

Global Routing Operations
Internet-Draft
Updates: [7854](#) (if approved)
Intended status: Standards Track
Expires: August 27, 2018

T. Evens
S. Bayraktar
M. Bhardwaj
Cisco Systems
P. Lucente
NTT Communications
February 23, 2018

Support for Local RIB in BGP Monitoring Protocol (BMP)
draft-ietf-grow-bmp-local-rib-01

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). 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 August 27, 2018.

Copyright Notice

Copyright (c) 2018 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
1.1.	Current Method to Monitor Loc-RIB	5
2.	Terminology	7
3.	Definitions	7
4.	Per-Peer Header	8
4.1.	Peer Type	8
4.2.	Peer Flags	8
5.	Loc-RIB Monitoring	9
5.1.	Per-Peer Header	9
5.2.	Peer UP Notification	9
5.2.1.	Peer UP Information	10
5.3.	Peer Down Notification	10
5.4.	Route Monitoring	10
5.4.1.	ASN Encoding	11
5.4.2.	Granularity	11
5.5.	Route Mirroring	11
5.6.	Statistics Report	11
6.	Other Considerations	11
6.1.	Loc-RIB Implementation	11
6.1.1.	Multiple Loc-RIB Peers	12
6.1.2.	Filtering Loc-RIB to BMP Receivers	12
7.	Security Considerations	12
8.	IANA Considerations	12
8.1.	BMP Peer Type	12
8.2.	BMP Peer Flags	13
8.3.	Peer UP Information TLV	13
9.	References	13
9.1.	Normative References	13
9.2.	URIs	13
	Acknowledgements	13
	Authors' Addresses	14

[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.

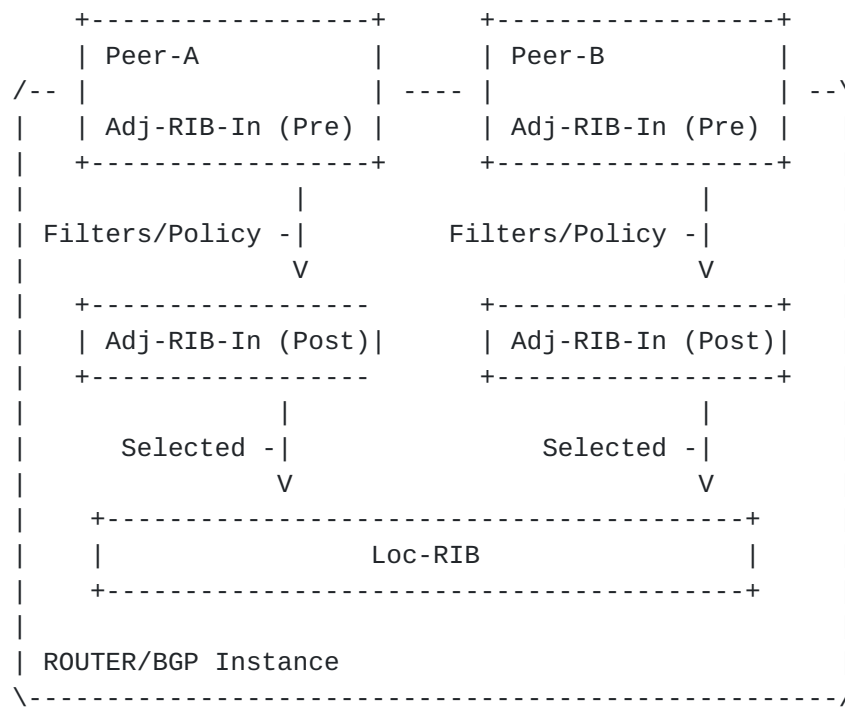


Figure 1: BGP peering Adj-RIBs-In into Loc-RIB

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.

- 0 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.

- o 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. At scale this becomes overly complex and error prone.

