Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: November 23, 2015

J. Scudder
Juniper Networks
R. Fernando
Cisco Systems
S. Stuart
Google

May 22, 2015

BGP Monitoring Protocol
draft-ietf-grow-bmp-08

Abstract

This document defines a 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 November 23, 2015.

Copyright Notice

Table of Contents

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. Prior to introduction of BMP, this data could 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. From a high level, BMP can be thought of as the result of multiplexing together the messages received on the various monitored BGP sessions.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. 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.

- Post-Policy Adj-RIB-In: The result of applying inbound policy to an Adj-RIB-In, but prior to the application of route selection to form the Loc-RIB.

3. Overview of BMP Operation

3.1. BMP Messages

The following are the messages provided by BMP.

- 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.

- 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.

- Stats Reports (SR): 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.

- 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.

- Initiation: A means for the monitored router to inform the monitoring station of its vendor, software version, and so on.

- Termination: A means for the monitored router to inform the monitoring station of why it is closing a BMP session.

- Route Mirroring: a means for the monitored router to send verbatim duplicates of messages as received. Can be used to exactly mirror a monitored BGP session. Can also be used to report malformed BGP PDUs.

3.2. Connection Establishment and Termination

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 (for example by half-closing the TCP session or setting its window size to zero) or it MAY silently discard any data sent by the monitoring station.

The router may be monitored by one or more monitoring stations. With respect to each (router, monitoring station) pair, one party is active with respect to TCP session establishment, and the other party is passive. Which party is active and which is passive is controlled by configuration.

The passive party is configured to listen on a particular TCP port and the active party is configured to establish a connection to that port. If the active party is unable to connect to the passive party, it periodically retries the connection. Retries MUST be subject to some variety of backoff. Exponential backoff with a default initial backoff of 30 seconds and a maximum of 720 seconds is suggested.

The router MAY restrict the set of IP addresses from which it will accept connections. It SHOULD restrict the number of simultaneous connections it will permit from a given IP address. The default value for this restriction SHOULD be 1, though an implementation MAY permit this restriction to be disabled in configuration. The router MUST also restrict the rate at which sessions may be established. A suggested default is an establishment rate of 2 sessions per minute.

A router (or management station) MAY implement logic to detect

redundant connections, as might occur if both parties are configured to be active, and MAY elect to terminate redundant connections. A Termination reason code is defined for this purpose.

Once a connection is established, the router sends messages over it. There is no initialization or handshaking phase, messages are simply sent as soon as the connection is established.

If the monitoring station intends to restart BMP processing, it simply drops the connection, optionally with a Termination message.

3.3. Lifecycle of a BMP Session

A router is configured to speak BMP with one or more monitoring stations. It MAY be configured to send monitoring information for only a subset of its BGP peers. Otherwise, all BGP peers are assumed to be monitored.

A BMP session begins when the active party (either router or management station, as determined by configuration) successfully opens a TCP session (the "BMP session"). Once the session is up, the router begins to send BMP messages. It MUST begin by sending an Initiation message. It subsequently sends a Peer Up message over the BMP session for each of its monitored BGP peers which are in Established state. It follows by sending the contents of its Adj-RIBs-In (pre-policy, post-policy or both, see Section 5) encapsulated

in Route Monitoring messages. Once it has sent all the routes for a given peer, it MUST send a End-of-RIB message for that peer; when End-of-RIB has been sent for each monitored 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.)

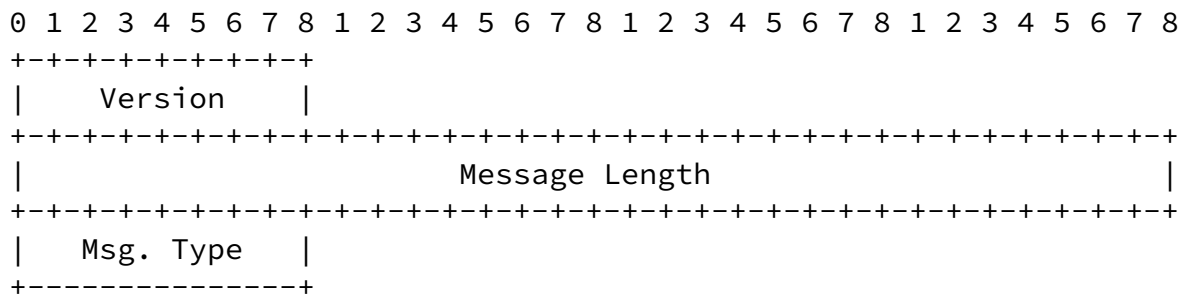
Following the initial table dump, the router sends incremental updates encapsulated in Route Monitoring messages. It MAY periodically send Stats Reports or even new Initiation messages, according to configuration. If any new monitored BGP peers become Established, corresponding Peer Up messages are sent. 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. The router MAY send a Termination message prior to closing the session.

