Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 15, 2021

Y. Gu
Huawei
H. Chen
China Telecom Co., Ltd.
D. Ma
ZDNS
S. Zhuang
Huawei

September 11, 2020

BMP for BGP Route Leak Detection
draft-gu-grow-bmp-route-leak-detection-04

Abstract

According to RFC7908 [RFC7908], Route leaks refer to the case that the delivery range of route advertisements is beyond the expected range. For many current security protection solutions, the ISPs (Internet Service Providers) are focusing on finding ways to prevent the happening of BGP [RFC4271] route leaks. However, the real-time route leak detection if any occurs is important as well, and serves as the basis for leak mitigation. This document extends the BGP Monitoring Protocol (BMP) [RFC7854] to provide a routing security scheme suitable for ISPs to detect BGP route leaks at the prefix level.

Requirements 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 RFC 2119 [RFC2119].

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 March 15, 2021.

Table of Contents

1. Terminology

BMP: BGP Monitoring Protocol

BMS: BGP Monitoring Station

C2P: Customer to Provider

ISP: Internet Service Provider

P2C: Provider to Customer

P2P: Peer to Peer

RIB: Routing Information Base

RLP: Route Leak Protection

RLD: Route Leak Detection

2. Introduction

RFC7908 [RFC7908]defines "Route Leak" as: A route leak is the propagation of routing announcement(s) beyond their intended scope, which can result in possible situations such as eavesdropping, device overload, routing black hole and so on. More specifically, the intended scope of route announcements is usually defined by local route filtering/distribution policies within devices. These policies are designed to realise the pair-wise peering business relationships between ASes (autonomous systems), which include Customer to Provider (C2P), Peer to Peer (Peer to Peer), and Provider to Customer (P2C). In a C2P relationship, the customer pays the transit provider for traffic sent between the two ASes. In return, the customer gains access to the ASes that the transit provider can reach, including those which the transit provider reaches through its own transit providers. In a P2P relationship, the peering ASes gain access to each other's customers, typically without either AS paying the other AS Relationships, Customer Cones, and Validation [Luckie].

More precisely, the route leaks we discuss in this draft, referring to Type 1, 2, 3, and 4 Route Leaks defined in RFC7908 [RFC7908], can be summarized as: a route leak occurs when a route received from a transit provider or a lateral peer is propagated to another transit provider or a lateral peer.

2.1. Actions Against Route Leaks

There are serveral types of approaches against route leak from different perspectives. In this draft, we mainly discuss the following three types:

- Route leak prevention: The approach to prevent route leak from happening. Commonly used methods inlcude inbound/outbound prefix/peer/AS filtering policies configured at the ingress/egress nodes of ASes per the propagation of BGP routes.

- Route leak detection: The approach to detect the existence of route leaks that happen at either the local AS, or upstream AS per the propagation of BGP routes. An intuitive way of detecting route leak is by checking the business relationship information at

both the ingress and egress nodes of the local AS along the BGP route propagation path with the route leak violation rules defined in RFC7908 [RFC7908]. Thus, it requires the knowledge of the actual route propagation trace, as well as the resulting business relationship information at the ingress and egress nodes. With the above information collected, the analysis can be done by the routing device or a centralized server. This draft specifies one such method.

- Route leak mitigation: The approach to mitigate route leaks that already happened at either the local AS, or upstream AS per the propagation of BGP routes. Commonly used methods include reject, drop or stop propagating the invalid routes once detected the existence of leaks.

The above mentioned actions can be used seperately or combinely, depending on the entities (routing devices, network manager) that execute the actions, and the relative positions of the executing entities from the leaking point (local or downstream).

2.2. Challenges of the Current Actions against Route Leaks

draft-ietf-idr-bgp-open-policy [I-D.ietf-idr-bgp-open-policy] updates the BGP Open negotiation process with a new BGP capability to exchange the BGP Roles between two BGP speakers, and also proposes to use a new BGP attribute, called the iOTC (Internal Only To Customer) Path attribute to mark routes according to the BGP Roles established in Open Message. The iOTC attribute of the incoming route is set at the ingress node of the local AS, and is conveyed with the BGP Update to the egress node of the local AS for outbound filtering to prevent route leaks in the local AS. This attribute is removed at the egress node before the BGP Update is sent to eBGP neighbors. For representation simplification,we use iOTC to refer to the method specified in draft-ietf-idr-bgp-open-policy [I-D.ietf-idr-bgp-open-policy].

draft-ietf-grow-route-leak-detection-mitigation [I-D.ietf-grow-route-leak-detection-mitigation] describes a route leak detection and mitigation solution based on conveying route-leak protection (RLP) information in a well-know transitive BGP community, called the RLP community. The RLP community helps with detection and mitigation of route leaks that happen at the upstream AS (per the BGP routes propagation), as an inter-AS solution. For representation simplification,we use RLP to refer to the method specified in draft-ietf-grow-route-leak-detection-mitigation [I-D.ietf-grow-route-leak-detection-mitigation].

The above two drafts provide solutions for route leak prevention, detection and mitigation. To summarize:

- iOTC is used for route leak prevention of the local AS. It does not provide the detection or mitigation of route leaks of either local As or upstream AS per the BGP routes propagation.

- iOTC is peer/AS-level route leak prevention, due to the fact the BGP Role negotiation is peer-level. It's does not provide prefix-level route leak prevention.

- RLP is used for route leak detection and mitigation of route leak that happens in the upstream AS (per the BGP Update distribution). It is prefix-level detection and mitigation.

Thus, there lacks method for local AS route leak detection.

3. Route Leak Detection (RLD) Design Considerations

Considering the challenges facing the existing approaches, this draft proposes a method called Route Leak Detection (RLD). It utilizes the BGP Monitoring Protocol (BMP) to convey the RLD information from to the BMP server to realize centralized leak detection. BMP is currently deployed by OTT and carriers to monitor the BGP routes, such as monitoring BGP Adj-RIB-In using the process defined in RFC7854 [RFC7854], and monitoring BGP Adj-RIB-Out using the process defined in RFC8671 [RFC8671]. On the other hand, the RLD information is in fact a representation of the business relationships between the local AS and its neighboring AS. It does not involve any information disclosure issue regarding third parties. Thus, a single ISP can deploy RLD without relying on any information from either other ISPs or other third parties.

4. BMP Support for RLD

4.1. RLD TLV Format

A RLD TLV is defined for the BMP Route Monitoring Message. Considering that the AS relationships are sometims per route based instead of per peer/AS based, this TLV is appended to each route, following the BGP Update Message. The order of placing the RLD TLV among other BMP supported TLVs is out of the scope of this draft. The TLV format is defined as follows:

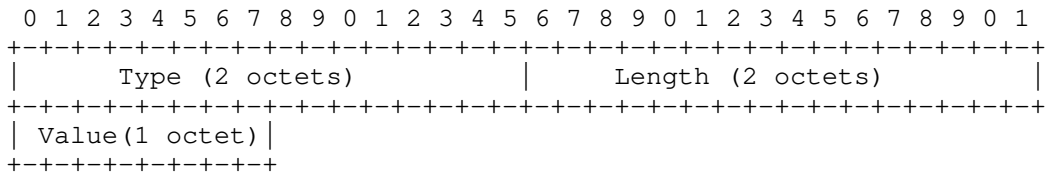


Figure 1: RLD TLV

- Type (2 octets) = TBD1, the RLD TLV represents the prefix-level business relationship between the transmitter AS and the receiver AS. The local AS is a transmitter or a receiver, depending on if the route is an incoming route from a neighbor AS or an outgoing route to a neighbor AS.

- Length (2 octets): Defines the length of the Value filed. It SHOULD be set to 0x01, considering the Value field is of 1 octect fixed length.

- Value (1 octet): Currently 4 values are defined to represent the business relationships, which are specified in Table 1.

Value	Business Relationship
0	P2C
1	C2P
2	P2P
3	I2I

Table 1: Business relationship value

4.2. RLD TLV Usage

The RLD TLV, presenting the business relationship between the neighbor AS and the local AS of the incoming route, SHOULD be prepended to the Adj-rib-in at the ingress node of the local AS. The RLD TLV, representing the business relationship between the local AS and the neighbor AS of the outgoing route, SHOULD also be prepended to the Adj-rib-out at the egress node of the local AS. The BMP server, by analyzing the above two RLD TLVs of the same route, can use the rules defined in RFC7908 [RFC7908] to detect the existence of any route leak. As example is shown in Figure 2.

