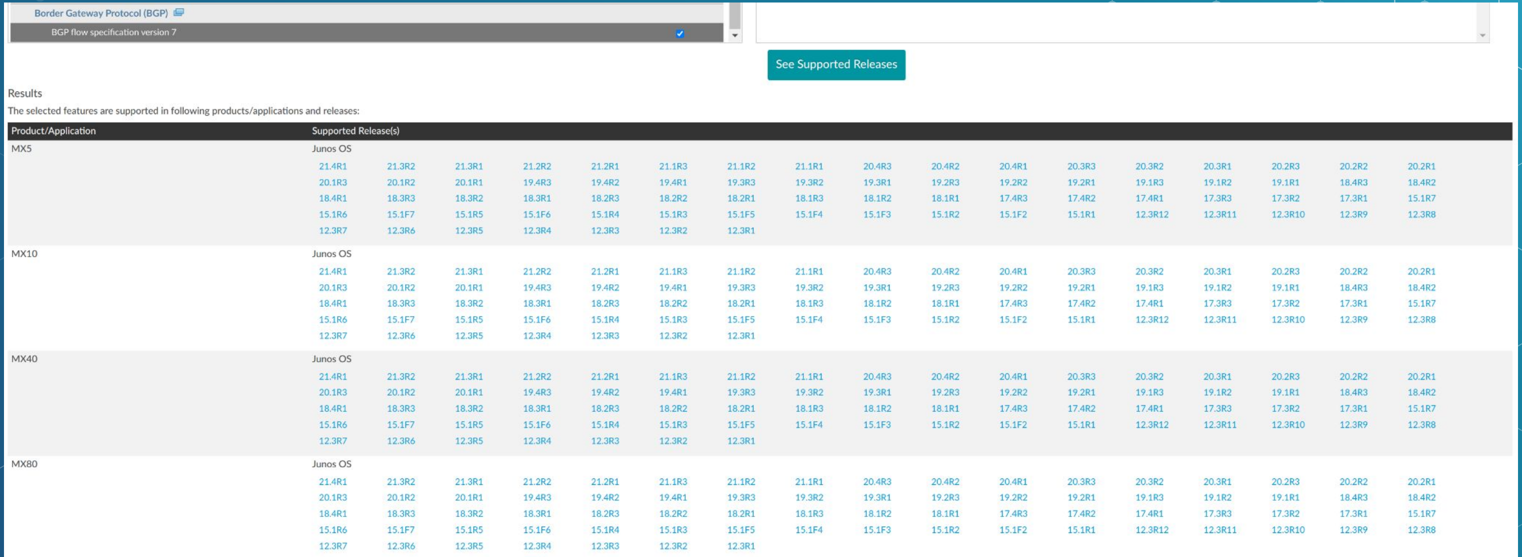
Dissemination of Flow Specification Rules

draft-ietf-idr-flow-spec-09

The information below is for an old version of the document that is already published as an RFC

Document	**Type**	RFC Internet-Draft (idr WG)
	Authors	Pedro Marques ✉, Jared Mauch ✉, Nischal Sheth ✉, Barry Greene ✉, Robert Raszuk ✉, Danny McPherson ✉
	Last updated	2020-01-21 (latest revision 2009-05-26)
	Replaces	draft-marques-idr-flow-spec
	Stream	Internet Engineering Task Force (IETF)
	Formats	plain text htmlized pdfized bibtex
	Reviews	SECDIR Last Call Review
Stream	**WG state**	WG Document
	Document shepherd	No shepherd assigned
IESG	**IESG state**	RFC 5575 (Proposed Standard)
	Consensus Boilerplate	Unknown
	Telechat date	
	Responsible AD	Adrian Farrel
	Send notices to	(None)

Support on Juniper, JunOS 12.3, March 2012?



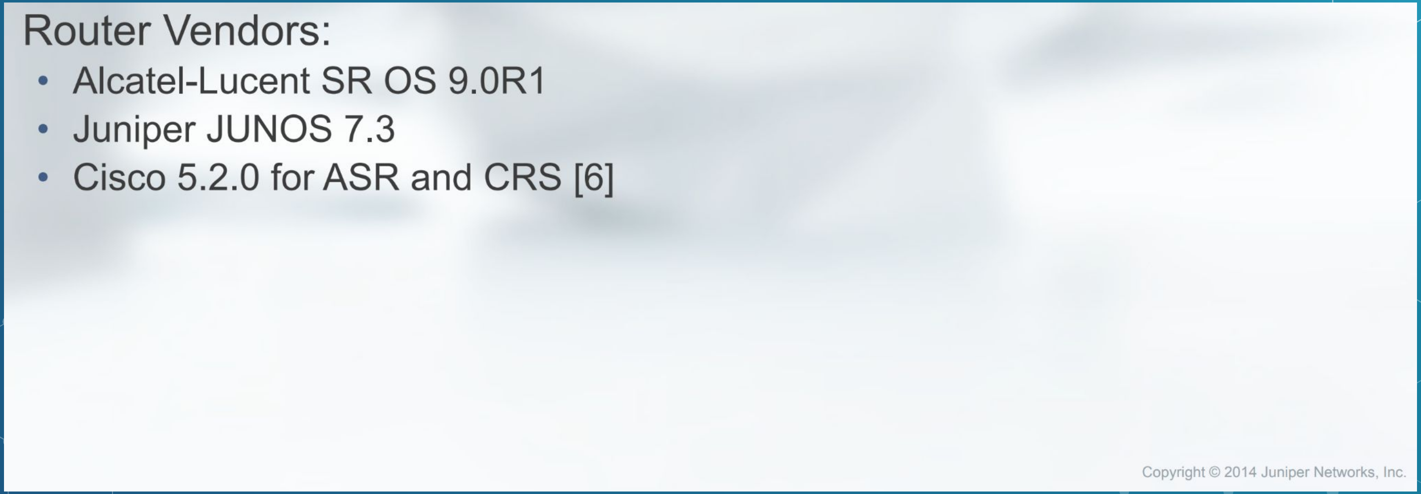
Border Gateway Protocol (BGP)
BGP flow specification version 7 ☑
See Supported Releases

Results

The selected features are supported in following products/applications and releases:

Product/Application	Supported Release(s)																
MX5	Junos OS																
	21.4R1	21.3R2	21.3R1	21.2R2	21.2R1	21.1R3	21.1R2	21.1R1	20.4R3	20.4R2	20.4R1	20.3R3	20.3R2	20.3R1	20.2R3	20.2R2	20.2R1
	20.1R3	20.1R2	20.1R1	19.4R3	19.4R2	19.4R1	19.3R3	19.3R2	19.3R1	19.2R3	19.2R2	19.2R1	19.1R3	19.1R2	19.1R1	18.4R3	18.4R2
	18.4R1	18.3R3	18.3R2	18.3R1	18.2R3	18.2R2	18.2R1	18.1R3	18.1R2	18.1R1	17.4R3	17.4R2	17.4R1	17.3R3	17.3R2	17.3R1	15.1R7
	15.1R6	15.1F7	15.1R5	15.1F6	15.1R4	15.1R3	15.1F5	15.1F4	15.1F3	15.1R2	15.1F2	15.1R1	12.3R12	12.3R11	12.3R10	12.3R9	12.3R8
	12.3R7	12.3R6	12.3R5	12.3R4	12.3R3	12.3R2	12.3R1										
MX10	Junos OS																
	21.4R1	21.3R2	21.3R1	21.2R2	21.2R1	21.1R3	21.1R2	21.1R1	20.4R3	20.4R2	20.4R1	20.3R3	20.3R2	20.3R1	20.2R3	20.2R2	20.2R1
	20.1R3	20.1R2	20.1R1	19.4R3	19.4R2	19.4R1	19.3R3	19.3R2	19.3R1	19.2R3	19.2R2	19.2R1	19.1R3	19.1R2	19.1R1	18.4R3	18.4R2
	18.4R1	18.3R3	18.3R2	18.3R1	18.2R3	18.2R2	18.2R1	18.1R3	18.1R2	18.1R1	17.4R3	17.4R2	17.4R1	17.3R3	17.3R2	17.3R1	15.1R7
	15.1R6	15.1F7	15.1R5	15.1F6	15.1R4	15.1R3	15.1F5	15.1F4	15.1F3	15.1R2	15.1F2	15.1R1	12.3R12	12.3R11	12.3R10	12.3R9	12.3R8
	12.3R7	12.3R6	12.3R5	12.3R4	12.3R3	12.3R2	12.3R1										
MX40	Junos OS																
	21.4R1	21.3R2	21.3R1	21.2R2	21.2R1	21.1R3	21.1R2	21.1R1	20.4R3	20.4R2	20.4R1	20.3R3	20.3R2	20.3R1	20.2R3	20.2R2	20.2R1
	20.1R3	20.1R2	20.1R1	19.4R3	19.4R2	19.4R1	19.3R3	19.3R2	19.3R1	19.2R3	19.2R2	19.2R1	19.1R3	19.1R2	19.1R1	18.4R3	18.4R2
	18.4R1	18.3R3	18.3R2	18.3R1	18.2R3	18.2R2	18.2R1	18.1R3	18.1R2	18.1R1	17.4R3	17.4R2	17.4R1	17.3R3	17.3R2	17.3R1	15.1R7
	15.1R6	15.1F7	15.1R5	15.1F6	15.1R4	15.1R3	15.1F5	15.1F4	15.1F3	15.1R2	15.1F2	15.1R1	12.3R12	12.3R11	12.3R10	12.3R9	12.3R8
	12.3R7	12.3R6	12.3R5	12.3R4	12.3R3	12.3R2	12.3R1										
MX80	Junos OS																
	21.4R1	21.3R2	21.3R1	21.2R2	21.2R1	21.1R3	21.1R2	21.1R1	20.4R3	20.4R2	20.4R1	20.3R3	20.3R2	20.3R1	20.2R3	20.2R2	20.2R1
	20.1R3	20.1R2	20.1R1	19.4R3	19.4R2	19.4R1	19.3R3	19.3R2	19.3R1	19.2R3	19.2R2	19.2R1	19.1R3	19.1R2	19.1R1	18.4R3	18.4R2
	18.4R1	18.3R3	18.3R2	18.3R1	18.2R3	18.2R2	18.2R1	18.1R3	18.1R2	18.1R1	17.4R3	17.4R2	17.4R1	17.3R3	17.3R2	17.3R1	15.1R7
	15.1R6	15.1F7	15.1R5	15.1F6	15.1R4	15.1R3	15.1F5	15.1F4	15.1F3	15.1R2	15.1F2	15.1R1	12.3R12	12.3R11	12.3R10	12.3R9	12.3R8
	12.3R7	12.3R6	12.3R5	12.3R4	12.3R3	12.3R2	12.3R1										

Support on Juniper, JunOS 7.3, August 2005?

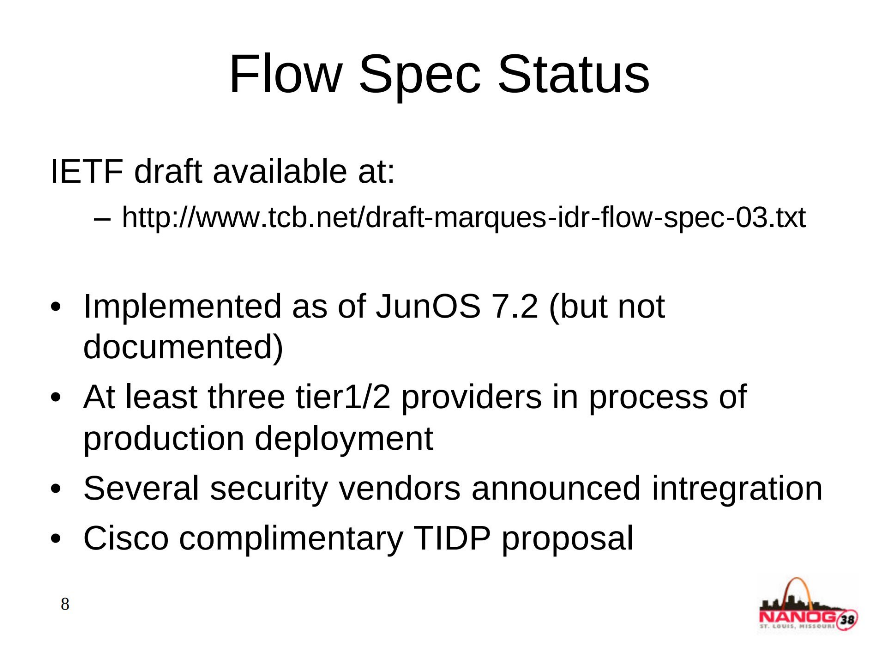


Router Vendors:

- Alcatel-Lucent SR OS 9.0R1
- Juniper JUNOS 7.3
- Cisco 5.2.0 for ASR and CRS [6]

https://archive.nanog.org/sites/default/files/tuesday_general_ddos_ryburn_63.16.pdf

Support on Juniper, JunOS 7.2, May 2005!