4. BMP Message Format

4.1. 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.



- 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.

- Message Length (4 bytes): Length of the message in bytes (including headers, data and encapsulated messages, if any).

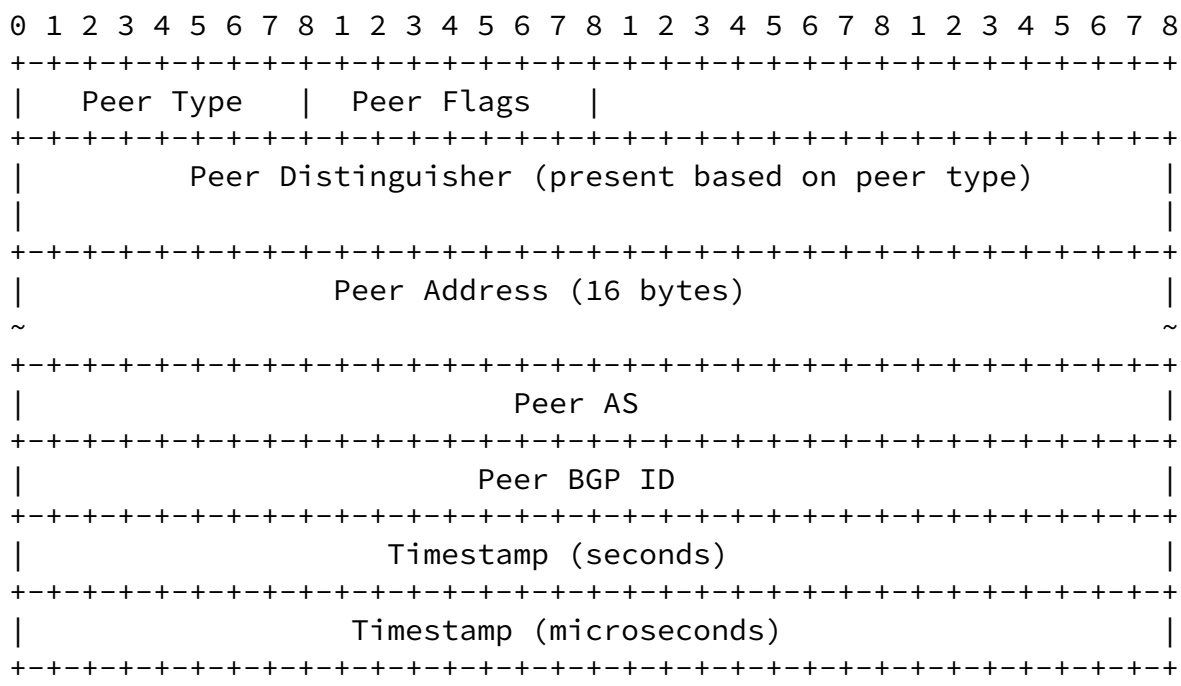
- Message Type (1 byte): This identifies the type of the BMP message. A BMP implementation MUST ignore unrecognized message types upon receipt.

  * Type = 0: Route Monitoring

  * Type = 1: Statistics Report
  * Type = 2: Peer Down Notification
  * Type = 3: Peer Up Notification
  * Type = 4: Initiation Message
  * Type = 5: Termination Message

4.2. Per-Peer Header

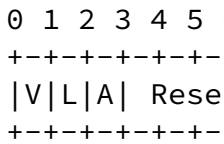
The per-peer header follows the common header for most BMP messages. The rest of the data in a BMP message is dependent on the "Message Type" field in the common header.



- 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

- Peer Flags (1 byte): These flags provide more information about the peer. The flags are defined as follows.



  * 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 post-policy Adj-RIB-In (i.e., its path attributes reflect the application of inbound policy). It is set to 0 if the message reflects the pre-policy Adj-RIB-In. Locally-sourced routes also carry an L flag of 1. See Section 5 for further detail. This flag has no significance when used with route mirroring messages (Section 4.7).
* The A flag, if set to 1, indicates that the message is formatted using the legacy two-byte AS_PATH format. If set to 0, the message is formatted using the four-byte AS_PATH format [RFC6793]. A BMP speaker MAY choose to propagate the AS_PATH information as received from its peer, or it MAY choose to reformat all AS_PATH information into four-byte format regardless of how it was received from the peer. In the latter case, AS4_PATH or AS4_AGGREGATOR path attributes SHOULD NOT be sent in the BMP UPDATE message. This flag has no significance when used with route mirroring messages (Section 4.7).
* The remaining bits are reserved for future use.

