

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: June 2, 2012

J. Scudder
Juniper Networks
R. Fernando
Cisco Systems
S. Stuart
Google
November 30, 2011

BGP Monitoring Protocol
draft-ietf-grow-bmp-06

Abstract

This document proposes a simple protocol, BMP, which can be used to monitor BGP sessions. BMP is intended to provide a more convenient interface for obtaining route views for research purpose than the screen-scraping approach in common use today. The design goals are to keep BMP simple, useful, easily implemented, and minimally service-affecting. BMP is not suitable for use as a routing protocol.

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 <http://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 June 2, 2012.

Copyright Notice

Copyright (c) 2011 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 (<http://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.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1.	Introduction	4
1.1.	Requirements Language	5
2.	Lifecycle of a BMP Session	5
3.	BMP Message Format	5
3.1.	Common Header	5
3.2.	Per-Peer Header	6
3.3.	Initiation Message	8
3.4.	Route Monitoring	9
3.5.	Stats Reports	9
3.6.	Peer Down Notification	10
3.7.	Peer Up Notification	11
4.	Route Monitoring	12
5.	Stat Reports	13
6.	Other Considerations	13
7.	Using BMP	14
8.	IANA Considerations	14
9.	Security Considerations	15
10.	Acknowledgements	15
11.	References	15
11.1.	Normative References	15
11.2.	Informative References	16
Appendix A.	Changes Between BMP Versions 1 and 2	16
Appendix B.	Changes Between BMP Versions 2 and 3	16
	Authors' Addresses	16

1. Introduction

Many researchers wish to have access to the contents of routers' BGP RIBs as well as a view of protocol updates that the router is receiving. This monitoring task cannot be realized by standard protocol mechanisms. At present, this data can only be obtained through screen-scraping.

The BMP protocol provides access to the Adj-RIB-In of a peer on an ongoing basis and a periodic dump of certain statistics that the monitoring station can use for further analysis. The following are the messages provided by BMP.

- o Route Monitoring (RM): An initial dump of all routes received from a peer as well as an ongoing mechanism that sends the incremental routes advertised and withdrawn by a peer to the monitoring station.
- o Peer Down Notification (PD): A message sent to indicate that a peering session has gone down with information indicating the reason for the session disconnect.
- o Stats Reports (SR): This is an ongoing dump of statistics that can be used by the monitoring station as a high level indication of the activity going on in the router.
- o Peer Up Notification (PU): A message sent to indicate that a peering session has come up. The message includes information regarding the data exchanged between the peers in their OPEN messages as well as information about the peering TCP session itself. In addition to being sent whenever a peer transitions to ESTABLISHED state, a Peer Up Notification is sent for each peer that is in ESTABLISHED state when the BMP session itself comes up.

BMP operates over TCP. All options are controlled by configuration on the monitored router. No message is ever sent from the monitoring station to the monitored router. The monitored router MAY take steps to prevent the monitoring station from sending data (e.g. by half-closing the TCP session or setting its window size to zero) or it MAY silently discard any data erroneously sent by the monitoring station.

The monitoring station is configured to listen on a particular TCP port and the router is configured to establish an active connection to that port and to send messages on that TCP connection. There is no initialization or handshaking phase, messages are simply sent as soon as the connection is established. If the router is unable to connect to the monitoring station, it periodically retries the connection. A suggested default retry period is 30 seconds.

If the monitoring station intends to restart BMP processing, it simply drops the connection. The router then re-establishes the connection and resends the messages.

1.1. 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](#)].

2. Lifecycle of a BMP Session

A BMP session begins when a router running BMP successfully opens a TCP session (the "BMP session") to the monitoring station it is configured to talk to. It MAY first send an Initiation message. It subsequently sends a Peer Up message over the BMP session for each of its BGP peers which are in Established state. It follows by sending the contents of its Adj-RIBs-In (or Loc-RIB, see [Section 4](#)) encapsulated in Route Monitoring messages. Once it has sent all the routes for a given peer, it sends an End-of-RIB message for that peer; when End-of-RIB has been sent for each peer, the initial table dump has completed. (A monitoring station that wishes only to gather a table dump could close the connection once it has gathered an End-of-RIB or Peer Down message corresponding to each Peer Up message.)