Flow Spec Status

IETF draft available at:

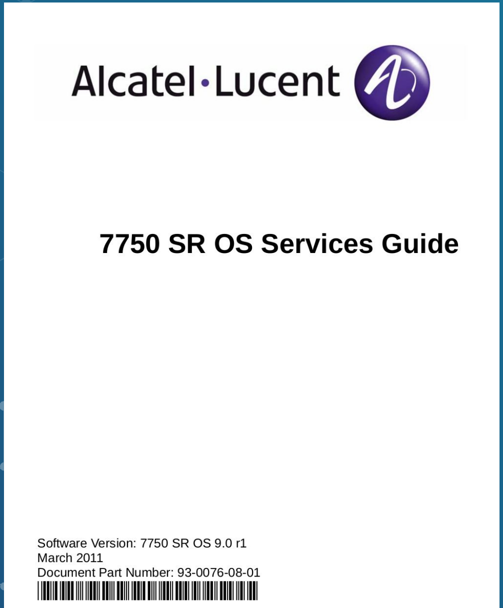
- http://www.tcb.net/draft-marques-idr-flow-spec-03.txt

- Implemented as of JunOS 7.2 (but not documented)
- At least three tier1/2 providers in process of production deployment
- Several security vendors announced intregration
- Cisco complimentary TIDP proposal

8

https://archive.nanog.org/meetings/nanog38/presentations/labovitz-bgp-flowspec.pdf

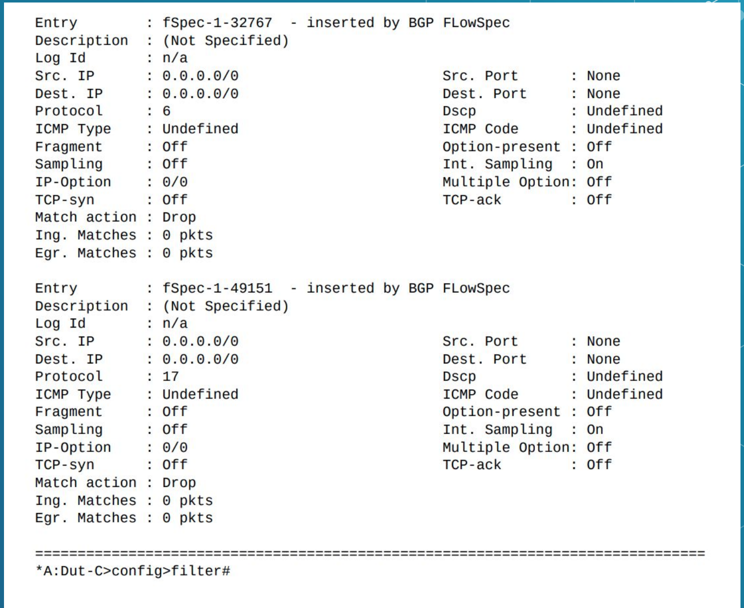
Support on Nokia, March 2011



Alcatel·Lucent

7750 SR OS Services Guide

Software Version: 7750 SR OS 9.0 r1
March 2011
Document Part Number: 93-0076-08-01

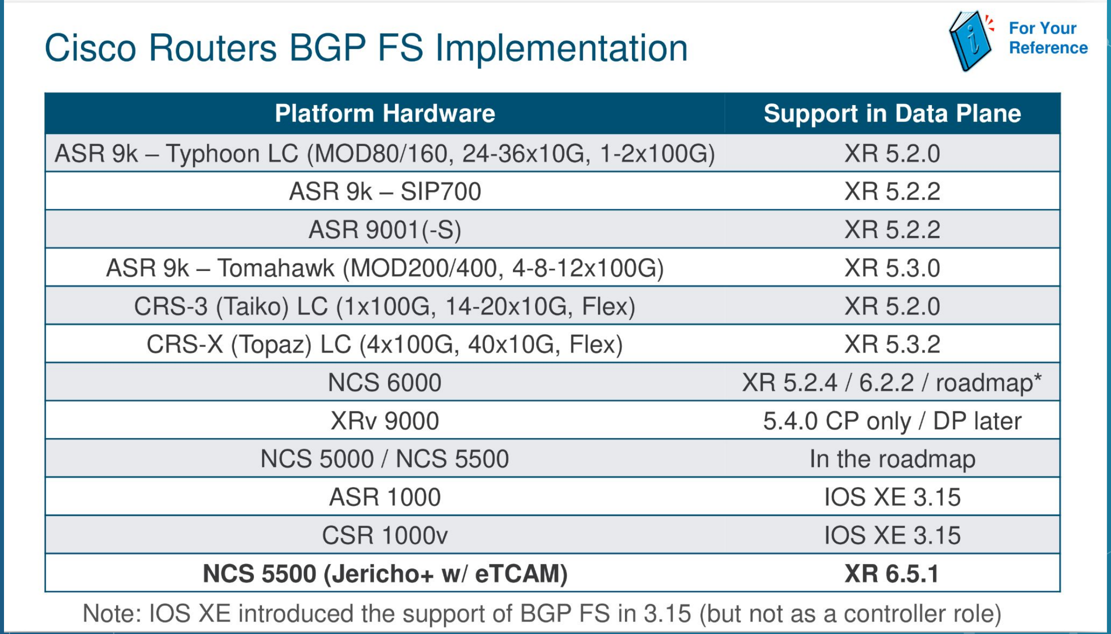


```
Entry        : fSpec-1-32767  - inserted by BGP FLowSpec
Description  : (Not Specified)
Log Id       : n/a
Src. IP      : 0.0.0.0/0                        Src. Port      : None
Dest. IP     : 0.0.0.0/0                        Dest. Port     : None
Protocol     : 6                                Dscp           : Undefined
ICMP Type    : Undefined                        ICMP Code      : Undefined
Fragment     : Off                              Option-present : Off
Sampling     : Off                              Int. Sampling  : On
IP-Option    : 0/0                              Multiple Option: Off
TCP-syn      : Off                              TCP-ack        : Off
Match action : Drop
Ing. Matches : 0 pkts
Egr. Matches : 0 pkts

Entry        : fSpec-1-49151  - inserted by BGP FLowSpec
Description  : (Not Specified)
Log Id       : n/a
Src. IP      : 0.0.0.0/0                        Src. Port      : None
Dest. IP     : 0.0.0.0/0                        Dest. Port     : None
Protocol     : 17                               Dscp           : Undefined
ICMP Type    : Undefined                        ICMP Code      : Undefined
Fragment     : Off                              Option-present : Off
Sampling     : Off                              Int. Sampling  : On
IP-Option    : 0/0                              Multiple Option: Off
TCP-syn      : Off                              TCP-ack        : Off
Match action : Drop
Ing. Matches : 0 pkts
Egr. Matches : 0 pkts

===============================================================================
*A:Dut-C>config>filter#
```