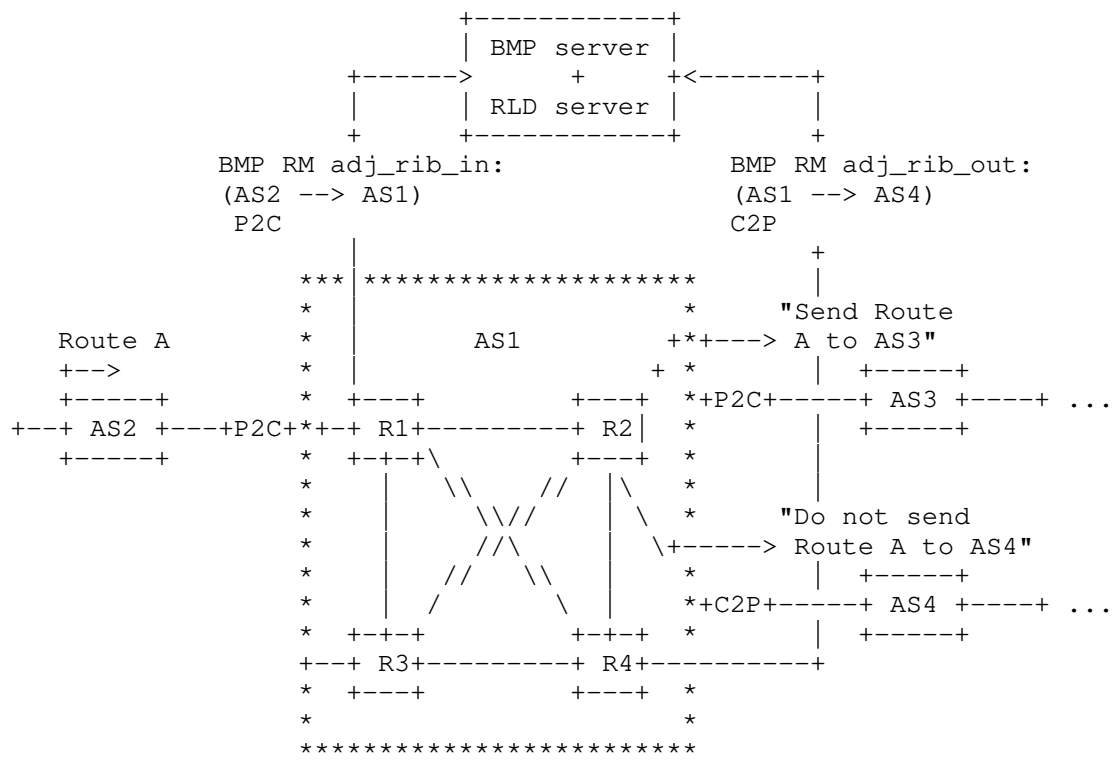


Figure 2: RLD depolyment by a single ISP

As shown in Figure 2, with the RLD TLV attached to each Route Monitoring Message, the RLD server (also working as the BMP server) combines the BMP adj_rib_in message collected from R1 and the BMP adj_rib_out message collected from R4 to decide if there's a route leak. For example, if the RLD TLV in R1's adj_rib_in message indicates a value of "0", and the RLD TLV in R4's adj_rib_out message indicates a value of "1", then the RLD server knows there exists a route leak.

4.3. Coordination with iOTC and RLP

RLD can be used as a complementary method to the existing methods against route leaks. More specifically, RLD can coordination with both iOTC and RLP.

- With the settlement of the iOTC draft, the iOTC attribute is naturally included in the BGP Update and can be collected to the BMP server without BMP extension. With the RLD TLV collected also

by BMP (more specifically, the iBGP adj-rib-in of the ingress node), the BMP server can do validate the consistency of the iOTC attribute with the RLD. If contradiction detected, the BMP server may further check the bussiness contract for the actual business relationship.

- For special prefixes that does not obey the peer/AS level business relationship negotiated through BGP Open Message, the BMP server can use the RLD TLV to detect such routes since the RLD TLV is set at prefix level.

- For devices that do not support RLP, using RLD to collect the BGP routes, which conveys the RLD information from upstream ASes, allows the BMP server to detect and mitigate the route leaks that happen in the upstream AS. In other words, the detection and mitigation process can be also done in the BMP server, should the BMP server collects the BGP Update messages at the ingress or egress nodes.

5. Acknowledgements

6. Contributors

Haibo Wang

Huawei

Email: rainsword.wang@huawei.com

7. IANA Considerations

This document defines the following new BMP Route Monitoring message TLV type (Section 4.1):

- Type = TBD1, the RLD TLV represents the prefix-level business relationship between the transmitter AS and the receiver AS. The local AS is a transmitter or a receiver, depending on if the route is an incoming route from a neighbor AS or an outgoing route to a neighbor AS.

8. Security Considerations

It is not believed that this document adds any additional security considerations.

9. References

9.1. Normative References

[I-D.ietf-grow-route-leak-detection-mitigation]
Sriram, K. and A. Azimov, "Methods for Detection and Mitigation of BGP Route Leaks", draft-ietf-grow-route-leak-detection-mitigation-03 (work in progress), July 2020.

[I-D.ietf-idr-bgp-open-policy]
Azimov, A., Bogomazov, E., Bush, R., Patel, K., and K. Sriram, "Route Leak Prevention using Roles in Update and Open messages", draft-ietf-idr-bgp-open-policy-13 (work in progress), July 2020.

[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>.

[RFC7854] Scudder, J., Ed., Fernando, R., and S. Stuart, "BGP Monitoring Protocol (BMP)", RFC 7854, DOI 10.17487/RFC7854, June 2016, <https://www.rfc-editor.org/info/rfc7854>.

[RFC7908] Sriram, K., Montgomery, D., McPherson, D., Osterweil, E., and B. Dickson, "Problem Definition and Classification of BGP Route Leaks", RFC 7908, DOI 10.17487/RFC7908, June 2016, <https://www.rfc-editor.org/info/rfc7908>.

[RFC8671] Evens, T., Bayraktar, S., Lucente, P., Mi, P., and S. Zhuang, "Support for Adj-RIB-Out in the BGP Monitoring Protocol (BMP)", RFC 8671, DOI 10.17487/RFC8671, November 2019, <https://www.rfc-editor.org/info/rfc8671>.

9.2. Informative References

[Luckie] claffy, M. L. M. L. A. D. V. G. K., "AS Relationships, Customer Cones, and Validation", October 2013.

[Siddiqui]
Ramirez, M. S. S. D. M. M. Y. R. S. X. M. W., "Route Leak Detection Using Real-Time Analytics on local BGP Information", 2014.

Authors' Addresses

Yunan Gu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: guyunan@huawei.com

Huanan Chen
China Telecom Co., Ltd.
109 Zhongshan W Ave
Guangzhou 510630
China

Email: chenhn8.gd@chinatelecom.cn

Di Ma
ZDNS
4 South 4th St. Zhongguancun
Beijing, Haidian
China

Email: madi@zdns.cn

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

Email: zhuangshunwan@huawei.com

Global Routing Operations
Internet-Draft
Updates: 7854 (if approved)
Intended status: Standards Track
Expires: February 6, 2020

T. Evens
S. Bayraktar
Cisco Systems
P. Lucente
NTT Communications
P. Mi
Tencent
S. Zhuang
Huawei

August 5, 2019

Support for Adj-RIB-Out in BGP Monitoring Protocol (BMP)
draft-ietf-grow-bmp-adj-rib-out-07

Abstract

The BGP Monitoring Protocol (BMP) defines access to only the Adj-RIB-In Routing Information Bases (RIBs). This document updates the BGP Monitoring Protocol (BMP) RFC 7854 by adding access to the Adj-RIB-Out RIBs. It adds a new flag to the peer header to distinguish Adj-RIB-In and Adj-RIB-Out.

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 February 6, 2020.

Table of Contents

1. Introduction

BGP Monitoring Protocol (BMP) defines monitoring of the received (e.g., Adj-RIB-In) Routing Information Bases (RIBs) per peer. The Adj-RIB-In pre-policy conveys to a BMP receiver all RIB data before any policy has been applied. The Adj-RIB-In post-policy conveys to a BMP receiver all RIB data after policy filters and/or modifications have been applied. An example of pre-policy versus post-policy is when an inbound policy applies attribute modification or filters. Pre-policy would contain information prior to the inbound policy changes or filters of data. Post policy would convey the changed data or would not contain the filtered data.

Monitoring the received updates that the router received before any policy has been applied is the primary level of monitoring for most use-cases. Inbound policy validation and auditing is the primary use-case for enabling post-policy monitoring.

In order for a BMP receiver to receive any BGP data, the BMP sender (e.g., router) needs to have an established BGP peering session and actively be receiving updates for an Adj-RIB-In.

Being able to only monitor the Adj-RIB-In puts a restriction on what data is available to BMP receivers via BMP senders (e.g., routers). This is an issue when the receiving end of the BGP peer is not enabled for BMP or when it is not accessible for administrative reasons. For example, a service provider advertises prefixes to a customer, but the service provider cannot see what it advertises via BMP. Asking the customer to enable BMP and monitoring of the Adj-RIB-In is not feasible.

BGP Monitoring Protocol (BMP) RFC 7854 [RFC7854] only defines Adj-RIB-In being sent to BMP receivers. This document updates the per-peer header in section 4.2 of [RFC7854] by adding a new flag to distinguish Adj-RIB-In versus Adj-RIB-Out. BMP senders use the new flag to send either Adj-RIB-In or Adj-RIB-Out.