- 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.

- 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.

- 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 [RFC6793], it should be padded with zeroes in the most significant bits.

- Peer BGP ID: The BGP Identifier of the peer from which the encapsulated PDU was received.

- 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.

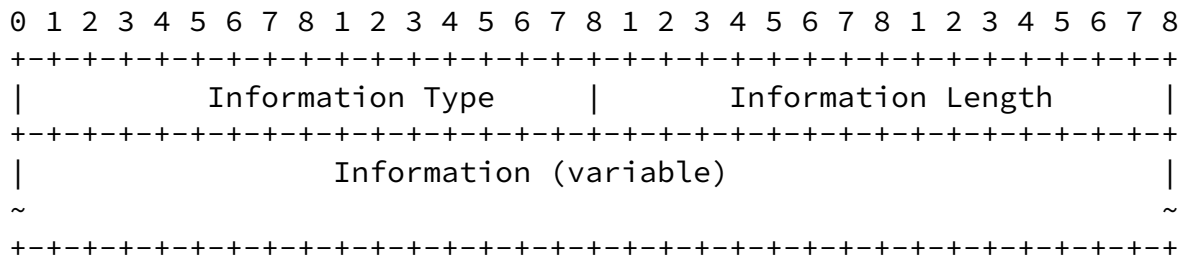
4.3. Initiation Message

The initiation message provides a means for the monitored router to inform the monitoring station of its vendor, software version, and so on. 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 two or more Information TLVs (Section 4.4) containing information about the monitored router. The sysDescr and sysName Information TLVs MUST be sent, any others are optional. The string TLV MAY be included multiple times.

4.4. Information TLV

The Information TLV is used by the Initiation (Section 4.3) and Peer Up (Section 4.10) messages.



- 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. The value is administratively assigned. Note that there is no requirement to terminate the string with a null (or any other particular) character -- the length field gives its termination. If multiple strings are included, their ordering MUST be preserved when they are reported.

  * Type = 1: sysDescr. The Information field contains an ASCII string whose value MUST be set to be equal to the value of the sysDescr MIB-II [RFC1213] object.

  * Type = 2: sysName. The Information field contains a ASCII

string whose value MUST be set to be equal to the value of the sysName MIB-II [RFC1213] object.

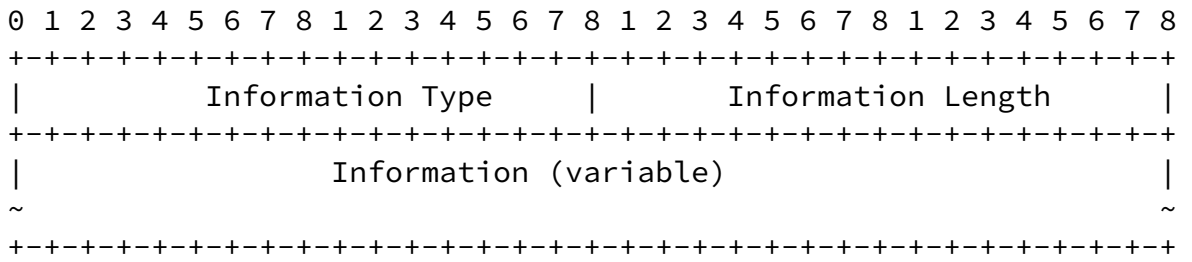
- Information Length (2 bytes): The length of the following Information field, in bytes.

- Information (variable): Information about the monitored router, according to the type.

4.5. Termination Message

The termination message provides a way for a monitored router to indicate why it is terminating a session. Although use of this message is RECOMMENDED, a monitoring station must always be prepared for the session to terminate with no message. Once the router has sent a termination message, it MUST close the TCP session without sending any further messages. Likewise, the monitoring station MUST close the TCP session after receiving a termination message.

The termination message consists of the common BMP header followed by one or more TLVs containing information about the reason for the termination, as follows:



- 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. Inclusion of this TLV is optional. It MAY be used to provide further detail for any of the defined reasons. Multiple String TLVs MAY be included in the message.