Support on Cisco, 2014



Cisco Routers BGP FS Implementation

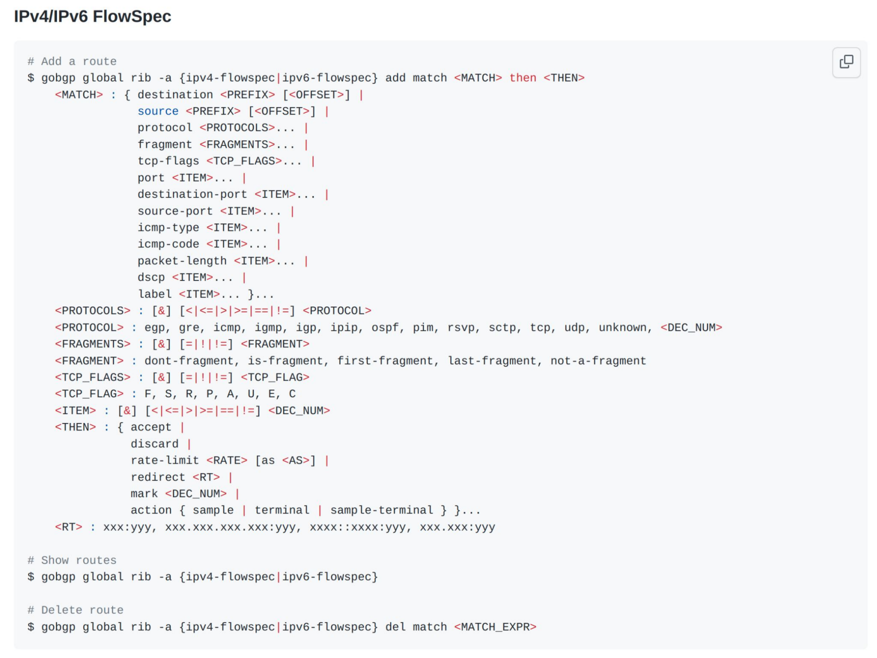
For Your Reference

Platform Hardware	Support in Data Plane
ASR 9k – Typhoon LC (MOD80/160, 24-36x10G, 1-2x100G)	XR 5.2.0
ASR 9k – SIP700	XR 5.2.2
ASR 9001(-S)	XR 5.2.2
ASR 9k – Tomahawk (MOD200/400, 4-8-12x100G)	XR 5.3.0
CRS-3 (Taiko) LC (1x100G, 14-20x10G, Flex)	XR 5.2.0
CRS-X (Topaz) LC (4x100G, 40x10G, Flex)	XR 5.3.2
NCS 6000	XR 5.2.4 / 6.2.2 / roadmap*
XRv 9000	5.4.0 CP only / DP later
NCS 5000 / NCS 5500	In the roadmap
ASR 1000	IOS XE 3.15
CSR 1000v	IOS XE 3.15
NCS 5500 (Jericho+ w/ eTCAM)	**XR 6.5.1**

Note: IOS XE introduced the support of BGP FS in 3.15 (but not as a controller role)