Adding Adj-RIB-Out provides the ability for a BMP sender to send to BMP receivers what it advertises to BGP peers, which can be used for outbound policy validation and to monitor routes that were advertised.

2. 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 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Definitions

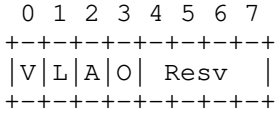
- Adj-RIB-Out: As defined in [RFC4271], "The Adj-RIBs-Out contains the routes for advertisement to specific peers by means of the local speaker's UPDATE messages."

- Pre-Policy Adj-RIB-Out: The result before applying the outbound policy to an Adj-RIB-Out. This normally would match what is in the local RIB.

- Post-Policy Adj-RIB-Out: The result of applying outbound policy to an Adj-RIB-Out. This MUST convey to the BMP receiver what is actually transmitted to the peer.

4. Per-Peer Header

The per-peer header has the same structure and flags as defined in section 4.2 of [RFC7854] with the following O flag addition:



- The O flag indicates Adj-RIB-In if set to 0 and Adj-RIB-Out if set to 1.

The existing flags are defined in section 4.2 of [RFC7854] and the remaining bits are reserved for future use. They MUST be transmitted as 0 and their values MUST be ignored on receipt.

When the O flag is set to 1, the following fields in the Per-Peer Header are redefined:

- Peer Address: The remote IP address associated with the TCP session over which the encapsulated PDU is sent.

- Peer AS: The Autonomous System number of the peer to which the encapsulated PDU is sent.

- Peer BGP ID: The BGP Identifier of the peer to which the encapsulated PDU is sent.

- Timestamp: The time when the encapsulated routes were advertised (one may also think of this as the time when they were installed in the Adj-RIB-Out), expressed in seconds and microseconds since midnight (zero hour), January 1, 1970 (UTC). If zero, the time is unavailable. Precision of the timestamp is implementation-dependent.

5. Adj-RIB-Out

5.1. Post-Policy

The primary use-case in monitoring Adj-RIB-Out is to monitor the updates transmitted to a BGP peer after outbound policy has been applied. These updates reflect the result after modifications and filters have been applied (e.g., Adj-RIB-Out Post-Policy). Some

attributes are set when the BGP message is transmitted, such as next-hop. Adj-RIB-Out Post-Policy MUST convey to the BMP receiver what is actually transmitted to the peer.

The L flag MUST be set to 1 to indicate post-policy.

5.2. Pre-Policy

Similarly to Adj-RIB-In policy validation, pre-policy Adj-RIB-Out can be used to validate and audit outbound policies. For example, a comparison between pre-policy and post-policy can be used to validate the outbound policy.

Depending on BGP peering session type (IBGP, IBGP route reflector client, EBGP, BGP confederations, Route Server Client) the candidate routes that make up the Pre-Policy Adj-RIB-Out do not contain all local-rib routes. Pre-Policy Adj-RIB-Out conveys only routes that are available based on the peering type. Post-Policy represents the filtered/changed routes from the available routes.

Some attributes are set only during transmission of the BGP message, i.e., Post-Policy. It is common that next-hop may be null, loopback, or similar during pre-policy phase. All mandatory attributes, such as next-hop, MUST be either ZERO or have an empty length if they are unknown at the Pre-Policy phase completion. The BMP receiver will treat zero or empty mandatory attributes as self-originated.

The L flag MUST be set to 0 to indicate pre-policy.

6. BMP Messages

Many BMP messages have a per-peer header but some are not applicable to Adj-RIB-In or Adj-RIB-Out monitoring, such as peer up and down notifications. Unless otherwise defined, the O flag should be set to 0 in the per-peer header in BMP messages.

6.1. Route Monitoring and Route Mirroring

The O flag MUST be set accordingly to indicate if the route monitor or route mirroring message conveys Adj-RIB-In or Adj-RIB-Out.

6.2. Statistics Report

The Statistics report message has a Stat Type field to indicate the statistic carried in the Stat Data field. Statistics report messages are not specific to Adj-RIB-In or Adj-RIB-Out and MUST have the O flag set to zero. The O flag SHOULD be ignored by the BMP receiver.

The following new statistic types are added:

- Stat Type = 14: (64-bit Gauge) Number of routes in Adj-RIBs-Out Pre-Policy.

- Stat Type = 15: (64-bit Gauge) Number of routes in Adj-RIBs-Out Post-Policy.

- Stat Type = 16: Number of routes in per-AFI/SAFI Adj-RIB-Out Pre-Policy. The value is structured as: 2-byte Address Family Identifier (AFI), 1-byte Subsequent Address Family Identifier (SAFI), followed by a 64-bit Gauge.

- Stat Type = 17: Number of routes in per-AFI/SAFI Adj-RIB-Out Post-Policy. The value is structured as: 2-byte Address Family Identifier (AFI), 1-byte Subsequent Address Family Identifier (SAFI), followed by a 64-bit Gauge.

6.3. Peer Down and Up Notifications

Peer Up and Down notifications convey BGP peering session state to BMP receivers. The state is independent of whether or not route monitoring or route mirroring messages will be sent for Adj-RIB-In, Adj-RIB-Out, or both. BMP receiver implementations SHOULD ignore the O flag in Peer Up and Down notifications.

6.3.1. Peer Up Information

The following Peer Up message Information TLV type is added:

- Type = 4: Admin Label. The Information field contains a free-form UTF-8 string whose byte length is given by the Information Length field. The value is administratively assigned. There is no requirement to terminate the string with null or any other character.

  Multiple admin labels can be included in the Peer Up notification. When multiple admin labels are included the BMP receiver MUST preserve their order.

  The TLV is optional.

7. Other Considerations

7.1. Peer and Update Groups

Peer and update groups are used to group updates shared by many peers. This is a level of efficiency in implementations, not a true representation of what is conveyed to a peer in either Pre-Policy or Post-Policy.

One of the use-cases to monitor Adj-RIB-Out Post-Policy is to validate and continually ensure the egress updates match what is expected. For example, wholesale peers should never have routes with community X:Y sent to them. In this use-case, there may be hundreds of wholesale peers but a single peer could have represented the group.

From a BMP perspective, this should be simple to include a group name in the Peer Up, but it is more complex than that. BGP implementations have evolved to provide comprehensive and structured policy grouping, such as session, AFI/SAFI, and template-based based group policy inheritances.

This level of structure and inheritance of polices does not provide a simple peer group name or ID, such as wholesale peer.

Instead of requiring a group name to be used, a new administrative label informational TLV (Section 6.3.1) is added to the Peer Up message. These labels have administrative scope relevance. For example, labels "type=wholesale" and "region=west" could be used to monitor expected policies.

Configuration and assignment of labels to peers is BGP implementation specific.

8. Security Considerations

The same considerations as in section 11 of [RFC7854] apply to this document. Implementations of this protocol SHOULD require to establish sessions with authorized and trusted monitoring devices. It is also believed that this document does not add any additional security considerations.

9. IANA Considerations

This document requests that IANA assign the following new parameters to the BMP parameters name space [1].

9.1. BMP Peer Flags

This document defines the following per-peer header flags (Section 4):

- Flag 3 as O flag: The O flag indicates Adj-RIB-In if set to 0 and Adj-RIB-Out if set to 1.

9.2. BMP Statistics Types

This document defines four statistic types for statistics reporting (Section 6.2):

- Stat Type = 14: (64-bit Gauge) Number of routes in Adj-RIBs-Out Pre-Policy.

- Stat Type = 15: (64-bit Gauge) Number of routes in Adj-RIBs-Out Post-Policy.

- Stat Type = 16: Number of routes in per-AFI/SAFI Adj-RIB-Out Pre-Policy. The value is structured as: 2-byte Address Family Identifier (AFI), 1-byte Subsequent Address Family Identifier (SAFI), followed by a 64-bit Gauge.

- Stat Type = 17: Number of routes in per-AFI/SAFI Adj-RIB-Out Post-Policy. The value is structured as: 2-byte Address Family Identifier (AFI), 1-byte Subsequent Address Family Identifier (SAFI), followed by a 64-bit Gauge.

9.3. Peer Up Information TLV

This document defines the following BMP Peer Up Information TLV types (Section 6.3.1):

- Type = 4: Admin Label. The Information field contains a free-form UTF-8 string whose byte length is given by the Information Length field. The value is administratively assigned. There is no requirement to terminate the string with null or any other character.

  Multiple admin labels can be included in the Peer Up notification. When multiple admin labels are included the BMP receiver MUST preserve their order.

  The TLV is optional.

10. References

10.1. 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>.

[RFC7854] Scudder, J., Ed., Fernando, R., and S. Stuart, "BGP Monitoring Protocol (BMP)", RFC 7854, DOI 10.17487/RFC7854, June 2016, <https://www.rfc-editor.org/info/rfc7854>.

[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>.

