Network Working Group                                          J. Uttaro
Internet-Draft                                                      AT&T
Updates: 8955 (if approved)                                   J. Alcaide
Intended status: Standards Track                             C. Filsfils
Expires: November 14, 2021                                     D. Smith
                                                                  Cisco
                                                           P. Mohapatra
                                                       Sproute Networks
                                                           May 13, 2021

                Revised Validation Procedure for BGP Flow Specifications
                        draft-ietf-idr-bgp-flowspec-oid-14

Abstract

   This document describes a modification to the validation procedure
   defined for the dissemination of BGP Flow Specifications.  The
   dissemination of BGP Flow Specifications requires that the originator
   of the Flow Specification matches the originator of the best-match
   unicast route for the destination prefix embedded in the Flow
   Specification.  For an iBGP received route, the originator is
   typically a border router within the same autonomous system.  The
   objective is to allow only BGP speakers within the data forwarding
   path to originate BGP Flow Specifications.  Sometimes it is desirable
   to originate the BGP Flow Specification from any place within the
   autonomous system itself, for example, from a centralized BGP route
   controller.  However, the validation procedure will fail in this
   scenario.  The modification proposed herein relaxes the validation
   rule to enable Flow Specifications to be originated within the same
   autonomous system as the BGP speaker performing the validation.
   Additionally, this document revises the AS_PATH validation rules so
   Flow Specifications received from an eBGP peer can be validated when
   such peer is a BGP route server.

   This document updates the validation procedure in [RFC8955].

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 November 14, 2021.

Copyright Notice

Table of Contents

## 1.  Requirements Language

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.

## 2.  Introduction

[RFC8955] defines a BGP NLRI [RFC4271] that can be used to distribute traffic Flow Specifications amongst BGP speakers in support of traffic filtering.  The primary intention of [RFC8955] is to enable downstream autonomous systems to signal traffic filtering policies to upstream autonomous systems.  In this way, traffic is filtered closer to the source and the upstream autonomous system(s) avoid carrying the traffic to the downstream autonomous system only to be discarded.  [RFC8955] also enables more granular traffic filtering based upon upper layer protocol information (e.g., protocol port numbers) as opposed to coarse IP destination prefix-based filtering.  Flow specification NLRIs received from a BGP peer are subject to validity checks before being considered feasible and subsequently installed within the respective Adj-RIB-In.

The validation procedure defined within [RFC8955] requires that the originator of the Flow Specification NLRI matches the originator of the best-match unicast route for the destination prefix embedded in the Flow Specification.  The aim is making sure that only speakers on the forwarding path can originate the Flow Specification.  Let's consider the particular case where the Flow Specification is originated in any location within the same autonomous system than the speaker performing the validation (for example by a centralized BGP route controller), and the best-match unicast route is originated in another autonomous system.  In order for the validation to succeed for a Flow Specification received from an iBGP peer, it could be possible to disseminate such Flow Specification NLRIs directly from the specific border router (within the local autonomous system) that is advertising the corresponding best-match unicast route to the local autonomous system.  Those border routers would be acting as de facto router controllers.  This approach would be, however, operationally cumbersome in an autonomous system with a large number of border routers having complex BGP policies.

Figure 1 illustrates this principle.  R1 (the upstream router) and RR need to validate the Flow Specification whose embedded destination prefix has a best-match unicast route (dest-route) originated by ASBR2.  ASBR2 could originate the Flow Specification, and it would be validated when received by RR and R1.  Sometimes the Flow Specification needs to be originated inside AS1.  ASBR1 could originate it, and the Flow Specification would still be validated.  In both cases, the Flow Specification is originated by a router in the same forwarding path as the dest-route.  For the case where AS1 has thousands of ASBRs, it becomes impractical to originate different Flow Specification rules on each ASBR in AS1 based on which ASBR each dest-route is learned from.  The objective is to advertise all the Flow Specifications from the same route-controller.
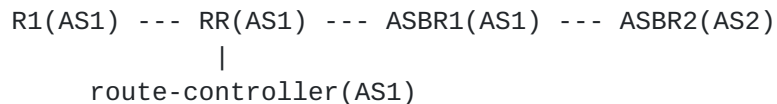
```
        R1(AS1) --- RR(AS1) --- ASBR1(AS1) --- ASBR2(AS2)
                        |
              route-controller(AS1)
```