Following the initial table dump, the router sends incremental updates encapsulated in Route Monitoring messages. It MAY periodically send Stats Reports or even new Initiation messages, according to configuration. If any new BGP peers become Established, corresponding Peer Up messages are sent. If any BGP peers for which Peer Up messages were sent transition out of the Established state, corresponding Peer Down messages are sent.

A BMP session ends when the TCP session that carries it is closed for any reason.

3. BMP Message Format

3.1. Common Header

The following common header appears in all BMP messages. The rest of the data in a BMP message is dependent on the "Message Type" field in the common header.


```

0 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8
+---+---+---+---+---+---+---+
|   Version   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Message Length                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   Msg. Type   |
+-----+

```

- o Version (1 byte): Indicates the BMP version. This is set to '3' for all messages defined in this specification. Version 0 is reserved and MUST NOT be sent.
- o Message Length (4 bytes): Length of the message in bytes (including headers, data and encapsulated messages, if any).
- o Message Type (1 byte): This identifies the type of the BMP message. A BMP implementation MUST ignore unrecognized message types upon receipt.
 - * Type = 0: Route Monitoring
 - * Type = 1: Statistics Report
 - * Type = 2: Peer Down Notification
 - * Type = 3: Peer Up Notification
 - * Type = 4: Initiation Message

3.2. Per-Peer Header

The per-peer header follows the common header for most BMP messages. The rest of the data in a BMP message is dependent on the "Message Type" field in the common header.


```

0 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Peer Type | Peer Flags |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Peer Distinguisher (present based on peer type) |
|
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Peer Address (16 bytes) |
~
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Peer AS |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Peer BGP ID |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Timestamp (seconds) |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Timestamp (microseconds) |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

- o Peer Type (1 byte): These bits identify the type of the peer. Currently only two types of peers are identified,
 - * Peer Type = 0: Global Instance Peer
 - * Peer Type = 1: L3 VPN Instance Peer
- o Peer Flags (1 byte): These flags provide more information about the peer. The flags are defined as follows.

```

0 1 2 3 4 5 6 7 8
+---+---+---+---+---+---+---+---+
|V|L| Reserved |
+---+---+---+---+---+---+---+---+

```

- * The V flag indicates the the Peer address is an IPv6 address. For IPv4 peers this is set to 0.
 - * The L flag, if set to 1, indicates that the message reflects the Loc-RIB (i.e., it reflects the application of inbound policy). It is set to 0 if the message reflects the Adj-RIB-In. See [Section 4](#) for further detail.
 - * The remaining bits are reserved for future use.
- o Peer Distinguisher (8 bytes): Routers today can have multiple instances (example L3VPNs). This field is present to distinguish peers that belong to one address domain from the other.

If the peer is a "Global Instance Peer", this field is zero

filled. If the peer is a "L3VPN Instance Peer", it is set to the route distinguisher of the particular L3VPN instance that the peer belongs to.

- o Peer Address: The remote IP address associated with the TCP session over which the encapsulated PDU was received. It is 4 bytes long if an IPv4 address is carried in this field (with most significant bytes zero filled) and 16 bytes long if an IPv6 address is carried in this field.
- o Peer AS: The Autonomous System number of the peer from which the encapsulated PDU was received. If a 16 bit AS number is stored in this field [[RFC4893](#)], it should be padded with zeroes in the most significant bits.
- o Peer BGP ID: The BGP Identifier of the peer from which the encapsulated PDU was received.
- o Timestamp: The time when the encapsulated routes were received (one may also think of this as the time when they were installed in the Adj-RIB-In), 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.

3.3. Initiation Message

The initiation message provides a means for the monitored router to inform the monitoring station of its vendor, software version, and so on. The initiation message is OPTIONAL. When used, an initiation message MUST be sent as the first message after the TCP session comes up. An initiation message MAY be sent at any point thereafter, if warranted by a change on the monitored router.

The initiation message consists of the common BMP header followed by one or more TLVs containing information about the monitored router, as follows:

```

0 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Information Type           |           Information Length           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Information (variable)           |
~                                           ~
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```


- o Information Type (2 bytes): Type of information provided. Defined types are:
 - * Type = 0: String. The Information field contains a free-form UTF-8 string whose length is given by the "Information Length" field.
- o Information Length (2 bytes): The length of the following Information field, in bytes.
- o Information (variable): Information about the monitored router, according to the type.

3.4. Route Monitoring

Route Monitoring messages are used for initial synchronization of ADJ-RIB-In. They are also used for ongoing monitoring of received advertisements and withdraws. This is discussed in more detail in subsequent sections.

Following the common BMP header and per-peer header is a BGP PDU.

3.5. Stats Reports

These messages contain information that could be used by the monitoring station to observe interesting events that occur on the router. 'Stats Report' messages have a message type of '3'.

The transmission of the SR messages could be timer triggered or event driven (for example, when a significant event occurs or a threshold is reached). This specification does not impose any timing restrictions on when and on what event these reports have to be transmitted. It is left to the implementation to determine transmission timings -- however, configuration control should be provided of the timer and/or threshold values. This document only specifies the form and content of SR messages.

Following the common BMP header and per-peer header is a 4-byte field that indicates the number of counters in the stats message where each counter is encoded as a TLV.

```

0 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|                                     Stats Count                             |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+

```

Each counter is encoded as follows,


```

0 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Stat Type           |           Stat Len           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                               Stat Data                               |
~                                                                       ~
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

- o Stat Type (2 bytes): Defines the type of the statistic carried in the "Stat Data" field.
- o Stat Len (2 bytes): Defines the length of the "Stat Data" Field.

This specification defines the following statistics. All statistics are 4-byte quantities and the stats data are counters. A BMP implementation **MUST** ignore unrecognized stat types on receipt, and likewise **MUST** ignore unexpected data in the Stat Data field.

- o Stat Type = 0: Number of prefixes rejected by inbound policy.
- o Stat Type = 1: Number of (known) duplicate prefix advertisements.
- o Stat Type = 2: Number of (known) duplicate withdraws.
- o Stat Type = 3: Number of updates invalidated due to CLUSTER_LIST loop.
- o Stat Type = 4: Number of updates invalidated due to AS_PATH loop.
- o Stat Type = 5: Number of updates invalidated due to ORIGINATOR_ID.
- o Stat Type = 6: Number of updates invalidated due to AS_CONFED loop.

Note that the current specification only specifies 4-byte counters as "Stat Data". This does not preclude future versions from incorporating more complex TLV-type "Stat Data" (for example, one which can carry prefix specific data). SR messages are optional. However if an SR message is transmitted, this specification requires at least one statistic to be carried in it.

3.6. Peer Down Notification

This message is used to indicate that a peering session was terminated. The type of this message is 4.