https://www.ciscolive.com/c/dam/r/ciscolive/emea/docs/2018/pdf/BRKSPG-3012.pdf

Support on GoBGP, 2015



IPv4/IPv6 FlowSpec

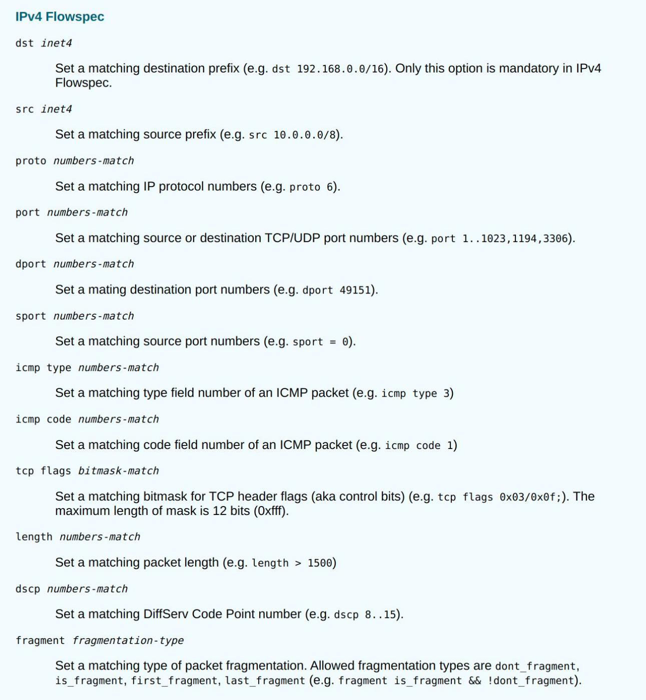
```
# Add a route
$ gobgp global rib -a {ipv4-flowspec|ipv6-flowspec} add match <MATCH> then <THEN>
    <MATCH> : { destination <PREFIX> [<OFFSET>] |
                source <PREFIX> [<OFFSET>] |
                protocol <PROTOCOLS>... |
                fragment <FRAGMENTS>... |
                tcp-flags <TCP_FLAGS>... |
                port <ITEM>... |
                destination-port <ITEM>... |
                source-port <ITEM>... |
                icmp-type <ITEM>... |
                icmp-code <ITEM>... |
                packet-length <ITEM>... |
                dscp <ITEM>... |
                label <ITEM>... }...
    <PROTOCOLS> : [&] [<|<=|>|>=|==|!=] <PROTOCOL>
    <PROTOCOL> : egp, gre, icmp, igmp, igp, ipip, ospf, pim, rsvp, sctp, tcp, udp, unknown, <DEC_NUM>
    <FRAGMENTS> : [&] [=|!|!=] <FRAGMENT>
    <FRAGMENT> : dont-fragment, is-fragment, first-fragment, last-fragment, not-a-fragment
    <TCP_FLAGS> : [&] [=|!|!=] <TCP_FLAG>
    <TCP_FLAG> : F, S, R, P, A, U, E, C
    <ITEM> : [&] [<|<=|>|>=|==|!=] <DEC_NUM>
    <THEN> : { accept |
               discard |
               rate-limit <RATE> [as <AS>] |
               redirect <RT> |
               mark <DEC_NUM> |
               action { sample | terminal | sample-terminal } }...
    <RT> : xxx:yyy, xxx.xxx.xxx.xxx:yyy, xxxx::xxxx:yyy, xxx.xxx:yyy

# Show routes
$ gobgp global rib -a {ipv4-flowspec|ipv6-flowspec}

# Delete route
$ gobgp global rib -a {ipv4-flowspec|ipv6-flowspec} del match <MATCH_EXPR>
```

https://ripe71.ripe.net/presentations/135-RIPE71_GoBGP.pdf

Support on Bird 2, 2017



IPv4 Flowspec

`dst` *`inet4`*

Set a matching destination prefix (e.g. `dst 192.168.0.0/16`). Only this option is mandatory in IPv4 Flowspec.

`src` *`inet4`*

Set a matching source prefix (e.g. `src 10.0.0.0/8`).

`proto` *`numbers-match`*

Set a matching IP protocol numbers (e.g. `proto 6`).

`port` *`numbers-match`*

Set a matching source or destination TCP/UDP port numbers (e.g. `port 1..1023,1194,3306`).

`dport` *`numbers-match`*

Set a mating destination port numbers (e.g. `dport 49151`).

`sport` *`numbers-match`*

Set a matching source port numbers (e.g. `sport = 0`).

`icmp type` *`numbers-match`*

Set a matching type field number of an ICMP packet (e.g. `icmp type 3`)

`icmp code` *`numbers-match`*

Set a matching code field number of an ICMP packet (e.g. `icmp code 1`)

`tcp flags` *`bitmask-match`*

Set a matching bitmask for TCP header flags (aka control bits) (e.g. `tcp flags 0x03/0x0f;`). The maximum length of mask is 12 bits (0xfff).

`length` *`numbers-match`*

Set a matching packet length (e.g. `length > 1500`)

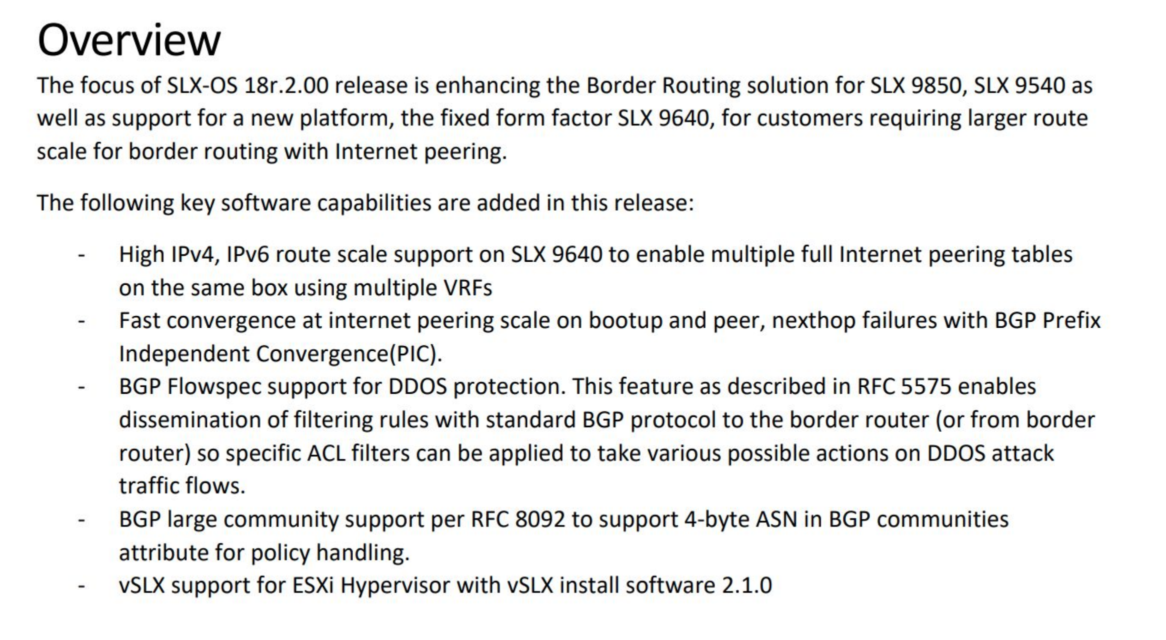
`dscp` *`numbers-match`*

Set a matching DiffServ Code Point number (e.g. `dscp 8..15`).

`fragment` *`fragmentation-type`*

Set a matching type of packet fragmentation. Allowed fragmentation types are `dont_fragment`, `is_fragment`, `first_fragment`, `last_fragment` (e.g. `fragment is_fragment && !dont_fragment`).