10.2. URIs

[1] https://www.iana.org/assignments/bmp-parameters/bmp-parameters.xhtml

Acknowledgements

The authors would like to thank John Scudder and Mukul Srivastava for their valuable input.

Contributors

Manish Bhardwaj
Cisco Systems
3700 Cisco Way
San Jose, CA 95134
USA

Email: manbhard@cisco.com

Xianyuzheng
Tencent
Tencent Building, Kejizhongyi Avenue,
Hi-techPark, Nanshan District,Shenzhen 518057, P.R.China

Weiguo
Tencent
Tencent Building, Kejizhongyi Avenue,
Hi-techPark, Nanshan District,Shenzhen 518057, P.R.China

Shugang cheng
H3C

Authors' Addresses

Tim Evens
Cisco Systems
2901 Third Avenue, Suite 600
Seattle, WA 98121
USA

Email: tievens@cisco.com

Serpil Bayraktar
Cisco Systems
3700 Cisco Way
San Jose, CA 95134
USA

Email: serpil@cisco.com

Paolo Lucente
NTT Communications
Siriusdreef 70-72
Hoofddorp, WT 2132
NL

Email: paolo@ntt.net

Penghui Mi
Tencent
Tengyun Building,Tower A ,No. 397 Tianlin Road
Shanghai 200233
China

Email: kevinmi@tencent.com

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

Email: zhuangshunwan@huawei.com

Global Routing Operations
Internet-Draft
Updates: 7854 (if approved)
Intended status: Standards Track
Expires: 20 May 2021

T. Evens
S. Bayraktar
M. Bhardwaj
Cisco Systems
P. Lucente
NTT Communications

16 November 2020

Support for Local RIB in BGP Monitoring Protocol (BMP)
draft-ietf-grow-bmp-local-rib-08

Abstract

The BGP Monitoring Protocol (BMP) defines access to various Routing Information Bases (RIBs). This document updates BMP (RFC 7854) by adding access to the Local Routing Information Base (Loc-RIB), as defined in RFC 4271. The Loc-RIB contains the routes that have been selected by the local BGP speaker's Decision Process.

Table of Contents

1. Introduction

This document defines a mechanism to monitor the BGP Loc-RIB state of remote BGP instances without the need to establish BGP peering sessions. BMP [RFC7854] does not define a method to send the BGP instance Loc-RIB. It does define in section 8.2 of [RFC7854] locally originated routes, but these routes are defined as the routes originated into BGP. For example, locally sourced routes that are redistributed.

Figure 1 shows the flow of received routes from one or more BGP peers into the Loc-RIB.

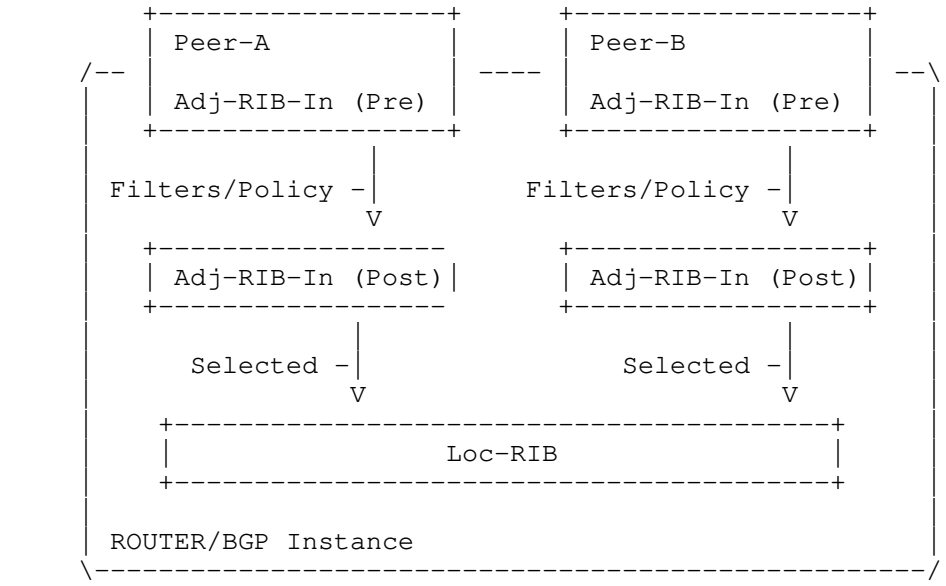


Figure 1: BGP peering Adj-RIBs-In into Loc-RIB

As shown in Figure 2, Locally originated section 9.4 of [RFC4271] follows a similar flow where the redistributed or otherwise originated routes get installed into the Loc-RIB based on the decision process selection.

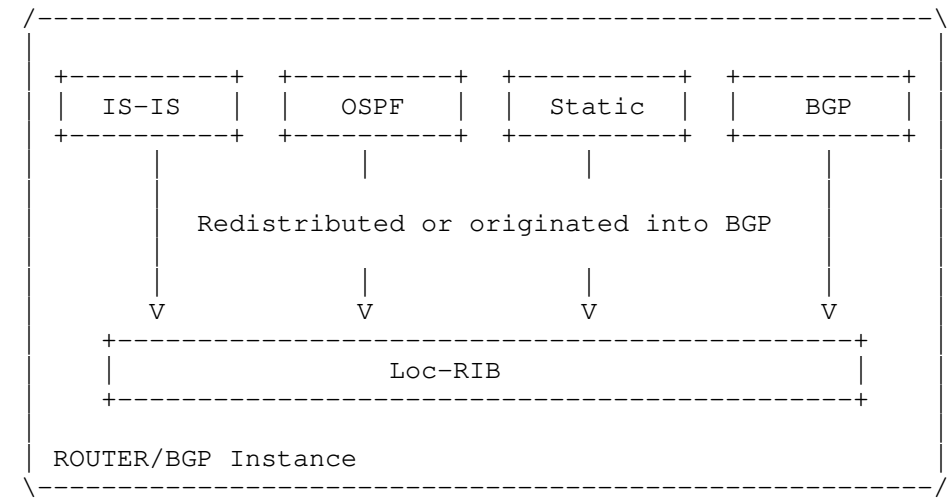


Figure 2: Locally Originated into Loc-RIB

The following are some use-cases for Loc-RIB access:

* The Adj-RIB-In for a given peer Post-Policy may contain hundreds of thousands of routes, with only a handful of routes selected and installed in the Loc-RIB after best-path selection. Some monitoring applications, such as ones that need only to correlate flow records to Loc-RIB entries, only need to collect and monitor the routes that are actually selected and used.

  Requiring the applications to collect all Adj-RIB-In Post-Policy data forces the applications to receive a potentially large unwanted data set and to perform the BGP decision process selection, which includes having access to the IGP next-hop metrics. While it is possible to obtain the IGP topology information using BGP-LS, it requires the application to implement SPF and possibly CSPF based on additional policies. This is overly complex for such a simple application that only needed to have access to the Loc-RIB.

* It is common to see frequent changes over many BGP peers, but those changes do not always result in the router's Loc-RIB changing. The change in the Loc-RIB can have a direct impact on the forwarding state. It can greatly reduce time to troubleshoot and resolve issues if operators had the history of Loc-RIB changes. For example, a performance issue might have been seen for only a duration of 5 minutes. Post troubleshooting this issue without Loc-RIB history hides any decision based routing changes that might have happened during those five minutes.

* Operators may wish to validate the impact of policies applied to Adj-RIB-In by analyzing the final decision made by the router when installing into the Loc-RIB. For example, in order to validate if multi-path prefixes are installed as expected for all advertising peers, the Adj-RIB-In Post-Policy and Loc-RIB needs to be compared. This is only possible if the Loc-RIB is available. Monitoring the Adj-RIB-In for this router from another router to derive the Loc-RIB is likely to not show same installed prefixes. For example, the received Adj-RIB-In will be different if add-paths is not enabled or if maximum number of equal paths are different from Loc-RIB to routes advertised.

This document adds Loc-RIB to the BGP Monitoring Protocol and replaces Section 8.2 of [RFC7854] Locally Originated Routes.

1.1. Alternative Method to Monitor Loc-RIB

Loc-RIB is used to build Adj-RIB-Out when advertising routes to a peer. It is therefore possible to derive the Loc-RIB of a router by monitoring the Adj-RIB-In Pre-Policy from another router. At scale this becomes overly complex and error prone.