  * Type = 1: Reason. The Information field contains a two-byte code indicating the reason the connection was terminated. Some

reasons may have further TLVs associated with them. Inclusion of this TLV is REQUIRED. Defined reasons are:

+ Reason = 0: Session administratively closed. The session might be re-initiated.

+ Reason = 1: Unspecified reason.

+ Reason = 2: Out of resources. The router has exhausted resources available for the BMP session.

+ Reason = 3: Redundant connection. The router has determined that this connection is redundant with another one.

+ Reason = 4: Session permanently administratively closed, will not be re-initiated. Collector should reduce (potentially to zero) the rate at which it attempts reconnection to the monitored router.

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.

4.6. Route Monitoring

Route Monitoring messages are used for initial synchronization of ADJ-RIBs-In. They are also used for ongoing monitoring of received advertisements and withdraws. Route monitoring messages are state-compressed. This is all discussed in more detail in Section 5.

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

4.7. Route Mirroring

Route Mirroring messages are used for verbatim duplication of messages as received. A possible use for mirroring is exact mirroring of one or more monitored BGP sessions, without state

compression. Another possible use is mirroring of messages that have been treated-as-withdraw [I-D.ietf-idr-error-handling], for debugging purposes. Mirrored messages may be sampled, or may provide complete fidelity. The Messages Lost Information code is provided to allow this to be communicated. Section 6 provides more detail.

Following the common BMP header and per-peer header is a set of TLVs that contain information about a message or set of messages. Each TLV comprises a two-byte type code, a two-byte length field, and a variable-length value. Inclusion of any given TLV is OPTIONAL, however at least one TLV SHOULD be included, otherwise what's the point of sending the message? Defined TLVs are as follows:

- Type = 0: BGP Message. A BGP PDU. This PDU may or may not be an Update message. If the BGP Message TLV occurs in the Route Mirroring message, it MUST occur last in the list of TLVs.

- Type = 1: Information. A two-byte code that provides information about the mirrored message or message stream. Defined codes are:

  * Code = 0: Errored PDU. The contained message was found to have some error that made it unusable, causing it to be treated-as-withdraw [I-D.ietf-idr-error-handling]. A BGP Message TLV MUST also occur in the TLV list.

  * Code = 1: Messages Lost. One or more messages may have been lost. This could occur, for example, if an implementation runs out of available buffer space to queue mirroring messages.

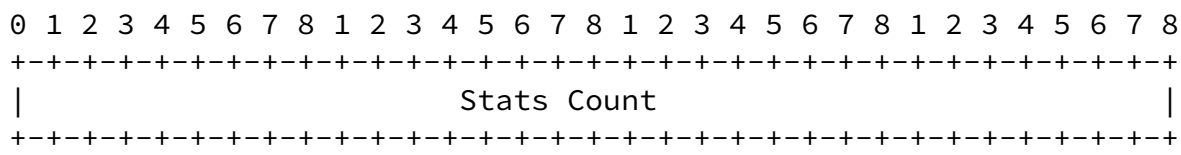
4.8. Stats Reports

These messages contain information that could be used by the monitoring station to observe interesting events that occur on the router.

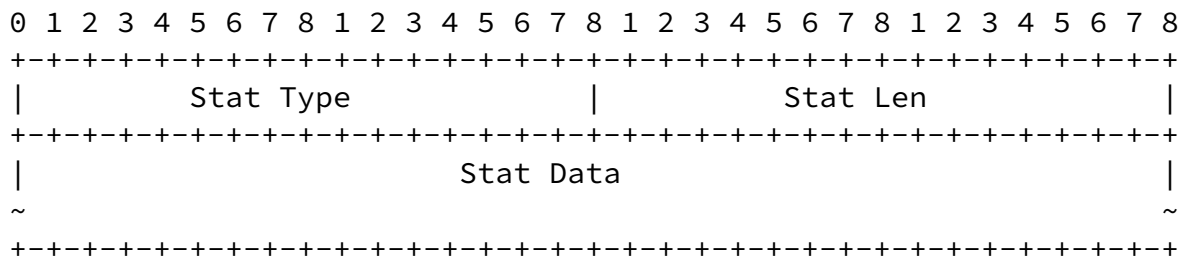
Transmission of SR messages could be timer triggered or event driven (for example, when a significant event occurs or a threshold is reached). This specification does not impose any timing restrictions on when and on what event these reports have to be transmitted. It is left to the implementation to 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.



Each counter is encoded as follows,



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

- Stat Len (2 bytes): Defines the length of the "Stat Data" Field.