Support on Extreme, December 2018



Overview

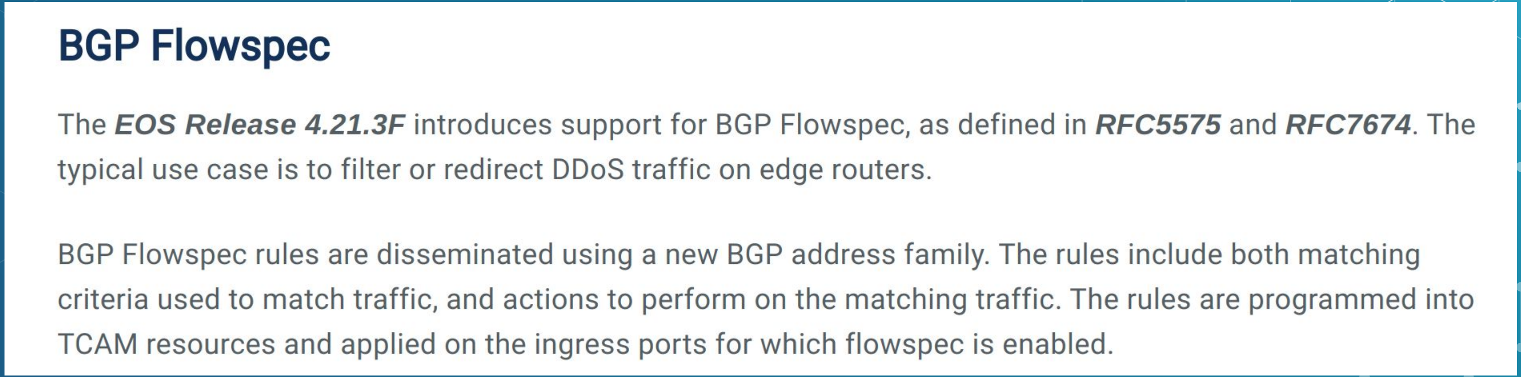
The focus of SLX-OS 18r.2.00 release is enhancing the Border Routing solution for SLX 9850, SLX 9540 as well as support for a new platform, the fixed form factor SLX 9640, for customers requiring larger route scale for border routing with Internet peering.

The following key software capabilities are added in this release:

- High IPv4, IPv6 route scale support on SLX 9640 to enable multiple full Internet peering tables on the same box using multiple VRFs
- Fast convergence at internet peering scale on bootup and peer, nexthop failures with BGP Prefix Independent Convergence(PIC).
- BGP Flowspec support for DDOS protection. This feature as described in RFC 5575 enables dissemination of filtering rules with standard BGP protocol to the border router (or from border router) so specific ACL filters can be applied to take various possible actions on DDOS attack traffic flows.
- BGP large community support per RFC 8092 to support 4-byte ASN in BGP communities attribute for policy handling.
- vSLX support for ESXi Hypervisor with vSLX install software 2.1.0

https://documentation.extremenetworks.com/release_notes/slxos/18r.2.00/SLX-OS_18r.2.00_v3_ReleaseNotes.pdf

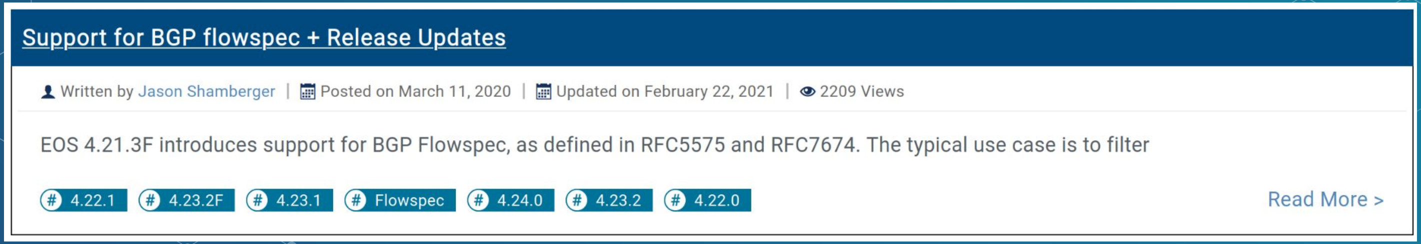
Support on Arista, March 2020



BGP Flowspec

The ***EOS Release 4.21.3F*** introduces support for BGP Flowspec, as defined in ***RFC5575*** and ***RFC7674***. The typical use case is to filter or redirect DDoS traffic on edge routers.

BGP Flowspec rules are disseminated using a new BGP address family. The rules include both matching criteria used to match traffic, and actions to perform on the matching traffic. The rules are programmed into TCAM resources and applied on the ingress ports for which flowspec is enabled.



Support for BGP flowspec + Release Updates

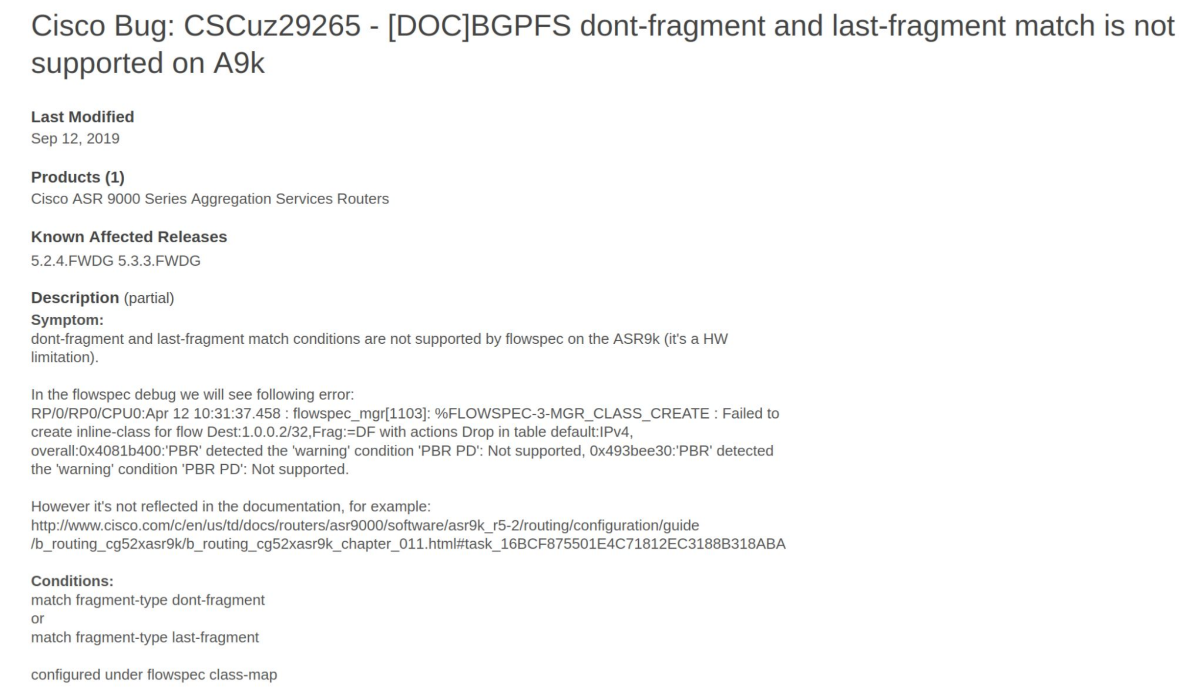
Written by Jason Shamberger | Posted on March 11, 2020 | Updated on February 22, 2021 | 2209 Views

