                     BGP Monitoring Protocol (BMP)

Abstract

   This document defines the BGP Monitoring Protocol (BMP), which can be
   used to monitor BGP sessions. BMP is intended to provide a
   convenient interface for obtaining route views. Prior to the
   introduction of BMP, screen scraping was the most commonly used
   approach to obtaining such views. 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 is an Internet Standards Track document.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Further information on
   Internet Standards is available in Section 2 of RFC 7841.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   http://www.rfc-editor.org/info/rfc7854.

Table of Contents

1. Introduction

   Many researchers and network operators wish to have access to the
   contents of routers' BGP Routing Information Bases (RIBs) as well as
   a view of protocol updates the router is receiving. This monitoring
   task cannot be realized by standard protocol mechanisms. Prior to
   the introduction of BMP, this data could only be obtained through
   screen scraping.

   BMP provides access to the Adj-RIB-In of a peer on an ongoing basis
   and a periodic dump of certain statistics 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

   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 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:

   o Route Monitoring (RM): Used to provide 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: 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): 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: 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 the
      Established state, a Peer Up Notification is sent for each peer in
      the Established state when the BMP session itself comes up.

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

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

   o 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
      Protocol Data Units (PDUs).

3.2. Connection Establishment and Termination

   BMP operates over TCP. All options are controlled by configuration
   on the monitored router. No BMP 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 0) 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 and 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 end or restart BMP processing,
it simply drops the connection.

3.3. Lifecycle of a BMP Session

A router is configured to speak BMP to 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 that is in the
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 an 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 only
wants 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 peer becomes
Established, a corresponding Peer Up message is 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.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+
   |    Version    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                        Message Length                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Msg. Type   |
   +---------------+
```
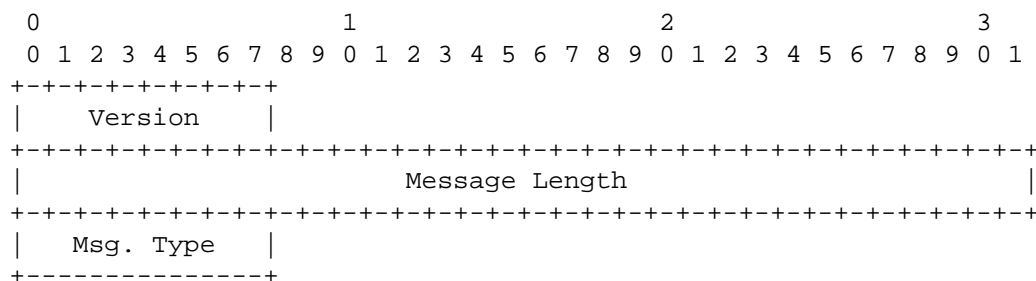
   o Version (1 byte): Indicates the BMP version. This is set to '3'
      for all messages defined in this specification. ('1' and '2' were
      used by draft versions of this document.) Version 0 is reserved
      and MUST NOT be sent.

   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
      * Type = 5: Termination Message
      * Type = 6: Route Mirroring 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.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Peer Type   |  Peer Flags   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Peer Distinguisher (present based on peer type)       |
   |                                                              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 Peer Address (16 bytes)                      |
   ~                                                              ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                         Peer AS                              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                       Peer BGP ID                            |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Timestamp (seconds)                       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                  Timestamp (microseconds)                    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
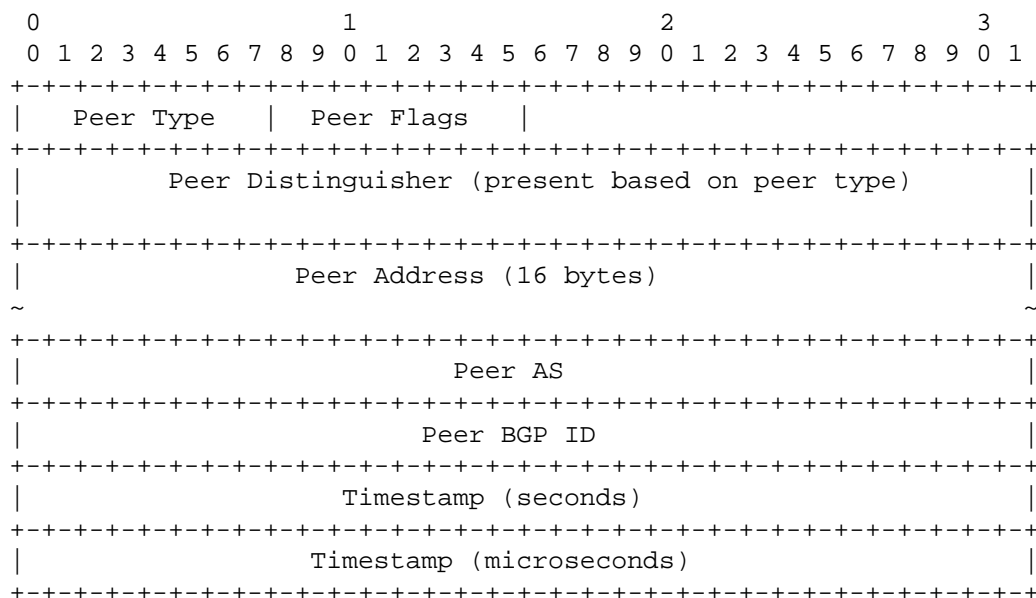
   o Peer Type (1 byte): Identifies the type of peer. Currently, three
      types of peers are identified:

      * Peer Type = 0: Global Instance Peer
      * Peer Type = 1: RD Instance Peer
      * Peer Type = 2: Local 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
                       +-+-+-+-+-+-+-+-+
                       |V|L|A| Reserved|
                       +-+-+-+-+-+-+-+-+
```
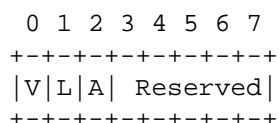
      * The V flag indicates that the Peer address is an IPv6 address.
         For IPv4 peers, this is set to 0.