This specification defines the following statistics. A BMP implementation MUST ignore unrecognized stat types on receipt, and likewise MUST ignore unexpected data in the Stat Data field.

Stats are either counters or gauges, defined as follows after the examples of [RFC1155] Section 3.2.3.3 and [RFC2856] Section 4 respectively:

32-bit Counter: A non-negative integer which monotonically increases until it reaches a maximum value, when it wraps around and starts increasing again from zero. It has a maximum value of 2^32-1 (4294967295 decimal).

64-bit Gauge: non-negative integer, which may increase or decrease, but shall never exceed a maximum value, nor fall below a minimum value. The maximum value can not be greater than 2^64-1 (18446744073709551615 decimal), and the minimum value can not be smaller than 0. The value has its maximum value whenever the information being modeled is greater than or equal to its maximum value, and has its minimum value whenever the information being modeled is smaller than or equal to its minimum value. If the information being modeled subsequently decreases below (increases above) the maximum (minimum) value, the 64-bit Gauge also decreases (increases).

- Stat Type = 0: (32-bit Counter) Number of prefixes rejected by inbound policy.

- Stat Type = 1: (32-bit Counter) Number of (known) duplicate prefix advertisements.

- Stat Type = 2: (32-bit Counter) Number of (known) duplicate withdraws.

- Stat Type = 3: (32-bit Counter) Number of updates invalidated due to CLUSTER_LIST loop.

- Stat Type = 4: (32-bit Counter) Number of updates invalidated due to AS_PATH loop.

- Stat Type = 5: (32-bit Counter) Number of updates invalidated due to ORIGINATOR_ID.

- Stat Type = 6: (32-bit Counter) Number of updates invalidated due to AS_CONFED loop.

- Stat Type = 7: (64-bit Gauge) Number of routes in Adj-RIBs-In.

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

- Stat Type = 9: Number of routes in per-AFI/SAFI Adj-RIB-In. The value is structured as: AFI (2 bytes), SAFI (1 byte), followed by a 64-bit Gauge.

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

- Stat Type = 11: (32-bit Counter) Number of updates subjected to treat-as-withdraw treatment [I-D.ietf-idr-error-handling].

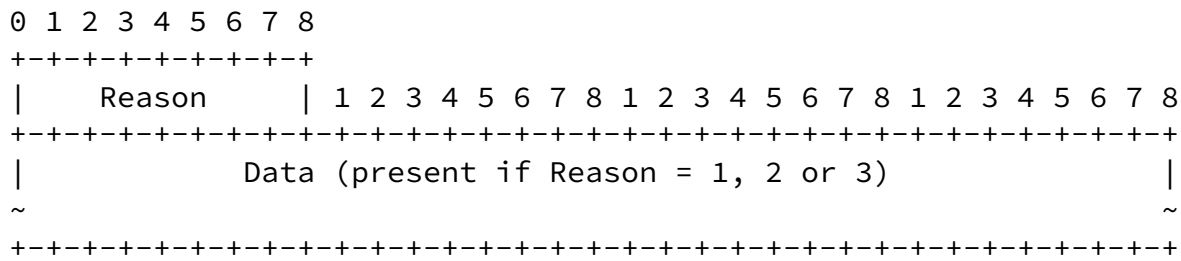
- Stat Type = 12: (32-bit Counter) Number of prefixes subjected to treat-as-withdraw treatment [I-D.ietf-idr-error-handling].

- Stat Type = 13: (32-bit Counter) Number of duplicate update messages received.

Note that although the current specification only specifies 4-byte counters and 8-byte gauges 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, at least one statistic MUST be carried in it.

4.9. Peer Down Notification

This message is used to indicate that a peering session was terminated.



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

- 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.

- 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]). Two bytes both set to zero are used to indicate that no relevant Event code is defined.

- 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.

