Global Routing Operations
Internet-Draft
Intended Status: Standards Track
Expires: September 11, 2017
March 10, 2017

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

Support for Local RIB in BGP Monitoring Protocol (BMP)

draft-evens-grow-bmp-local-rib-00

Abstract

The BGP Monitoring Protocol (BMP) defines access to the Adj-RIB-In and locally originated routes (e.g. routes distributed into BGP from protocols such as static) but not access to the BGP instance Loc-RIB. This document updates the BGP Monitoring Protocol (BMP) RFC 7854 by adding access to the BGP instance Local-RIB, as defined in RFC 4271 the routes that have been selected by the local BGP speaker's Decision Process. These are the routes over all peers, locally originated, and after best-path selection.

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), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on September 11, 2017.

Copyright and License Notice

Table of Contents

1. Introduction

The BGP Monitoring Protocol (BMP) suggests that locally originated routes are locally sourced routes, such as redistributed or otherwise added routes to the BGP instance by the local router. It does not specify routes that are in the BGP instance Loc-RIB, such as routes after best-path selection.

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

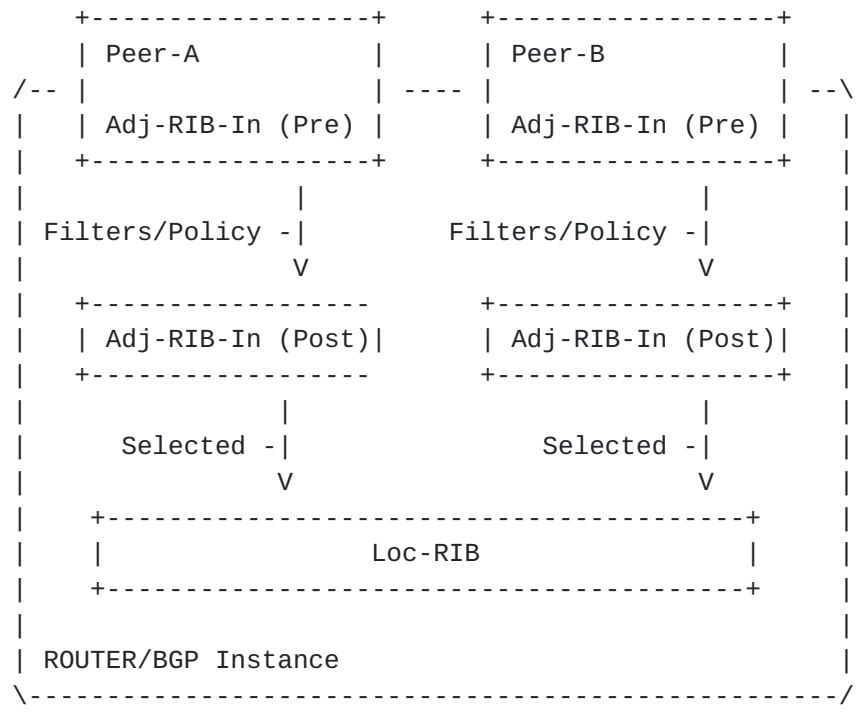


Figure 1: BGP peering Adj-RIBs-In into Loc-RIB

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

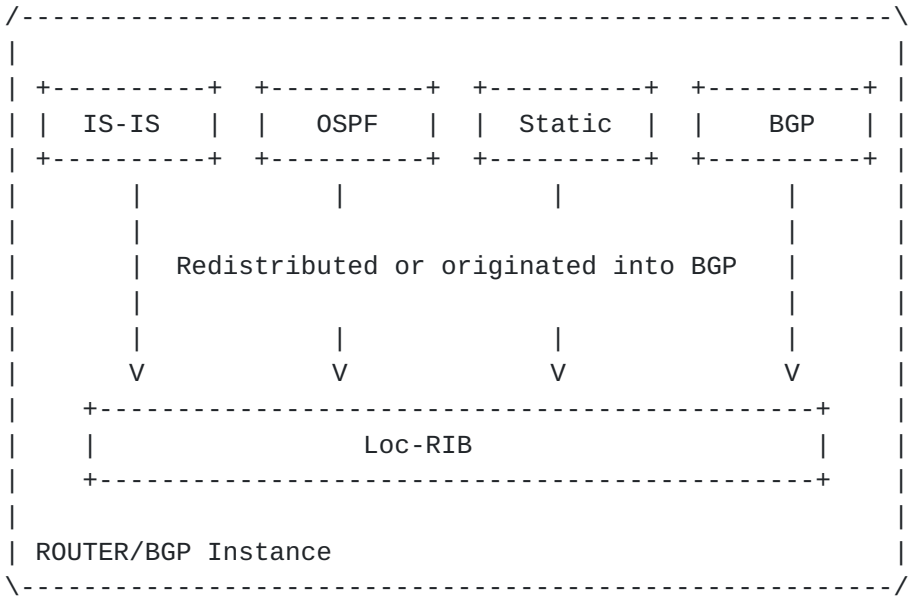


Figure 2: Locally Originated into Loc-RIB

BGP instance Loc-RIB usually provides a similar, if not exact, forwarding information base (FIB) view of the routes from BGP that the router will use. The following are some use-cases for Loc-RIB access:

- Adj-RIBs-In Post-Policy may still contain hundreds of thousands of routes per-peer but only a handful are selected and installed in the Loc-RIB as part of the 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 [RFC7854] Locally Originated Routes.

1.1. Current 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. While it is possible to derive the Loc-RIB, it is also error prone and complex.

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. 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.

Unnecessary resources needed for current method:

- 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, VRF's 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 erroneous 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-ID's). This leads to error prone correlations.

- The BGP-ID's 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-ID's 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", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Definitions

- **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 [RFC4271], "The Loc-RIB contains the routes that have been selected by the local BGP speaker's Decision Process." It is further defined that the routes selected include locally originated and routes from all peers.

- 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 be what is actually sent to the peer.