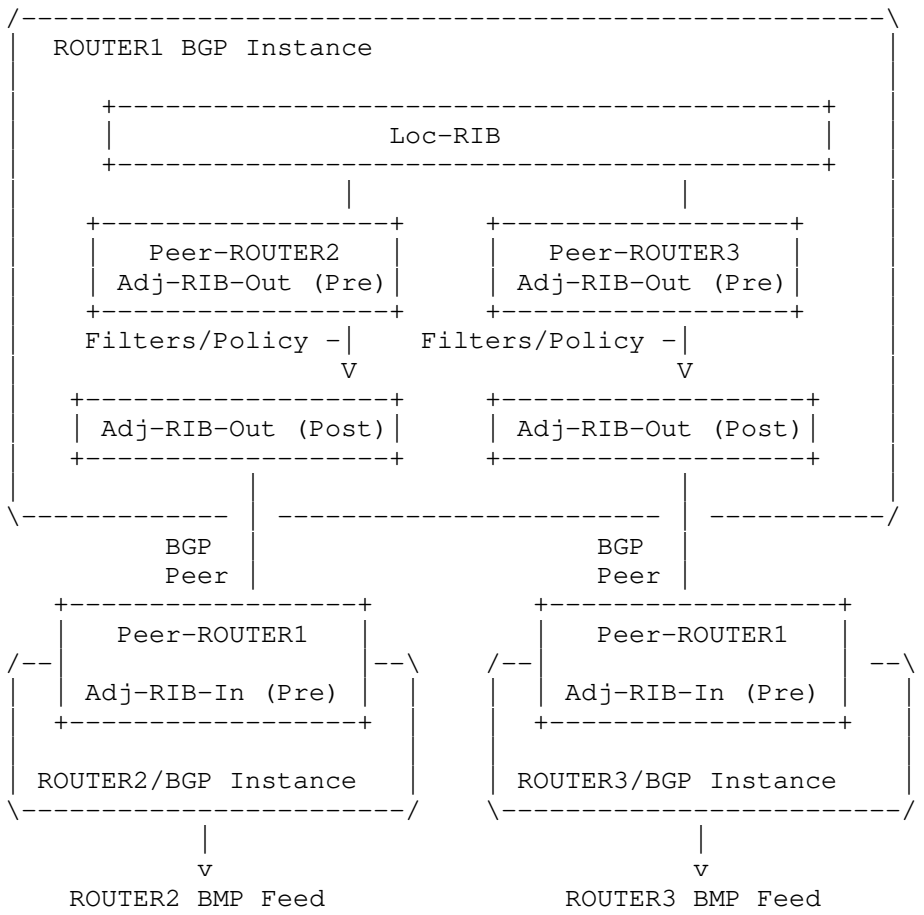


Figure 3: Alternative method to monitor Loc-RIB

The setup needed to monitor the Loc-RIB of a router requires another router with a peering session to the target router that is to be monitored. As shown in Figure 3, the target router Loc-RIB is advertised via Adj-RIB-Out to the BMP router over a standard BGP peering session. The BMP router then forwards Adj-RIB-In Pre-Policy to the BMP receiver.

The current method introduces the need for additional resources:

* Requires at least two routers when only one router was to be monitored.

* Requires additional BGP peering to collect the received updates when peering may have not even been required in the first place. For example, VRFs with no peers, redistributed BGP-LS with no peers, segment routing egress peer engineering where no peers have link-state address family enabled.

Complexities introduced with current method in order to derive (e.g. correlate) peer to router Loc-RIB:

* Adj-RIB-Out received as Adj-RIB-In from another router may have a policy applied that filters, generates aggregates, suppresses more specifics, manipulates attributes, or filters routes. Not only does this invalidate the Loc-RIB view, it adds complexity when multiple BMP routers may have peering sessions to the same router. The BMP receiver user is left with the error prone task of identifying which peering session is the best representative of the Loc-RIB.

* BGP peering is designed to work between administrative domains and therefore does not need to include internal system level information of each peering router (e.g. the system name or version information). In order to derive a Loc-RIB to a router, the router name or other system information is needed. The BMP receiver and user are forced to do some type of correlation using what information is available in the peering session (e.g. peering addresses, ASNs, and BGP-IDs). This leads to error prone correlations.

* The BGP-IDs and session addresses to router correlation requires additional data, such as router inventory. This additional data provides the BMP receiver the ability to map and correlate the BGP-IDs and/or session addresses, but requires the BMP receiver to somehow obtain this data outside of BMP. How this data is obtained and the accuracy of the data directly effects the integrity of the correlation.

2. 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 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Definitions

* BGP Instance: it refers to an instance of an instance of BGP-4 [RFC4271] and considerations in section 8.1 of [RFC7854] do apply to it.

* Adj-RIB-In: As defined in [RFC4271], "The Adj-RIBs-In contains unprocessed routing information that has been advertised to the local BGP speaker by its peers." This is also referred to as the pre-policy Adj-RIB-In in this document.

* Adj-RIB-Out: As defined in [RFC4271], "The Adj-RIBs-Out contains the routes for advertisement to specific peers by means of the local speaker's UPDATE messages."

* Loc-RIB: As defined in section 9.4 of [RFC4271], "The Loc-RIB contains the routes that have been selected by the local BGP speaker's Decision Process." Note that the Loc-RIB state as monitored through BMP might also contain routes imported from other routing protocols such as an IGP, or local static routes.

* Pre-Policy Adj-RIB-Out: The result before applying the outbound policy to an Adj-RIB-Out. This normally represents a similar view of the Loc-RIB but may contain additional routes based on BGP peering configuration.

* Post-Policy Adj-RIB-Out: The result of applying outbound policy to an Adj-RIB-Out. This MUST be what is actually sent to the peer.

4. Per-Peer Header

4.1. Peer Type

A new peer type is defined for Loc-RIB to distinguish that it represents Loc-RIB with or without RD and local instances. Section 4.2 of [RFC7854] defines a Local Instance Peer type, which is for the case of non-RD peers that have an instance identifier.

This document defines the following new peer type:

* Peer Type = 3: Loc-RIB Instance Peer

4.2. Peer Flags

In section 4.2 of [RFC7854], the "locally sourced routes" comment under the L flag description is removed. Locally sourced routes MUST be conveyed using the Loc-RIB instance peer type.

The per-peer header flags for Loc-RIB Instance Peer type are defined as follows:

* The F flag indicates that the Loc-RIB is filtered. This MUST be set when only a subset of Loc-RIB routes is sent to the BMP collector.

  The remaining bits are reserved for future use. They MUST be transmitted as 0 and their values MUST be ignored on receipt.

5. Loc-RIB Monitoring

The Loc-RIB contains all routes selected by the BGP protocol Decision Process section 9.1 of [RFC4271]. These routes include those learned from BGP peers via its Adj-RIBs-In post-policy, as well as routes learned by other means section 9.4 of [RFC4271]. Examples of these include redistribution of routes from other protocols into BGP or otherwise locally originated (ie. aggregate routes).

As mentioned in Section 4.2 a subset of Loc-RIB routes MAY be sent to a BMP collector by setting the F flag.

5.1. Per-Peer Header

All peer messages that include a per-peer header MUST use the following values:

* Peer Type: Set to 3 to indicate Loc-RIB Instance Peer.

* Peer Distinguisher: Zero filled if the Loc-RIB represents the global instance. Otherwise set to the route distinguisher or unique locally defined value of the particular instance the Loc-RIB belongs to.

* Peer Address: Zero-filled. Remote peer address is not applicable. The V flag is not applicable with Loc-RIB Instance peer type considering addresses are zero-filed.

* Peer AS: Set to the BGP instance global or default ASN value.

* Peer BGP ID: Set to the BGP instance global or RD (e.g. VRF) specific router-id section 1.1 of [RFC7854].

* Timestamp: The time when the encapsulated routes were installed in The Loc-RIB, expressed in seconds and microseconds since midnight (zero hour), January 1, 1970 (UTC). If zero, the time is unavailable. Precision of the timestamp is implementation-dependent.

5.2. Peer UP Notification

Peer UP notifications follow section 4.10 of [RFC7854] with the following clarifications:

* Local Address: Zero-filled, local address is not applicable.

* Local Port: Set to 0, local port is not applicable.

* Remote Port: Set to 0, remote port is not applicable.

* Sent OPEN Message: This is a fabricated BGP OPEN message. Capabilities MUST include 4-octet ASN and all necessary capabilities to represent the Loc-RIB route monitoring messages. Only include capabilities if they will be used for Loc-RIB monitoring messages. For example, if add-paths is enabled for IPv6 and Loc-RIB contains additional paths, the add-paths capability should be included for IPv6. In the case of add-paths, the capability intent of advertise, receive or both can be ignored since the presence of the capability indicates enough that add-paths will be used for IPv6.

* Received OPEN Message: Repeat of the same Sent Open Message. The duplication allows the BMP receiver to use existing parsing.

5.2.1. Peer UP Information

The following Peer UP information TLV type is added:

* Type = 3: VRF/Table Name. The Information field contains a UTF-8 string whose value MUST be equal to the value of the VRF or table name (e.g. RD instance name) being conveyed. The string size MUST be within the range of 1 to 255 bytes.

  The VRF/Table Name TLV is optionally included. For consistency, it is RECOMMENDED that the VRF/Table Name always be included. The default value of "global" MUST be used for the default Loc-RIB instance with a zero-filled distinguisher. If the TLV is included, then it MUST also be included in the Peer Down notification.

Multiple TLVs of the same type can be repeated as part of the same message, for example to convey a filtered view of a VRF. A BMP receiver should append multiple TLVs of the same type to a set in order to support alternate or additional names for the same peer. If multiple strings are included, their ordering MUST be preserved when they are reported.