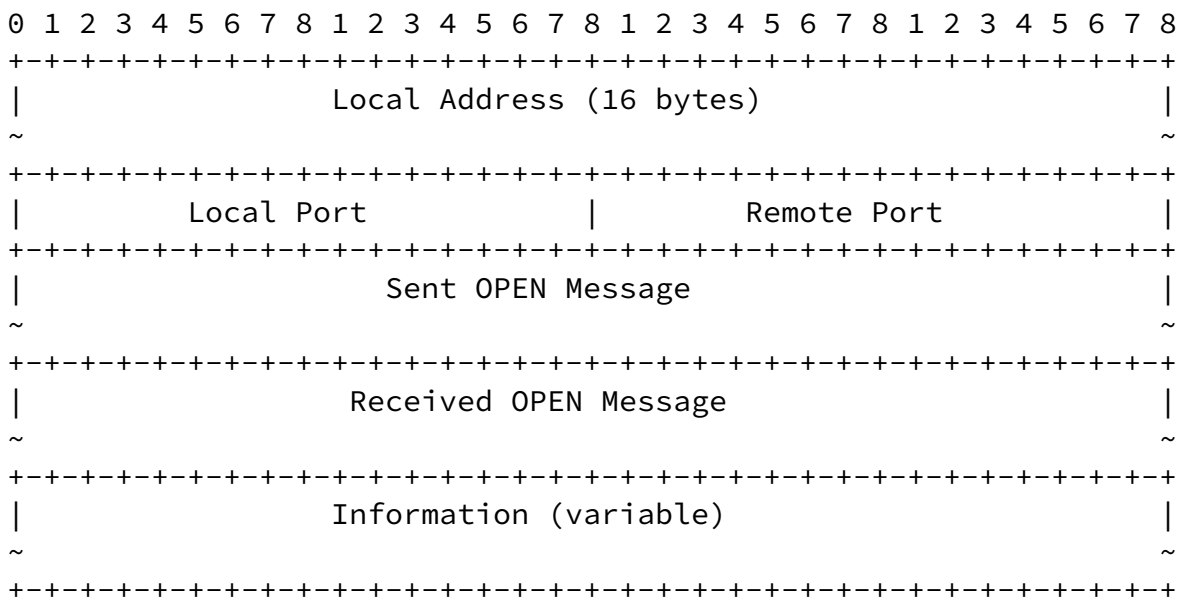
- Reason 4: The remote system closed the session without a notification message.

- Reason 5: Information for this peer will no longer be sent to the monitoring station for configuration reasons. This does not, strictly speaking, indicate that the peer has gone down, but it does indicate that the monitoring station will not receive updates for the peer.

A Peer Down message implicitly withdraws all routes that had been associated with the peer in question. A BMP implementation MAY omit sending explicit withdraws for such routes.

4.10. 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:



- 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.

- Local Port: The local port number associated with the peering TCP session, or zero if no TCP session actually exists (see Section 8.2).

- Remote Port: The remote port number associated with the peering TCP session, or zero if no TCP session actually exists (see Section 8.2). (Note that the remote address can be found in the Peer Address field of the fixed header.)

- Sent OPEN Message: The full OPEN message transmitted by the monitored router to its peer.

- Received OPEN Message: The full OPEN message received by the monitored router from its peer.

- Information: Information about the peer, using the Information TLV (Section 4.4) format. Only the string type is defined in this context; it may be repeated. Inclusion of the Information field is OPTIONAL. Its presence or absence can be inferred by inspection of the Message Length in the common header.

5. Route Monitoring

In BMP's normal operating mode, after the BMP session is up, Route Monitoring messages are used to provide a snapshot of the Adj-RIB-In of each monitored peer. This is done by sending all routes stored in the Adj-RIB-In of those peers using standard BGP Update messages, encapsulated in Route Monitoring messages. There is no requirement on the ordering of messages in the peer dumps. When the initial dump is completed for a given peer, this MUST be indicated by sending an End-of-RIB marker for that peer (as specified in Section 2 of [RFC4724], plus the BMP encapsulation header). See also Section 9.

A BMP speaker may send pre-policy routes, post-policy routes, or both. The selection may be due to implementation constraints (it is possible that a BGP implementation may not store, for example, routes which have been filtered out by policy). Pre-policy routes MUST have their L flag clear in the BMP header (see Section 4), post-policy routes MUST have their L flag set. When an implementation chooses to send both pre- and post-policy routes, it is effectively multiplexing two update streams onto the BMP session. The streams are distinguished by their L flags.

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.

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. As discussed above, it MAY generate either an update with the L flag clear, with it set, or two updates, one with the L flag clear and the other with the L flag set. When a route is withdrawn by a peer, a corresponding withdraw is sent to the monitor. The withdraw MUST have its L flag set to correspond to that of any previous announcement; if the route in question was previously announced with L flag both clear and set, the withdraw MUST similarly be sent twice, with L flag clear and set. 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 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. This is sometimes known as "state compression". 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.