                             Figure 1

   This document describes a modification to the [RFC8955] validation
   procedure, allowing Flow Specification NLRIs to be originated from a
   centralized BGP route controller located within the local autonomous
   system and not necessarily in the data forwarding path.  While the
   proposed modification cannot be used for inter-domain coordination of
   traffic filtering, it greatly simplifies distribution of intra-domain
   traffic filtering policies within an autonomous system which has a
   large number of border routers having complex BGP policies.  By
   relaxing the validation procedure for iBGP, the proposed modification
   allows Flow Specifications to be distributed in a standard and
   scalable manner throughout an autonomous system.

   Throughout this document, some references are made to
   AS_CONFED_SEQUENCE segments; see Sections 4.1 and 5.  If
   AS_CONFED_SET segments are also present in the AS_PATH, the same
   considerations apply to them.  Note, however, that the use of
   AS_CONFED_SET segments is not recommended [RFC6472].  Refer to
   [I-D.ietf-idr-deprecate-as-set-confed-set] as well.

## 3.  Motivation

   Step (b) of the validation procedure in [RFC8955], section 6 is
   defined with the underlying assumption that the Flow Specification
   NLRI traverses the same path, in the inter-domain and intra-domain
   route distribution graph, as that of the longest-match unicast route
   for the destination prefix embedded in the Flow Specification.

   In the case of inter-domain traffic filtering, the Flow Specification
   originator at the egress border routers of an AS (e.g.  RTR-D and
   RTR-E of AS1 in Figure 2) matches the eBGP neighbor that advertised
   the longest match destination prefix (see RTR-F and RTR-G
   respectively in Figure 2).

   Similarly, at the upstream routers of an AS (see RTR-A and RTR-B of
   AS1 in Figure 2), the Flow Specification originator matches the
   egress iBGP border routers that had advertised the unicast route for
   the best-match destination prefix (see RTR-D and RTR-E respectively
   in Figure 2).  This is true even when upstream routers select paths
   from different egress border routers as best route based upon IGP
   distance.  For example, in Figure 2:

      RTR-A chooses RTR-D as the best route

RTR-B chooses RTR-E as the best route

```
          / - - - - - - - - - - - - - -
          |              AS1                |
           +-------+          +-------+
          | |       |        |        |   | |
          | | RTR-A |        | RTR-B |
          | |       |        |        |   | |
           +-------+          +-------+
          |       \          /            |
            iBGP    \        / iBGP
          |          \      /             |
                   +-------+
          |        |       |              |
                   | RTR-C |
          |        |  RC   |              |
                   +-------+
          |           /   \              |
                      /     \
          |   iBGP   /       \ iBGP       |
            +-------+          +-------+
          | | RTR-D |        | RTR-E |  |
          |  |       |        |       |
          | |       |        |       |   | |
            +-------+          +-------+
          |    |                    |    |
           - - -|- - - - - - - - - -|- - -/
               | eBGP    eBGP |
           - - -|- - - - - - - - - -|- - -/
          |    |                    |    |
            +-------+          +-------+
          | |       |        |       |   | |
            | RTR-F |        | RTR-G |
          | |       |        |       |   | |
            +-------+          +-------+
          |            AS2                 |
          / - - - - - - - - - - - - - -
```
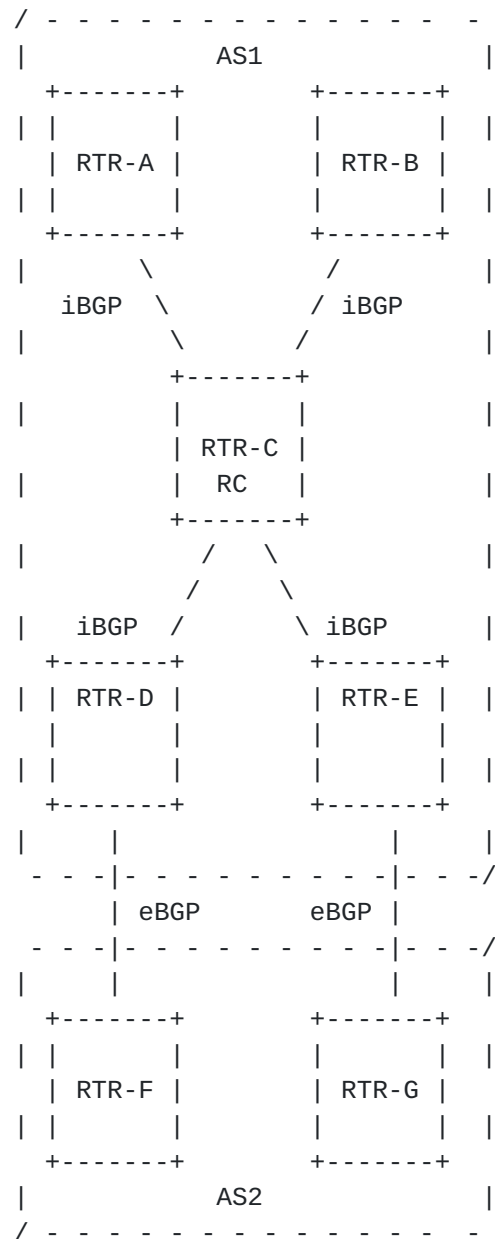
                         Figure 2