```

0 1 2 3 4 5 6 7 8
+--+--+--+--+--+--+--+
|      Reason      | 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|      Data (present if Reason = 1, 2 or 3)      |
~
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+

```

Reason indicates why the session was closed. Defined values are:

- 0 Reason 1: The local system closed the session. Following the Reason is a BGP PDU containing a BGP NOTIFICATION message that would have been sent to the peer.
- 0 Reason 2: The local system closed the session. No notification message was sent. Following the reason code is a two-byte field containing the code corresponding to the FSM Event which caused the system to close the session (see [Section 8.1 of \[RFC4271\]](#)). Zero is used to indicate that no relevant Event code is defined.
- 0 Reason 3: The remote system closed the session with a notification message. Following the Reason is a BGP PDU containing the BGP NOTIFICATION message as received from the peer.
- 0 Reason 4: The remote system closed the session without a notification message.

3.7. Peer Up Notification

The Peer Up message is used to indicate that a peering session has come up (i.e., has transitioned into ESTABLISHED state). Following the common BMP header and per-peer header is the following:

[illegible]

- o Local Address: The local IP address associated with the peering TCP session. It is 4 bytes long if an IPv4 address is carried in this field, as determined by the V flag (with most significant bytes zero filled) and 16 bytes long if an IPv6 address is carried in this field.
- o Local Port: The local port number associated with the peering TCP session.
- o Remote Port: The remote port number associated with the peering TCP session. (Note that the remote address can be found in the Peer Address field of the fixed header.)
- o Sent OPEN Message: The full OPEN message transmitted by the monitored router to its peer.
- o Received OPEN Message: The full OPEN message received by the monitored router from its peer.

4. Route Monitoring

After the BMP session is up, Route Monitoring messages are used to provide a snapshot of the Adj-RIB-In of a particular peer. This is done by sending all routes stored in the Adj-RIB-In of that peer using standard BGP Update messages. There is no requirement on the ordering of messages in the peer dump. When the initial peer dump is completed, this MUST be indicated by sending an End-of-RIB marker (as specified in [Section 2 of \[RFC4724\]](#), plus the BMP encapsulation header).

Depending on the implementation or configuration, it may only be possible to send the Loc-RIB (post-policy routes) instead of the Adj-RIB-In. This is because it is possible that a BGP implementation may not store, for example, routes which have been filtered out by policy. If this is the case, the implementation may send the Loc-RIB path that pertains to a particular peer in the route monitor message. Such paths MUST have the L flag set in the BMP header (see [Section 3](#)).

If the implementation is able to provide information about when routes were received, it MAY provide such information in the BMP timestamp field. Otherwise, the BMP timestamp field MUST be set to zero, indicating that time is not available.

AS Numbers in the BMP UPDATE message MUST be sent as 4-octet quantities, as described in [\[RFC4893\]](#). This affects the AS_PATH and AGGREGATOR path attributes. AS4_PATH or AS4_AGGREGATOR path

attributes MUST NOT be sent in a BMP UPDATE message, as it makes no sense to do so.

Ongoing monitoring is accomplished by propagating route changes in BGP UPDATE PDUs and forwarding those PDUs to the monitoring station, again using RM messages. When a change occurs to a route, such as an attribute change, the router must update the monitor with the new attribute. When a route is withdrawn by a peer, a corresponding withdraw is sent to the monitor. Multiple changed routes MAY be grouped into a single BGP UPDATE PDU when feasible, exactly as in the standard BGP protocol.

It's important to note that RM messages are not real time replicated messages received from a peer. While the router should attempt to generate updates as soon as they are received there is a finite time that could elapse between reception of an update and the generation an RM message and its transmission to the monitoring station. If there are state changes in the interim for that prefix, it is acceptable that the router generate the final state of that prefix to the monitoring station. The actual PDU generated and transmitted to the station might also differ from the exact PDU received from the peer, for example due to differences between how different implementations format path attributes.

5. Stat Reports

As outlined above, SR messages are used to monitor specific events and counters on the monitored router. One type of monitoring could be to find out if there are an undue number of route advertisements and withdraws happening (churn) on the monitored router. Another metric is to evaluate the number of looped AS-Paths on the router.