6. Route Mirroring

Route Mirroring messages are provided for two primary reasons: First, to enable an implementation to operate in a mode where it provides a full-fidelity view of all messages received from its peers, without state compression. As we note in Section 5, BMP's normal operational mode cannot provide this. Implementors are strongly cautioned that without state compression, an implementation could require unbounded storage to buffer messages queued to be mirrored. This requirement, and concomitant performance implications, means that this mode of operation is unlikely to be suitable for implementation in conventional routers, and its use is NOT RECOMMENDED except in cases where implementors have carefully considered the tradeoffs.

The second application for Route Mirroring is for error reporting and diagnosis. When [I-D.ietf-idr-error-handling] is in use, a router can process BGP messages that are determined to contain errors, without resetting the BGP session. Such messages MAY be mirrored. The buffering used for such mirroring SHOULD be limited. If an errored message is unable to be mirrored due to buffer exhaustion, a message with the "Messages Lost" code SHOULD be sent to indicate this. (This implies that a buffer should be reserved for this use.)

7. 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.

8. Other Considerations

8.1. Multiple Instances

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 BGP instance).

8.2. Locally-Originated Routes

Some consideration is required for routes that are originated into BGP by the local router, whether as a result of redistribution from a another protocol or for some other reason.

Such routes can be modeled as having been sent by the router to itself, placing the router's own address in the Peer Address field of the header. It is RECOMMENDED that when doing so the router should use the same address it has used as its local address for the BMP session. Since in this case no transport session actually exists the Local and Remote Port fields of the Peer Up message MUST be set to zero. Clearly the OPEN Message fields of the Peer Up message will equally not have been physically transmitted, but should represent the relevant capabilities of the local router.

Also recall that the L flag is used to indicate locally-sourced routes, see Section 4.2.

9. 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.

10. IANA Considerations

IANA is requested to create the following registries.

10.1. BMP Message Types

This document defines five message types for transferring BGP messages between cooperating systems (Section 4):

- Type 0: Route Monitor
- Type 1: Statistics Report
- Type 2: Peer Down Notification
- Type 3: Peer Up Notification
- Type 4: Initiation
- Type 5: Termination
- Type 6: Mirroring

Type values 7 through 128 MUST be assigned using the "Standards Action" policy, and values 129 through 255 using the "Specification Required" policy defined in [RFC5226].

10.2. BMP Statistics Types

This document defines nine statistics types for statistics reporting (Section 4.8):

- Stat Type = 0: Number of prefixes rejected by inbound policy.
- Stat Type = 1: Number of (known) duplicate prefix advertisements.
- Stat Type = 2: Number of (known) duplicate withdraws.
- Stat Type = 3: Number of updates invalidated due to CLUSTER_LIST loop.
- Stat Type = 4: Number of updates invalidated due to AS_PATH loop.
- Stat Type = 5: Number of updates invalidated due to ORIGINATOR_ID.
- Stat Type = 6: Number of updates invalidated due to a loop found in AS_CONFED_SEQUENCE or AS_CONFED_SET.
- Stat Type = 7: Number of routes in Adj-RIBs-In.
- Stat Type = 8: Number of routes in Loc-RIB.
- Stat Type = 9: Number of routes in per-AFI/SAFI Adj-RIB-In.
- Stat Type = 10: Number of routes in per-AFI/SAFI Loc-RIB.
- Stat Type = 11: Number of updates subjected to treat-as-withdraw.
- Stat Type = 12: Number of prefixes subjected to treat-as-withdraw.
- Stat Type = 13: Number of duplicate update messages received.

Stat Type values 14 through 32767 MUST be assigned using the "Standards Action" policy, and values 32768 through 65535 using the "Specification Required" policy, defined in [RFC5226].

10.3. BMP Initiation Message TLVs

This document defines three types for information carried in the Initiation message (Section 4.3):

- Type = 0: String.
- Type = 1: sysDescr.
- Type = 2: sysName.

Information type values 3 through 32767 MUST be assigned using the "Standards Action" policy, and values 32768 through 65535 using the "Specification Required" policy, defined in [RFC5226].

10.4. BMP Termination Message TLVs

This document defines two types for information carried in the Termination message (Section 4.5):

- Type = 0: String.
- Type = 1: Reason.

Information type values 2 through 32767 MUST be assigned using the "Standards Action" policy, and values 32768 through 65535 using the "Specification Required" policy, defined in [RFC5226].

10.5. BMP Termination Message Reason Codes

This document defines four types for information carried in the Termination message (Section 4.5) Reason code,:

- Type = 0: Administratively closed.
- Type = 1: Unspecified reason.
- Type = 2: Out of resources.
- Type = 3: Redundant connection.
- Type = 4: Permanently administratively closed.