   It is highly desirable that mechanisms exist to protect each AS
   independently from network security attacks using the BGP Flow
   Specification NLRI for intra-AS purposes only.  Network operators
   often deploy a dedicated Security Operations Center (SOC) within
   their AS to monitor and detect such security attacks.  To mitigate
   attacks within an AS, operators require the ability to originate
   intra-AS Flow Specification NLRIs from a central BGP route controller
   that is not within the data forwarding plane.  In this way, operators

can direct border routers within their AS with specific attack
mitigation actions (drop the traffic, forward to a clean-pipe center,
etc.).

In addition, an operator may extend the requirements above for a
group of ASes via policy.  This is described in section (b.2.3) of
the validation procedure.

A central BGP route controller that originates a Flow Specification
NLRI should be able to avoid the complexity of having to determine
the egress border router whose path was chosen as the best for each
of its neighbors.  When a central BGP route controller originates a
Flow Specification NLRI, the rest of the speakers within the AS will
see the BGP controller as the originator of the Flow Specification in
terms of the validation procedure rules.  Thus, it is necessary to
modify step (b) of the [RFC8955] validation procedure such that an
iBGP peer that is not within the data forwarding plane may originate
Flow Specification NLRIs.

## 4.  Revised Validation Procedure

### 4.1.  Revision of Route Feasibility

Step (b) of the validation procedure specified in [RFC8955], section
6 is redefined as follows:

b) One of the following conditions MUST hold true:

   1.  The originator of the Flow Specification matches the
       originator of the best-match 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].

       1.  This condition SHOULD be enabled by default (it may be
           disabled by explicit configuration as described on the
           next list item (b.2.2)).

       2.  This condition MAY be disabled by explicit configuration
           on a BGP speaker.

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

Explanation:

   In this context, a local domain includes the local AS or the local
   confederation [RFC5065].  Receiving either an empty AS_PATH or one
   with only an AS_CONFED_SEQUENCE segment indicates that the Flow
   Specification was originated inside the local domain.

   With the above modification to the [RFC8955] validation procedure,
   a BGP peer within the local domain that is not within the data
   forwarding path can originate a Flow Specification.

   Disabling the new condition above (b.2.2) could be a good practice
   if the operator knew with certainty that a Flow Specification
   would not be originated inside the local domain.  An additional
   case would be if it was known for a fact that only the right
   egress border routers (i.e. those that were also egress border
   routers for the best routes) were originating a Flow Specification
   NLRI.

   Also, policy may be useful to validate a specific set of non-empty
   AS_PATHs (b.2.3).  For example, it could validate a Flow
   Specification whose AS_PATH contained only an AS_SEQUENCE segment
   with ASes that were all known to belong to the same administrative
   domain.

## 4.2.  Revision of AS_PATH Validation

[RFC8955] states:

   BGP implementations MUST also enforce that the AS_PATH attribute
   of a route received via the External Border Gateway Protocol
   (eBGP) contains the neighboring AS in the left-most position of
   the AS_PATH attribute.  While this rule is optional in the BGP
   specification, it becomes necessary to enforce it here for
   security reasons.