EOS 4.21.3F introduces support for BGP Flowspec, as defined in RFC5575 and RFC7674. The typical use case is to filter

BGP Flow Spec challenges

- Limited number of BGP Flow Spec rules
- Lack of standard approach to retrieve packet and byte counters per rule
- Lack of proper rule validation
- Different hardware limitations
- Lack of interface to manage rules efficiently
- Weak integration with Netflow and IPFIX
- Lack of solid support for draft-ietf-idr-flowspec-redirect-ip-00

BGP Flow Spec hardware limits: ASR 9000



Cisco Bug: CSCuz29265 - [DOC]BGPFS dont-fragment and last-fragment match is not supported on A9k

Last Modified
Sep 12, 2019

Products (1)
Cisco ASR 9000 Series Aggregation Services Routers

Known Affected Releases
5.2.4.FWDG 5.3.3.FWDG

Description (partial)
Symptom:
dont-fragment and last-fragment match conditions are not supported by flowspec on the ASR9k (it's a HW limitation).

In the flowspec debug we will see following error:

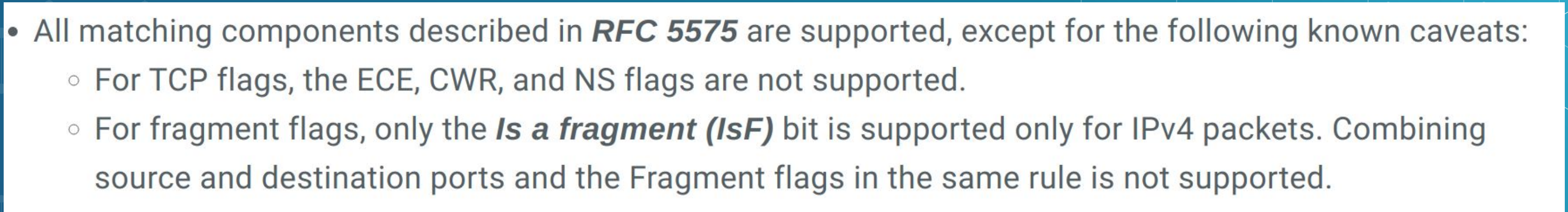
```
RP/0/RP0/CPU0:Apr 12 10:31:37.458 : flowspec_mgr[1103]: %FLOWSPEC-3-MGR_CLASS_CREATE : Failed to
create inline-class for flow Dest:1.0.0.2/32,Frag:=DF with actions Drop in table default:IPv4,
overall:0x4081b400:'PBR' detected the 'warning' condition 'PBR PD': Not supported, 0x493bee30:'PBR' detected
the 'warning' condition 'PBR PD': Not supported.
```

However it's not reflected in the documentation, for example:
http://www.cisco.com/c/en/us/td/docs/routers/asr9000/software/asr9k_r5-2/routing/configuration/guide/b_routing_cg52xasr9k/b_routing_cg52xasr9k_chapter_011.html#task_16BCF875501E4C71812EC3188B318ABA

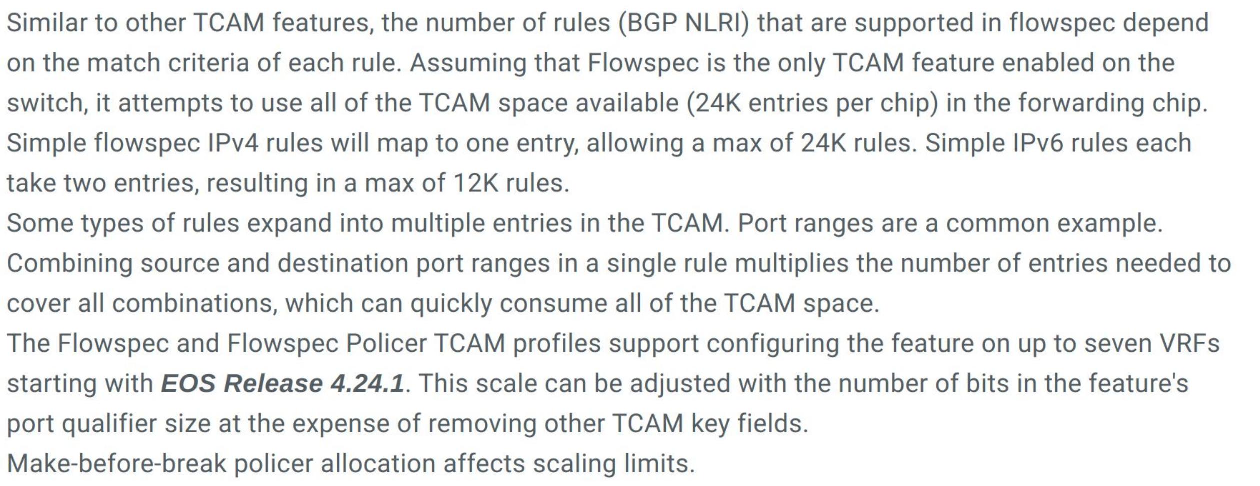
Conditions:
match fragment-type dont-fragment
or
match fragment-type last-fragment

configured under flowspec class-map

BGP Flow Spec hardware limits: Arista



- All matching components described in ***RFC 5575*** are supported, except for the following known caveats:
  - For TCP flags, the ECE, CWR, and NS flags are not supported.
  - For fragment flags, only the ***Is a fragment (IsF)*** bit is supported only for IPv4 packets. Combining source and destination ports and the Fragment flags in the same rule is not supported.