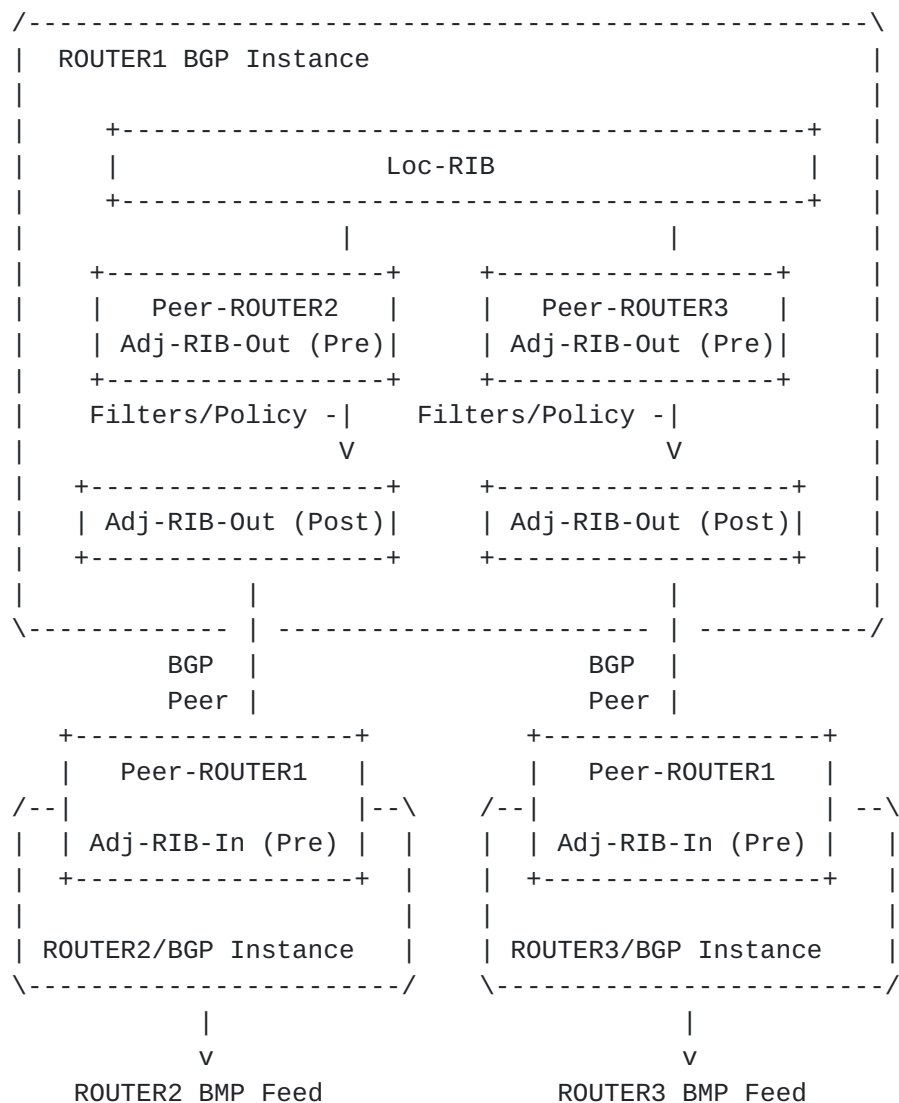


Figure 3: Current method to monitor Loc-RIB

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:

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

- o 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:

- o 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.
- o 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.
- o 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

- o 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.

- o 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."
- o 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.
- o 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.
- o 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 \[RFC7854\]](#) 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:

- o Peer Type = TBD: Loc-RIB Instance Peer

[4.2.](#) Peer Flags

In [section 4.2 \[RFC7854\]](#), 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:

```

      0 1 2 3 4 5 6 7
    +---+---+---+---+
    |F|  Reserved  |
    +---+---+---+---+

```

- o 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 SHOULD 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, originated by or into BGP, or received via Adj-RIB-In are not considered.

Loc-RIB in this context does not attempt to maintain a pre-policy and post-policy representation. Loc-RIB is the selected and used routes, which is equivalent to post-policy.

For example, VRF "Blue" imports several targets but filters out specific routes. The end result of VRF "Blue" Loc-RIB is conveyed. Even though the import is filtered, the result is complete for VRF "Blue" Loc-RIB. The F flag is not set in this case since the Loc-RIB is complete and not filtered to the BMP receiver.

5.1. Per-Peer Header

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

- o Peer Type: Set to TBD to indicate Loc-RIB Instance Peer.
- o 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.
- o Peer Address: Zero-filled. Remote peer address is not applicable. The V flag is not applicable with Local-RIB Instance peer type considering addresses are zero-filled.
- o Peer AS: Set to the BGP instance global or default ASN value.
- o 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:

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

- o Local Port: Set to 0, local port is not applicable.
- o Remote Port: Set to 0, remote port is not applicable.
- o 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.
- o 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 types are added:

- o Type = TBD: VRF/Table Name. The Information field contains an ASCII 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" SHOULD be used for the default Loc-RIB instance with a zero-filled distinguisher. If the TLV is included, then it SHOULD also be included in the Peer Down notification.

5.3. Peer Down Notification

Peer down notification SHOULD follow the [section 4.9 \[RFC7854\]](#) reason 2.

The VRF/Table Name informational TLV SHOULD 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 \[RFC7854\]](#), "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.

[5.6.](#) Statistics Report

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

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

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 parameters to the BMP parameters name space [[1](#)].

8.1. BMP Peer Type

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

- o Peer Type = TBD: 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:

- o 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](#)):

- o Type = TBD: VRF/Table Name. The Information field contains an ASCII 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.

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), DOI 10.17487/RFC4271, January 2006, <<https://www.rfc-editor.org/info/rfc4271>>.
- [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>>.

9.2. URIs

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

Acknowledgements

The authors would like to thank John Scudder for his valuable input.

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, WT 2132
NL

Email: paolo@ntt.net