Information type values 5 through 32767 MUST be assigned using the "Standards Action" policy, and values 32768 through 65535 using the "Specification Required" policy, defined in [RFC5226].

10.6. BMP Peer Down Reason Codes

This document defines five types for information carried in the Peer Down Notification (Section 4.9) Reason code:

- Type = 1: Local system closed, NOTIFICATION PDU follows.
- Type = 2: Local system closed, FSM Event follows.
- Type = 3: Remote system closed, NOTIFICATION PDU follows.
- Type = 4: Remote system closed, no data.
- Type = 5: Peer de-configured.

Information type values 6 through 32767 MUST be assigned using the "Standards Action" policy, and values 32768 through 65535 using the

"Specification Required" policy, defined in [RFC5226]. Value 0 is reserved.

10.7. Route Mirroring TLVs

This document defines two types for information carried in the Route Mirroring message (Section 4.7):

- Type = 0: BGP Message.
- Type = 1: Information.

Information type values 2 through 32767 MUST be assigned using the "Standards Action" policy, and values 32768 through 65535 using the "Specification Required" policy, defined in [RFC5226].

10.8. BMP Route Mirroring Information Codes

This document defines two types for information carried in the Route Mirroring Information (Section 4.7) code:

- Type = 0: Errored PDU.
- Type = 1: Messages Lost.

Information type values 2 through 32767 MUST be assigned using the "Standards Action" policy, and values 32768 through 65535 using the "Specification Required" policy, defined in [RFC5226]. Value 0 is reserved.

11. 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 transport mechanism, such as IPSec [RFC4303] or TCP-AO [RFC5925], in order to provide mutual authentication, data integrity and transport protection.

Unless a transport that provides mutual authentication is used, an attacker could masquerade as the monitored router and trick a monitoring station into accepting false information.

12. Acknowledgements

Thanks to Michael Axelrod, Tim Evens, Pierre Francois, John ji Ioannidis, Mack McBride, Danny McPherson, David Meyer, Dimitri Papadimitriou, Robert Raszuk, Erik Romijn, and the members of the GROW working group for their comments.

13. References

13.1. Normative References

[I-D.ietf-idr-error-handling]
Chen, E., Scudder, J., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", draft-ietf-idr-error-handling-19 (work in progress), April 2015.

[RFC1213] McCloghrie, K. and M. Rose, "Management Information Base for Network Management of TCP/IP-based internets:MIB-II", STD 17, RFC 1213, March 1991.

[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.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet Autonomous System (AS) Number Space", RFC 6793, December 2012.

13.2. Informative References

[RFC1155] Rose, M. and K. McCloghrie, "Structure and identification of management information for TCP/IP-based internets", STD 16, RFC 1155, May 1990.

[RFC2856] Bierman, A., McCloghrie, K., and R. Presuhn, "Textual Conventions for Additional High Capacity Data Types", RFC 2856, June 2000.

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

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.

Appendix A. Changes Between BMP Versions 1 and 2

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

Appendix B. Changes Between BMP Versions 2 and 3

- Added a 32-bit length field to the fixed header.
- Clarified error handling.
- Added new stat types: 5 (number of updates invalidated due to ORIGINATOR_ID), 6 (number of updates invalidated due to AS_CONFED_SEQUENCE/AS_CONFED_SET), 7 (number of routes in Adj-RIB-In), 8 (number of routes in Loc-RIB), 9 (number of routes in Adj-RIB-In, per AFI/SAFI), 10 (numer of routes in Loc-RIB, per AFI/SAFI), 11 (number of updates subjected to treat-as-withdraw treatment), 12 (number of prefixes subjected to treat-as-withdraw treatment), and 13 (number of duplicate update messages received).
- Defined counters and gauges for use with stat types.
- For peer down messages, the relevant FSM event is to be sent in type 2 messages. Added type 5 to indicate peer is no longer monitored.
- Added local address and local and remote ports to the peer up message. Also optional descriptive string.
- Require End-of-RIB marker after initial dump.
- Added Initiation message with string content.
- Permit multiplexing pre- and post-policy feeds onto a single BMP session.
- Changed assignment policy for IANA registries.
- Changed "Loc-RIB" references to refer to "Post-Policy Adj-RIB-In", plus other editorial changes.
- Introduced option for monitoring station to be active party in initiating connection.
- Introduced Termination message.
- Added "route mirroring" mode.
- Added "A" flag to identify AS Path format in use.

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