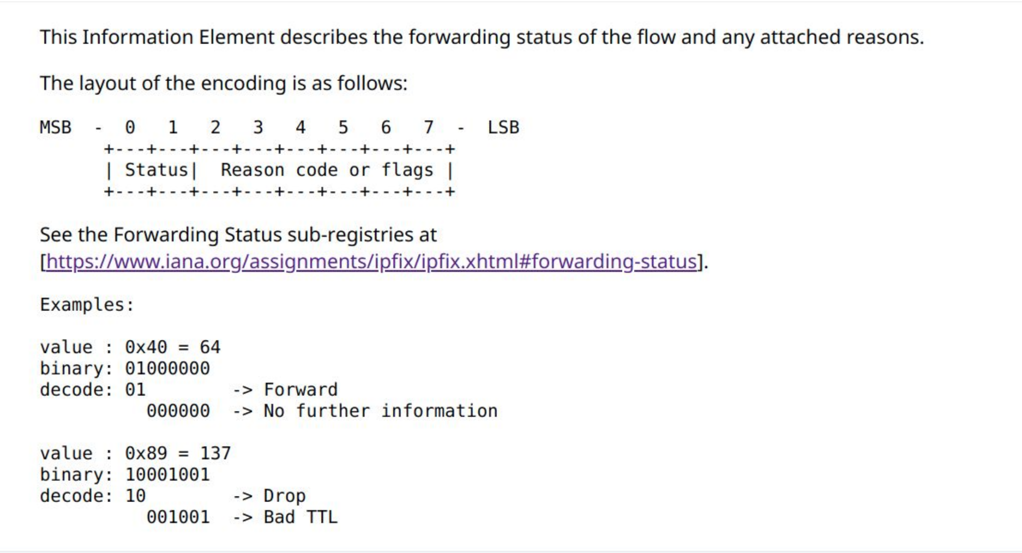
Similar to other TCAM features, the number of rules (BGP NLRI) that are supported in flowspec depend on the match criteria of each rule. Assuming that Flowspec is the only TCAM feature enabled on the switch, it attempts to use all of the TCAM space available (24K entries per chip) in the forwarding chip. Simple flowspec IPv4 rules will map to one entry, allowing a max of 24K rules. Simple IPv6 rules each take two entries, resulting in a max of 12K rules.

Some types of rules expand into multiple entries in the TCAM. Port ranges are a common example. Combining source and destination port ranges in a single rule multiplies the number of entries needed to cover all combinations, which can quickly consume all of the TCAM space.

The Flowspec and Flowspec Policer TCAM profiles support configuring the feature on up to seven VRFs starting with ***EOS Release 4.24.1***. This scale can be adjusted with the number of bits in the feature's port qualifier size at the expense of removing other TCAM key fields.

Make-before-break policer allocation affects scaling limits.

BGP Flow Spec and IPFIX, Netflow on Cisco



This Information Element describes the forwarding status of the flow and any attached reasons.

The layout of the encoding is as follows:

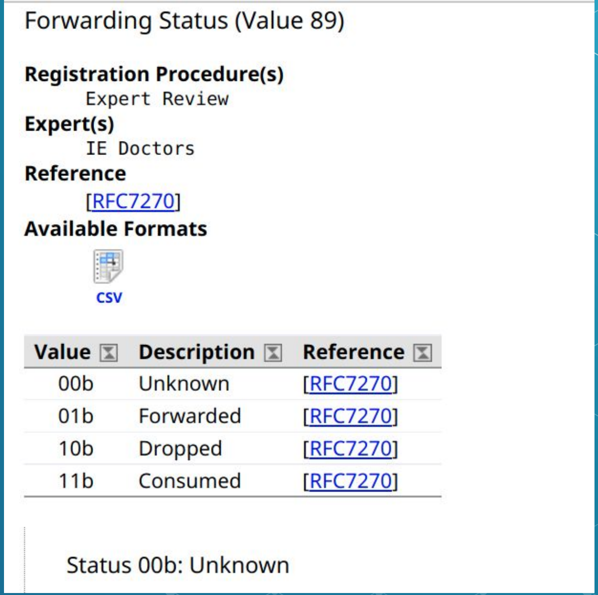
```
MSB  -  0   1   2   3   4   5   6   7  -  LSB
      +---+---+---+---+---+---+---+---+
      | Status|  Reason code or flags |
      +---+---+---+---+---+---+---+---+
```

See the Forwarding Status sub-registries at [https://www.iana.org/assignments/ipfix/ipfix.xhtml#forwarding-status].

```
Examples:

value : 0x40 = 64
binary: 01000000
decode: 01        -> Forward
          000000  -> No further information

value : 0x89 = 137
binary: 10001001
decode: 10        -> Drop
          001001  -> Bad TTL
```



Forwarding Status (Value 89)

Registration Procedure(s)

Expert Review

Expert(s)

IE Doctors

Reference

[RFC7270]

Available Formats

CSV

Value	Description	Reference
00b	Unknown	[RFC7270]
01b	Forwarded	[RFC7270]
10b	Dropped	[RFC7270]
11b	Consumed	[RFC7270]

Status 00b: Unknown

FastNetMon: our community

- Site: https://fastnetmon.com
- GitHub: https://github.com/pavel-odintsov/fastnetmon
- Slack: https://slack.fastnetmon.com/
- Telegram: https://t.me/fastnetmon
- IRC: #fastnetmon at Libra Chat
- Discord: https://discord.fastnetmon.com/
- LinkedIN: https://www.linkedin.com/company/fastnetmon/
- Facebook: https://www.facebook.com/fastnetmon/
- Twitter: https://twitter.com/fastnetmon

THANKS!

ANY QUESTIONS?

You can find me at:

- @odintsov_pavel
- pavel@fastnetmon.com
- linkedin.com/in/podintsov