While this document proposes a small set of counters to begin with, the authors envision this list may grow in the future with new applications that require BMP style monitoring.

6. Other Considerations

Some routers may support multiple instances of the BGP protocol, for example as "logical routers" or through some other facility. The BMP protocol relates to a single instance of BGP; thus, if a router supports multiple BGP instances it should also support multiple BMP instances (one per BMP instance).

7. Using BMP

Once the BMP session is established route monitoring starts dumping the current snapshot as well as incremental changes simultaneously.

It is fine to have these operations occur concurrently. If the initial dump visits a route and subsequently a withdraw is received, this will be forwarded to the monitoring station which would have to correlate and reflect the deletion of that route in its internal state. This is an operation a monitoring station would need to support regardless.

If the router receives a withdraw for a prefix even before the peer dump procedure visits that prefix, then the router would clean up that route from its internal state and will not forward it to the monitoring station. In this case, the monitoring station may receive a bogus withdraw which it can safely ignore.

8. IANA Considerations

This document defines five message types for transferring BGP messages between cooperating systems ([Section 3](#)):

- o Type 0: Route Monitor
- o Type 1: Statistics Report
- o Type 2: Peer Down Notification
- o Type 3: Peer Up Notification
- o Type 4: Initiation

Type values 5 through 128 MUST be assigned using the "Standards Action" policy, and values 129 through 255 using the "Specification Required" policy defined in [[RFC5226](#)].

This document defines five statistics types for statistics reporting ([Section 3.5](#)):

- o Stat Type = 0: Number of prefixes rejected by inbound policy.
- o Stat Type = 1: Number of (known) duplicate prefix advertisements.
- o Stat Type = 2: Number of (known) duplicate withdraws.
- o Stat Type = 3: Number of updates invalidated due to CLUSTER_LIST loop.
- o Stat Type = 4: Number of updates invalidated due to AS_PATH loop.
- o Stat Type = 5: Number of updates invalidated due to ORIGINATOR_ID.
- o Stat Type = 6: Number of updates invalidated due to AS_CONFED loop.

Stat Type values 7 through 32767 MUST be assigned using the

"Standards Action" policy, and values 32768 through 65535 using the "Specification Required" policy, defined in [[RFC5226](#)].

This document defines one type for information carried in the Initiation message ([Section 3.3](#)):

- o Type = 0: String.

Information type values 1 through 32767 MUST be assigned using the "Standards Action" policy, and values 32768 through 65535 using the "Specification Required" policy, defined in [[RFC5226](#)].

[9.](#) Security Considerations

This document defines a mechanism to obtain a full dump or provide continuous monitoring of a BGP speaker's local BGP table, including received BGP messages. This capability could allow an outside party to obtain information not otherwise obtainable.

Implementations of this protocol MUST require manual configuration of the monitored and monitoring devices.

Users of this protocol MAY use some type of secure transmission mechanism, such as IPSec [[RFC4303](#)], to transmit this data.

[10.](#) Acknowledgements

Thanks to John ji Ioannidis, Mack McBride, Danny McPherson, Dimitri Papadimitriou, Erik Romijn, and the members of the GROW working group for their comments.

[11.](#) References

[11.1.](#) Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", [RFC 4271](#), January 2006.
- [RFC4724] Sangli, S., Chen, E., Fernando, R., Scudder, J., and Y. Rekhter, "Graceful Restart Mechanism for BGP", [RFC 4724](#), January 2007.

- [RFC4893] Vohra, Q. and E. Chen, "BGP Support for Four-octet AS Number Space", [RFC 4893](#), May 2007.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", [BCP 26](#), [RFC 5226](#), May 2008.

11.2. Informative References

- [RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)", [RFC 4303](#), December 2005.

Appendix A. Changes Between BMP Versions 1 and 2

- o Added Peer Up Message
- o Added L flag
- o Editorial changes

Appendix B. Changes Between BMP Versions 2 and 3

- o Added a 16-bit length field to the fixed header.
- o Clarified error handling.
- o Added stat types 5 and 6 (number of updates invalidated due to ORIGINATOR_ID and AS_CONFED, respectively).
- o For peer down messages, the relevant FSM event is to be sent in type 2 messages.
- o Added local address and local and remote ports to the peer up message.
- o Require End-of-RIB marker after initial dump.
- o Added Initiation message with string content.
- o Changed assignment policy for IANA registries.
- o Editorial changes.

Authors' Addresses

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

Email: jgs@juniper.net

Rex Fernando
Cisco Systems
170 W. Tasman Dr.
San Jose, CA 95134
USA

Email: rex@cisco.com

Stephen Stuart
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
USA

Email: [sstuart@google.com](mailto:ssstuart@google.com)