5.3. Peer Down Notification

Peer down notification MUST use reason code TBD3. Following the reason is data in TLV format. The following peer Down information TLV type is defined:

* Type = 3: VRF/Table Name. The Information field contains a UTF-8 string whose value MUST be equal to the value of the VRF or table name (e.g. RD instance name) being conveyed. The string size MUST be within the range of 1 to 255 bytes. The VRF/Table Name informational TLV MUST be included if it was in the Peer UP.

5.4. Route Monitoring

Route Monitoring messages are used for initial synchronization of the Loc-RIB. They are also used to convey incremental Loc-RIB changes.

As defined in section 4.3 of [RFC7854], "Following the common BMP header and per-peer header is a BGP Update PDU."

5.4.1. ASN Encoding

Loc-RIB route monitor messages MUST use 4-byte ASN encoding as indicated in PEER UP sent OPEN message (Section 5.2) capability.

5.4.2. Granularity

State compression and throttling SHOULD be used by a BMP sender to reduce the amount of route monitoring messages that are transmitted to BMP receivers. With state compression, only the final resultant updates are sent.

For example, prefix 10.0.0.0/8 is updated in the Loc-RIB 5 times within 1 second. State compression of BMP route monitor messages results in only the final change being transmitted. The other 4 changes are suppressed because they fall within the compression interval. If no compression was being used, all 5 updates would have been transmitted.

A BMP receiver should expect that Loc-RIB route monitoring granularity can be different by BMP sender implementation.

5.5. Route Mirroring

Route mirroring is not applicable to Loc-RIB and Route Mirroring messages SHOULD be ignored.

5.6. Statistics Report

Not all Stat Types are relevant to Loc-RIB. The Stat Types that are relevant are listed below:

* Stat Type = 8: (64-bit Gauge) Number of routes in Loc-RIB.

* Stat Type = 10: Number of routes in per-AFI/SAFI Loc-RIB. The value is structured as: 2-byte AFI, 1-byte SAFI, followed by a 64-bit Gauge.

6. Other Considerations

6.1. Loc-RIB Implementation

There are several methods for a BGP speaker to implement Loc-RIB efficiently. In all methods, the implementation emulates a peer with Peer UP and DOWN messages to convey capabilities as well as Route Monitor messages to convey Loc-RIB. In this sense, the peer that conveys the Loc-RIB is a local router emulated peer.

6.1.1. Multiple Loc-RIB Peers

There MUST be multiple emulated peers for each Loc-RIB instance, such as with VRFs. The BMP receiver identifies the Loc-RIB by the peer header distinguisher and BGP ID. The BMP receiver uses the VRF/ Table Name from the PEER UP information to associate a name to the Loc-RIB.

In some implementations, it might be required to have more than one emulated peer for Loc-RIB to convey different address families for the same Loc-RIB. In this case, the peer distinguisher and BGP ID should be the same since it represents the same Loc-RIB instance. Each emulated peer instance MUST send a PEER UP with the OPEN message indicating the address family capabilities. A BMP receiver MUST process these capabilities to know which peer belongs to which address family.

6.1.2. Filtering Loc-RIB to BMP Receivers

There maybe be use-cases where BMP receivers should only receive specific routes from Loc-RIB. For example, IPv4 unicast routes may include IBGP, EBGP, and IGP but only routes from EBGP should be sent to the BMP receiver. Alternatively, it may be that only IBGP and EBGP that should be sent and IGP redistributed routes should be excluded. In these cases where the Loc-RIB is filtered, the F flag is set to 1 to indicate to the BMP receiver that the Loc-RIB is filtered. If multiple filters are associated to the same Loc-RIB, a Table Name MUST be used in order to allow a BMP receiver to make the right associations.

6.1.3. Changes to existing BMP sessions

In case of any change that results in the alteration of behaviour of an existing BMP session, ie. changes to filtering and table names, the session MUST be bounced with a Peer DOWN/Peer UP sequence.

7. Security Considerations

The same considerations as in section 11 of [RFC7854] apply to this document. Implementations of this protocol SHOULD require to establish sessions with authorized and trusted monitoring devices. It is also believed that this document does not add any additional security considerations.

8. IANA Considerations

This document requests that IANA assign the following new parameters to the BMP parameters name space (https://www.iana.org/assignments/bmp-parameters/bmp-parameters.xhtml).

8.1. BMP Peer Type

This document defines a new peer type (Section 4.1):

* Peer Type = 3: Loc-RIB Instance Peer

8.2. BMP Peer Flags

This document defines a new flag (Section 4.2) and proposes that peer flags are specific to the peer type:

* The F flag indicates that the Loc-RIB is filtered. This indicates that the Loc-RIB does not represent the complete routing table.

8.3. Peer UP Information TLV

This document defines the following new BMP PEER UP informational message TLV types (Section 5.2.1):

* Type = 3: VRF/Table Name. The Information field contains a UTF-8 string whose value MUST be equal to the value of the VRF or table name (e.g. RD instance name) being conveyed. The string size MUST be within the range of 1 to 255 bytes.

8.4. Peer Down Reason code

This document defines the following new BMP Peer Down reason code (Section 5.3):

* Type = TBD3: Local system closed, TLV data follows.

9. 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>.

[RFC7854] Scudder, J., Ed., Fernando, R., and S. Stuart, "BGP Monitoring Protocol (BMP)", RFC 7854, DOI 10.17487/RFC7854, June 2016, <https://www.rfc-editor.org/info/rfc7854>.

[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>.

Acknowledgements

The authors would like to thank John Scudder, Jeff Haas and Mukul Srivastava for their valuable input.

Authors' Addresses

Tim Evens
Cisco Systems
2901 Third Avenue, Suite 600

Seattle, WA 98121
United States of America

Email: tievens@cisco.com

Serpil Bayraktar
Cisco Systems
3700 Cisco Way
San Jose, CA 95134
United States of America

Email: serpil@cisco.com

Manish Bhardwaj
Cisco Systems
3700 Cisco Way
San Jose, CA 95134
United States of America

Email: manbhard@cisco.com

Paolo Lucente
NTT Communications
Siriusdreef 70-72
2132 Hoofddorp
Netherlands

Email: paolo@ntt.net

Global Routing Operations
Internet-Draft
Intended status: Informational
Expires: October 26, 2020

J. Snijders
NTT
M. Stucchi
Independent
M. Aelmans
Juniper Networks

April 24, 2020

RPKI Autonomous Systems Cones: A Profile To Define Sets of Autonomous Systems Numbers To Facilitate BGP Filtering
draft-ietf-grow-rpki-as-cones-02

Abstract

This document describes a way to define groups of Autonomous System numbers in RPKI [RFC6480]. We call them AS-Cones. AS-Cones provide a mechanism to be used by operators for filtering BGP-4 [RFC4271] announcements.

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.

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 October 26, 2020.

Table of Contents

1. Introduction

The main goal of the Resource Public Key Infrastructure (RPKI) system [RFC6480] is to support improved security for the global routing system. This is achieved through the use of information stored in a distributed repository system comprised of signed objects. A

commonly used object type is the Route Object Authorisation (ROAs), which describe the relation between a prefix and its originating ASNs.

There is however no method for an operator to assert the routes for its customer networks, making it difficult to use the information carried by RPKI to create meaningful BGP-4 filters without relying on RPSL [RFC2622] as-sets.

This document introduces a new attestation object, called an AS-Cone. An AS-Cone is a digitally signed object with the goal to enable operators to define a set of customer or downstream ASNs that can be found as "right adjacencies", or transit customer networks, facilitating the construction of prefix filters for a given ASN, thus making routing more secure.

The goal of AS-Cones is to be able to recursively define all the originating ASNs that define the customer base of a given ASN, including all the transit relationships. This means that through AS-Cones, it is possible to create a tree of all the neighbour relationships for the customers of a given Autonomous System.

2. Format of AS-Cone objects

AS-Cones are composed of two types of distinct objects:

- Policy definitions; and
- The AS-Cones themselves.

These objects are stored in ASN.1 format and are digitally signed according to the same rules and conventions applied for RPKI ROA Objects ([RFC6482]).

2.1. Policy definition object

A policy definition object contains a list of the upstream and peering relationships for a given Autonomous System that need an AS-Cone to be used for filtering. For each relationship, either an AS-Cone or a plain Autonomous System Number is referenced to indicate which networks will be announced to the other end of the relationship using BGP.

The default behaviour for a neighbour, if the relationship is not explicitly described in the policy, is to only accept the networks originated by the ASN. This means that a stub ASN neither has to set up any AS-Cone, description, nor policy.

The Policy Definition object contains a field called "ContactEmail" containing the E-Mail address for which all the communication related to this policy definition should be sent to.

Only one AS-Cone or Autonomous System Number can be supplied for a given relationship. If more than one AS-Cone needs to be announced in the relationship, then it is mandatory to create a third AS-Cone that includes those two. If more than one ASN needs to be referenced, then an AS-Cone for the relationship needs to be created.