4. Per-Peer Header

4.1. Peer Type

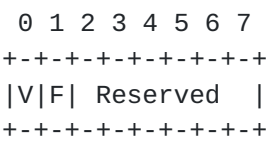
This document defines the following new peer type:

- Peer Type = 3: Loc-RIB Instance Peer

4.2. Peer Flags

In section 4.2 [RFC7854], the "locally sourced routes" comment in 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 V flag indicates that the Peer address is an IPv6 address. For IPv4 peers, this is set to 0.

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

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

Loc-RIB contains all routes from BGP peers as well as any and all routes redistributed or otherwise locally originated. In this context, only the BGP instance Loc-RIB is included. Routes from other routing protocols that have not been redistributed or received via Adj-RIB-In are not considered.

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 as remote peer address is not applicable.
- 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.

5.2. Peer UP Notification

Peer UP notifications follow section 4.10 [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 Name. The Information field contains an ASCII string whose value MUST be equal to the value of the VRF name (e.g. RD instance name) configured. This type is only relevant and used when the Loc-RIB represents a VRF/RD instance.

It is RECOMMENDED that the VRF Name be defined as "global" for the global/default Loc-RIB instance.

5.3. Peer Down Notification

Peer down notification SHOULD follow the section 4.9 [RFC7854] reason 2.

5.4. Route Monitoring

Route Monitoring messages are used for initial synchronization of the Loc-RIB. They are also used for incremental updates upon every change to the RIB. State compression on interval, such as 1 or greater seconds, can mask critical RIB changes. Therefore state compression SHOULD be avoided. If the Loc-RIB changes, a route monitor message should be sent.

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

5.5. Route Mirroring

Route mirroring is not applicable to Loc-RIB.

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 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 VRF's. The BMP receiver identifies the Loc-RIB's by the peer header distinguisher and BGP ID. The BMP receiver uses the VRF Name from the PEER UP to name 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 partial.

7. Security Considerations

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

8. IANA Considerations

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

o Peer Type = 3: Loc-RIB Instance Peer

9. References

9.1. URIs

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

9.2. Normative References

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

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

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

9.3. Informative References

[I-ID.ietf-grow-bmp-adj-rib-out] TBD.

Acknowledgments

TBD.

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

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

Email: manbhard@cisco.com

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

Email: paolo@ntt.net