This rule prevents the exchange of BGP Flow Specification NLRIs at
Internet exchanges with BGP route servers [RFC7947].  Therefore, this
document also redefines the [RFC8955] AS_PATH validation procedure
referenced above as follows:

   BGP Flow Specification implementations MUST enforce that the AS in
   the left-most position of the AS_PATH attribute of a Flow
   Specification route received via the External 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 NLRI.

Explanation:

   For clarity, the AS in the left-most position of the AS_PATH means
   the AS that was last added to the AS_SEQUENCE.

   This proposed modification enables the exchange of BGP Flow
   Specification NLRIs at Internet exchanges with BGP route servers
   while at the same time, for security reasons, prevents an eBGP
   peer from advertising an inter-domain Flow Specification for a
   destination prefix that it does not provide reachability
   information for.

   Comparing only the last ASes added is sufficient for eBGP learned
   Flow Specification NLRIs.  Requiring a full AS_PATH match would
   limit origination of inter-domain Flow Specifications to the
   origin AS of the best-match unicast route for the destination
   prefix embedded in the Flow Specification only.  As such, a full
   AS_PATH validity check may prevent transit ASes from originating
   inter-domain Flow Specifications, which is not desirable.

   Note, however, that not checking the full AS_PATH allows any rogue
   or misconfigured AS the ability to originate undesired Flow
   Specifications.  This is a security BGP threat, but out of the
   scope of this document.

   Redefinition of this AS_PATH validation rule for a Flow
   Specification does not mean that the original rule in [RFC8955]
   cannot be enforced as well.  Its enforcement remains optional per
   [RFC4271] section 6.3.  That is, a BGP speaker can enforce the
   first AS in the AS_PATH to be the same as the neighbor AS for any
   address-family route (including a Flow Specification address-
   family route).

   Using the new rule to validate a Flow Specification route received
   from an External Border Gateway Protocol (eBGP) peer belonging to
   the same local domain (in the case of a confederation) is out of
   the scope of this document.  Note that although it's possible, its
   utility is dubious.  Although it is conceivable that a router in
   the same local domain (both iBGP and eBGP within the same local
   domain) could send a rogue update, only eBGP (outside the local
   domain) risk is considered within this document (in the same
   spirit of the mentioned beforehand AS_PATH validation in
   [RFC4271]).

5.  Topology Considerations

   [RFC8955] indicates that the originator may refer to the originator
   path attribute (ORIGINATOR_ID) or (if the attribute is not present)
   the transport address of the peer from which the BGP speaker received
   the update.  If the latter applies, a network should be designed so
   it has a congruent topology amongst unicast routes and Flow
   Specification routes.  By congruent topology, it is understood that
   the two routes (i.e. the Flow Specification route and its best-match
   unicast route) are learned from the same peer across the AS.  That
   would likely not be true, for instance, if some peers only negotiated
   one type of address-family or if each address-family had a different
   set of policies.

   With the additional second condition (b.2) in the validation
   procedure, non-congruent topologies are supported within the local
   domain if the Flow Specification is originated within the local
   domain.

   Explanation:

      Consider the following scenarios of a non-congruent topology
      without the second condition (b.2) being added to the validation
      procedure:

      1.  Consider a topology with two BGP speakers with two iBGP
          peering sessions between them, one for unicast and one for
          Flow Specification.  This is a non-congruent topology.  Let's
          assume that the ORIGINATOR_ID attribute was not received (e.g.
          a route reflector receiving routes from its clients).  In this
          case, the Flow Specification validation procedure will fail
          because of the first condition (b.1).

      2.  Consider a confederation of ASes with local AS X and local AS
          Y (both belonging to the same local domain), and a given BGP
          speaker X1 inside local AS X.  The ORIGINATOR_ID attribute is
          not advertised when propagating routes across local ASes.
          Let's assume the Flow Specification route is received from
          peer Y1 and the best-match unicast route is received from peer
          Y2.  Both peers belong to local AS Y.  The Flow Specification
          validation procedure will also fail because of the first
          condition (b.1).

      In the scenarios above, if Flow Specifications are originated in
      the same local domain, the AS_PATH will be empty or contain only
      an AS_CONFED_SEQUENCE segment.  Condition (b.2) will evaluate to
      true.  Therefore, using the second condition (b.2), as defined by
      this document, guarantees that the overall validation procedure

will pass.  Thus, non-congruent topologies are supported if the
Flow Specification is originated in the same local domain.