2.1.1. Naming convention for Policy definition objects

A Policy object is referenced using the Autonomous System number it refers to, preceded by the string "AS".

2.1.2. ASN.1 format of a Policy Definition object

```
ASNPolicy DEFINITIONS ::=
BEGIN
Neighbours ::= SEQUENCE OF Neighbour

Neighbour ::= SEQUENCE
{
ASN INTEGER (1..42949672965),
ASCone  VisibleString
}

Version ::= INTEGER
LastModified ::= GeneralizedTime
Created ::= GeneralizedTime
ContactEmail ::= PrintableString(SIZE (1..75))
END
```

ASN.1 format of a Policy definition object

2.1.3. Naming convention for neighbour relationships

When referring to a neighbour relationship contained in a Policy definition object, the following convention should be used:

ASX:ASY

Where X is the number of the ASN holder and Y is the number of the ASN intended to use the AS-Cone object to generate a filter.

2.2. AS-Cone definition object

An AS-Cone contains a list of the downstream customer ASNs and AS-Cones of a given ASN. The list is used to create filter lists by the networks providing transit to or having a peering relationship with the ASN.

An AS-Cone can reference another AS-Cone.

2.2.1. Adding entries in an AS-Cone object

When an entry is added, it is in the Unverified status, and its "Verified" variable is set to 0.

If an ASN is added as an entry, it becomes directly visible and usable in building prefix lists, and a notification is sent to the E-mail address contained in the "ContactEmail" field of the AS-Cone Policy Object for that Autonomous System Number. The holder of the Autonomous System Number can acknowledge the notification, in which case the "Verified" field is switched to the value of 1.

If an AS-Cone is added to the object, a notification is sent to the E-Mail address contained in the "ContactEmail" field of the AS-Cone object that is being added. If the "ContactEmail" field is blank, the notification is sent to the E-mail address contained in the "ContactEmail" field of the AS-Cone Policy Object of the ASN of which the AS-Cone is part of. Only when an acknowledgement from the holder of the object is obtained, the "Verified" field is changed to a value of 1, and the AS-Cone becomes visible.

The value of the "Verified" field is fundamental for the creation of appropriate prefix filtering rules as described later.

2.2.2. Removal of entries from an AS-Cone object

The owner of an AS-Cone can remove any entry from its object without requesting any permission from the holders of the entries being removed.

The holder of an entry in a third party AS-Cone can remove the entry by performing authentication based on the E-mail address contained in the "ContactEmail" field of the resource itself. The RIRs MUST provide means to perform this authentication via an auth code, an API, or other means. The removal of an entry SHOULD be immediate upon successful authentication.

2.2.3. Naming convention for AS-Cone objects

AS-Cones MUST have a unique name for the ASN they belong to. Names are composed of ASCII strings up to 255 characters long and cannot contain spaces.

In order for AS-Cones to be unique in the global routing system, their string name is preceded by the AS number of the ASN they are part of, followed by ":". For example, AS-Cone "EuropeanCustomers" for ASN 65530 is represented as "AS65530:EuropeanCustomers" when referenced from a third party.

2.2.4. ASN.1 format of an AS-Cone

```
ASCone DEFINITIONS ::=
BEGIN
Entities ::= SEQUENCE OF Entity

Entity CHOICE
{
        ASN INTEGER (1..4294967295),
        OtherASCone VisibleString
        Verified ::= BOOLEAN
}

Version ::= INTEGER
LastModified ::= GeneralizedTime
Created ::= GeneralizedTime
ContactEmail ::= PrintableString(SIZE (1..75))
END
```

ASN.1 format of an AS-Cone

3. Validating an AS-Cone

In order to validate a full AS-Cone, a network operator MUST have access to the validated cache of an RPKI validator software containing all the Policy definition and AS-Cone objects. Validation occurs following the description in [RFC6488]:

In order to validate a full AS-Cone, an operator SHOULD perform the following steps:

1. For every downstream ASN, the operator verifies if a related policy definition (see Section 2.1) file exists. If no object exists, the status of the AS-Cone is "Unknown". If instead it

exists, it proceeds to collect a list of ASNs for the cone by looking at the following data, in exact order:

1. A policy for the specific relationship, in the form of ASX:ASY, where ASX is the downstream ASN, and ASY is the ASN of the operator validating the AS-Cone;

2. If there is no specific definition for the relationship, the ASX:Default policy;

If none of the two definitions above exists, then the operator should only consider the ASN of its downstream to be added to the list.

2. These objects can either point to:

    1. An AS-Cone; or

    2. An ASN

3. If the definition points to an AS-Cone, the operator looks for the object referenced, which should be contained in the validated cache;

4. If the validated cache does not contain the referenced object, then the validation moves on to the next downstream ASN;

5. If the validated cache contains the referenced object, the validation process evaluates every entry in the AS-Cone. For each entry:

    1. If there is a reference to an ASN, then the operator adds the ASN to the list for the given AS-Cone;

    2. If there is a reference to another AS-Cone, the validating process should recursively process all the entries in that AS-Cone first, with the same principles contained in this list.

    Since the goal is to build a list of ASNs announcing routes in the AS-Cone, then if an ASN or an AS-Cone are referenced more than once in the process, their contents should only be added once to the list. This is intended to avoid endless loops, and in order to avoid cross-reference of AS-Cones.

6. When all the AS-Cones referenced in the policies have been recursively iterated, and all the originating ASNs have been taken into account, the operator can then build a full prefix-

list with all the prefixes originated in its AS-Cone. This can be done by querying the RPKI validator software for all the networks originated by every ASN referenced in the AS-Cone.

4. Types of validation for AS-Cones

AS-Cones can be validated in 4 different ways:

Loose Validation. This is the method described in the procedure above;

Opportunistic Validation. This is similar to Loose validation, but it discards all the ASNs for which the "Validated" fields have a value of 0. The intent is to remove from the prefix list all the ASNs that haven't validated their entry in the customer cone for the operator;

Almost-Strict validation. In this method, whenever an entry with the "Validated" field set to 0 is found, the entire sub-tree (the AS-Cone) in which it is contained is discarded.

Strict Validation. In this method, only the entries with the "Validated" field set to 1 are considered. If even a single entry has a "Validated" field set to 0, the whole AS-Cone is discarded.

It is important to note that no AS-Cone with the "Validated" field set to 0 is going to be visible at any time, so they are automatically discarded. This protects AS-Cone holders from being considered customers of a third party without their consent.

5. Recommendations for use of AS-Cones at Internet Exchange points

When an operator is a member of an internet exchange point, it is recommended for it to create at least a Default policy.

In case of a peering session with a route server, the operator could publish a policy pointing to the ASN of the route server. A route server operator, then, could build strict prefix filtering rules for all the participants, and offer it as a service to its members.

For internet exchange points operators, the recommendation is to use Strict Filtering as explained in the previous section.

6. Publication of AS-Cones as IRR objects

AS-Cones are very similar to AS-Set RPSL Objects, so they could also be published in IRR Databases as AS-Set objects. Every ASN contained in an AS-Cone, and all the AS-Cones referenced should be considered

as member: attributes. The naming convention for AS-Cones (ASX:AS-Cone) should be maintained, in order to keep consistency between the two databases.

7. Security Considerations

TBW

8. IANA Considerations

This memo includes no request to IANA.

9. Contributors

The following people contributed significantly to the content of the document: Greg Skinner.

10. Acknowledgments

The authors would like to thank Randy Bush, Nick Hilliard and Aftab Siddiqui.

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, <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>.

[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>.

11.2. Informative References

[RFC2622] Alaettinoglu, C., Villamizar, C., Gerich, E., Kessens, D., Meyer, D., Bates, T., Karrenberg, D., and M. Terpstra, "Routing Policy Specification Language (RPSL)", RFC 2622, DOI 10.17487/RFC2622, June 1999, <https://www.rfc-editor.org/info/rfc2622>.

[RFC6480] Lepinski, M. and S. Kent, "An Infrastructure to Support Secure Internet Routing", RFC 6480, DOI 10.17487/RFC6480, February 2012, <https://www.rfc-editor.org/info/rfc6480>.

[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>.

[RFC6488] Lepinski, M., Chi, A., and S. Kent, "Signed Object Template for the Resource Public Key Infrastructure (RPKI)", RFC 6488, DOI 10.17487/RFC6488, February 2012, <https://www.rfc-editor.org/info/rfc6488>.

Authors' Addresses

Job Snijders
NTT Ltd.
Theodorus Majofskistraat 100
Amsterdam 1065 SZ
The Netherlands

Email: job@ntt.net

Massimiliano Stucchi
Independent

Email: max@stucchi.ch

Melchior Aelmans
Juniper Networks
Boeing Avenue 240
Schiphol-Rijk 1119 PZ
The Netherlands

Email: maelmans@juniper.net

Network Working Group
Internet-Draft
Updates: 1997 (if approved)
Intended status: Standards Track
Expires: December 15, 2019

J. Borkenhagen
AT&T
R. Bush
IIJ & Arrcus
R. Bonica
Juniper Networks
S. Bayraktar
Cisco Systems
June 13, 2019

Well-Known Community Policy Behavior
draft-ietf-grow-wkc-behavior-08

Abstract

Well-Known BGP Communities are manipulated differently across various current implementations; resulting in difficulties for operators. Network operators should deploy consistent community handling across their networks while taking the inconsistent behaviors from the various BGP implementations into consideration.. This document recommends specific actions to limit future inconsistency, namely BGP implementors must not create further inconsistencies from this point forward.