Flow Specifications originated in a different local domain sill
need a congruent topology.  The reason is that the second
condition (b.2) evaluates to false and only the first condition
(b.1) is evaluated.

## 6.  IANA Considerations

This memo includes no request to IANA.

## 7.  Security Considerations

This document updates the route feasibility validation procedures for
Flow Specifications learned from iBGP peers and through route
servers.  This change is in line with the procedures described in
[RFC8955] and, thus, the security characteristics equivalent to the
existing security properties of BGP unicast routing are maintained.

The security considerations discussed in [RFC8955] apply to this
specification as well.

This document makes the original AS_PATH validation rule ([RFC4271]
section 6.3) again optional (section 4.2) for Flow Specification
Address Family (the rule is no longer mandatory as it was specified
by [RFC8955]).  If that original rule is not enforced for Flow
Specification it may introduce some new security risks.  A peer (or a
client of a route server peer) in AS X could advertise a rogue Flow
Specification route whose first AS in AS_PATH was Y (assume Y is the
first AS in the AS_PATH of the best-match unicast route).  This risk
is impossible to prevent if the Flow Specification route is received
from a route server peer.  If that peer is known for a fact not to be
a route server, that optional rule SHOULD be enforced for Flow
Specification routes.  Note that identifying those peers that are
route servers may suppose an operational challenge.  If the condition
of the peer is unknown, the rule SHOULD not be enforced.

A route server itself may be in a good position to enforce the
AS_PATH validation rule described in the previous paragraph.  If a
route server knowns it's not peering with any other route server, it
can enforce the AS_PATH validation rule across all its peers.  If, in
addition to that, the route server is trusted, the security threat
described above disappears.

BGP updates learned from iBGP peers are considered trusted, so the
Traffic Flow Specifications contained in BGP updates are also
considered trusted.  Therefore, it is not required to validate that

the originator of an intra-domain Traffic Flow Specification matches
the originator of the best-match unicast route for the destination
prefix embedded in that Flow Specification.  Note that this
trustworthy consideration is not absolute and the new possibility
than an iBGP speaker could send a rogue Flow Specification is
introduced.

The changes in Section 4.1 don't affect the validation procedures for
eBGP-learned routes.

## 8.  Acknowledgements

The authors would like to thank Han Nguyen for his direction on this
work as well as Waqas Alam, Keyur Patel, Robert Raszuk, Eric Rosen
and Shyam Sethuram for their review comments.  Shyam Sethuram for
their review comments.

## 9.  References

### 9.1.  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,
            <https://www.rfc-editor.org/info/rfc2119>.

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

[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>.

[RFC7947]   Jasinska, E., Hilliard, N., Raszuk, R., and N. Bakker,
            "Internet Exchange BGP Route Server", RFC 7947,
            DOI 10.17487/RFC7947, September 2016,
            <https://www.rfc-editor.org/info/rfc7947>.

[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>.

## 9.2.  Informative References

   [I-D.ietf-idr-deprecate-as-set-confed-set]
              Kumari, W., Sriram, K., Hannachi, L., and J. Haas,
              "Deprecation of AS_SET and AS_CONFED_SET in BGP", draft-
              ietf-idr-deprecate-as-set-confed-set-05 (work in
              progress), March 2021.

   [RFC6472]  Kumari, W. and K. Sriram, "Recommendation for Not Using
              AS_SET and AS_CONFED_SET in BGP", BCP 172, RFC 6472,
              DOI 10.17487/RFC6472, December 2011,
              <https://www.rfc-editor.org/info/rfc6472>.

Authors' Addresses

   James Uttaro
   AT&T
   200 S. Laurel Ave
   Middletown, NJ  07748
   USA

   Email: ju1738@att.com


   Juan Alcaide
   Cisco
   7100 Kit Creek Road
   Research Triangle Park, NC  27709
   USA

   Email: jalcaide@cisco.com


   Clarence Filsfils
   Cisco

   Email: cf@cisco.com

David Smith
Cisco
111 Wood Ave South
Iselin, NJ  08830
USA

Email: djsmith@cisco.com


Pradosh Mohapatra
Sproute Networks

Email: mpradosh@yahoo.com