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.

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 December 15, 2019.

Table of Contents

1. Introduction

The BGP Communities Attribute was specified in [RFC1997] which introduced the concept of Well-Known Communities. In hindsight, [RFC1997] did not prescribe as fully as it should have how Well-Known Communities may be manipulated by policies applied by operators. Currently, implementations differ in this regard, and these differences can result in inconsistent behaviors that operators find difficult to identify and resolve.

This document describes the current behavioral differences in order to assist operators in generating consistent community-manipulation policies in a multi-vendor environment, and to prevent the introduction of additional divergence in implementations.

This document recommends specific actions to limit future inconsistency, namely BGP implementors MUST NOT create further inconsistencies from this point forward.

2. Manipulation of Communities by Policy

[RFC1997] says:

"A BGP speaker receiving a route with the COMMUNITIES path attribute may modify this attribute according to the local policy."

One basic operational need is to add or remove one or more communities to the set. The focus of this document is another common operational need, to replace all communities with a new set. To simplify this second case, most BGP policy implementations provide syntax to "set" community that operators use to mean "remove any/all communities present on the route, and apply this set of communities instead."

Some operators prefer to write explicit policy to delete unwanted communities rather than using "set;" i.e. using a "delete community *:*" and then "add community x:y ..." configuration statements in an attempt to replace all communities. The same community manipulation policy differences described in the following section exist in both "set" and "delete community *:*" syntax. For simplicity, the remainder of this document refers only to the "set" behaviors, which we refer to collectively as each implementation's '"set" directive.'

3. Community Manipulation Policy Differences

Vendor implementations differ in the treatment of certain Well-Known communities when modified using the syntax to "set" the community. Some replace all communities including the Well-Known ones with the new set, while others replace all non-Well-Known Communities but do not modify any Well-Known Communities that are present.

These differences result in what would appear to be identical policy configurations having very different results on different platforms.

4. Documentation of Vendor Implementations

In this section we document the syntax and observed behavior of the "set" directive in several popular BGP implementations to illustrate the severity of the problem operators face.

In Juniper Networks' Junos OS, "community set" removes all communities, Well-Known or otherwise.

In Cisco IOS XR, "set community" removes all communities except for the following:

Numeric	Common Name
0:0	internet
65535:0	graceful-shutdown
65535:1	accept-own rfc7611
65535:65281	NO_EXPORT
65535:65282	NO_ADVERTISE
65535:65283	NO_EXPORT_SUBCONFED (or local-AS)

Communities not removed by Cisco IOS XR

Table 1

Cisco IOS XR does allow Well-Known communities to be removed only by explicitly enumerating one at a time, not in the aggregate; for example, "delete community accept-own". Operators are advised to consult Cisco IOS XR documentation and/or Cisco support for full details.

On Extreme networks' Brocade NetIron: "set community X" removes all communities and sets X.

In Huawei's VRP product, "community set" removes all communities, Well-Known or otherwise.

In OpenBGPD, "set community" does not remove any communities, Well-Known or otherwise.

Nokia's SR OS has several directives that operate on communities. Its "set" directive is called using the "replace" keyword, replacing all communities, Well-Known or otherwise, with the specified communities.

4.1. Note on an Inconsistency

The IANA publishes a list of Well-Known Communities [IANA-WKC].

Cisco IOS XR's set of Well-Known communities that "set community" will not overwrite diverges from the IANA's list of Well-Known communities. Quite a few Well-Known communities from IANA's list do not receive special treatment in Cisco IOS XR, and at least one community on Cisco IOS XR's special treatment list, internet == 0:0,

is not formally a Well-Known Community as it is not in [IANA-WKC]; but taken from the Reserved range [0x00000000-0x0000FFFF].

This merely notes an inconsistency. It is not a plea to 'protect' the entire IANA list from "set community."

5. Note for Those Writing RFCs for New Community-Like Attributes

When establishing new [RFC1997]-like attributes (large communities, wide communities, etc.), RFC authors should state explicitly how the new attribute is to be handled.

6. Action Items

Network operators are encouraged to limit their use of the "set" directive (within reason), to improve consistency across platforms.

Unfortunately, it would be operationally disruptive for vendors to change their current implementations.

Vendors MUST clearly document the behavior of "set" directive in their implementations.

Vendors MUST ensure that their implementations' "set" directive treatment of any specific community does not change if/when that community becomes a new Well-Known Community through future standardization. For most implementations, this means that the "set" directive MUST continue to remove the community; for those implementations where the "set" directive removes no communities, that behavior MUST continue.

Given the implementation inconsistencies described in this document, network operators are urged never to rely on any implicit understanding of a neighbor ASN's BGP community handling. I.e., before announcing prefixes with NO_EXPORT or any other community to a neighbor ASN, the operator should confirm with that neighbor how the community will be treated.

7. Security Considerations

Surprising defaults and/or undocumented behaviors are not good for security. This document attempts to remedy that.

8. IANA Considerations

The IANA is requested to list this document as an additional reference for the [IANA-WKC] registry.

9. Acknowledgments

The authors thank Martijn Schmidt, Qin Wu for the Huawei data point, Greg Hankins, Job Snijders, David Farmer, John Heasley, and Jakob Heitz.

10. Normative References

[IANA-WKC] IANA, "Border Gateway Protocol (BGP) Well-Known Communities", <https://www.iana.org/assignments/bgp-well-known-communities>.

[RFC1997] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, DOI 10.17487/RFC1997, August 1996, <http://www.rfc-editor.org/info/rfc1997>.

[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>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <http://www.rfc-editor.org/info/rfc8174>.

Authors' Addresses

Jay Borkenhagen
AT&T
200 Laurel Avenue South
Middletown, NJ 07748
United States of America

Email: jayb@att.com

Randy Bush
IIJ & Arrcus
5147 Crystal Springs
Bainbridge Island, WA 98110
US

Email: randy@psg.com

Ron Bonica
Juniper Networks
2251 Corporate Park Drive
Herndon, VA 20171
US

Email: rbonica@juniper.net

Serpil Bayraktar
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
United States of America

Email: serpil@cisco.com

Network Working Group
Internet-Draft
Updates: 7854 (if approved)
Intended status: Standards Track
Expires: March 24, 2019

J. Scudder
Juniper Networks

September 20, 2018

Revision to Registration Procedures for Multiple BMP Registries
draft-scudder-grow-bmp-registries-change-00.txt

Abstract

This document updates RFC 7854, BGP Monitoring Protocol (BMP) by making a change to the registration procedures for several registries. Specifically, any BMP registry with a range of 32768-65530 designated "Specification Required" has that range re-designated as "First Come First Served".

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 March 24, 2019.

Copyright Notice

Table of Contents

1. Introduction

[RFC7854] creates a number of IANA registries that include a range of 32768-65530 designated "Specification Required". Each such registry also has a large range designated "Standards Action". Subsequent experience has shown two things. First, there is less difference between these two policies in practice than there is in theory (consider that [RFC8126] explains that for Specification Required, "Publication of an RFC is an ideal means of achieving this requirement"). Second, it's desirable to have a very low bar to registration, to avoid the risk of conflicts introduced by use of unregistered code points (so-called "code point squatting").

Accordingly, this document revises the registration procedures, as given in Section 2.

2. IANA Considerations

IANA is requested to revise the following registries within the BMP group:

- BMP Statistics Types
- BMP Initiation Message TLVs
- BMP Termination Message TLVs
- BMP Termination Message Reason Codes
- BMP Peer Down Reason Codes
- BMP Route Mirroring TLVs
- BMP Route Mirroring Information Codes

For each of these registries, the ranges 32768-65530 whose registration procedures were "Specification Required" are revised to have the registration procedures "First Come First Served".

3. Security Considerations

This revision to registration procedures does not change the underlying security issues inherent in the existing [RFC7854].

4. Acknowledgements

Thanks to Jeff Haas for review and encouragement.

5. References

5.1. Normative References

[RFC7854] Scudder, J., Ed., Fernando, R., and S. Stuart, "BGP Monitoring Protocol (BMP)", RFC 7854, DOI 10.17487/RFC7854, June 2016, <https://www.rfc-editor.org/info/rfc7854>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

5.2. Informative References

[RFC1997] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, DOI 10.17487/RFC1997, August 1996, <https://www.rfc-editor.org/info/rfc1997>.

Author's Address

John Scudder
Juniper Networks
1194 N. Mathilda Ave
Sunnyvale, CA 94089
USA

Email: jgs@juniper.net