          * 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 2-byte AS_PATH format. If set to 0,
             the message is formatted using the 4-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 a 4-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. They MUST be
             transmitted as 0 and their values MUST be ignored on receipt.

   o Peer Distinguisher (8 bytes): Routers today can have multiple
      instances (example: Layer 3 Virtual Private Networks (L3VPNs)
      [RFC4364]). 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 "RD Instance Peer", it is set to the
      route distinguisher of the particular instance the peer belongs
      to. If the peer is a "Local Instance Peer", it is set to a
      unique, locally defined value. In all cases, the effect is that
      the combination of the Peer Type and Peer Distinguisher is
      sufficient to disambiguate peers for which other identifying
      information might overlap.

   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 the
      12 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 [RFC6793], it should be padded with zeroes in the 16
      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-
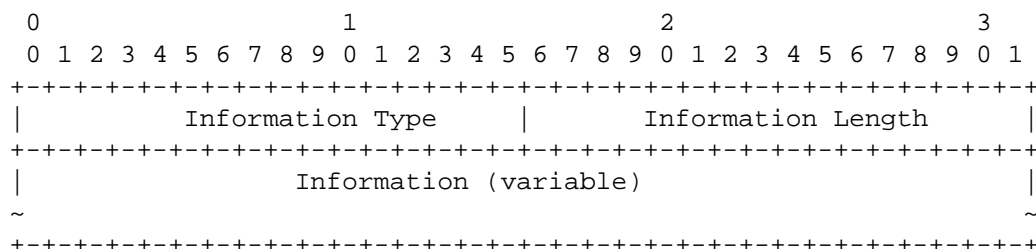   dependent.

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.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |          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. The value is administratively assigned. There is no
         requirement to terminate the string with a null (or any other
         particular) character -- the Information 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 an ASCII
            string whose value MUST be set to be equal to the value of the
            sysName MIB-II [RFC1213] object.

      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.

4.5. 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:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         Information Type     |       Information Length       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   Information (variable)                     |
    ~                                                              ~
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
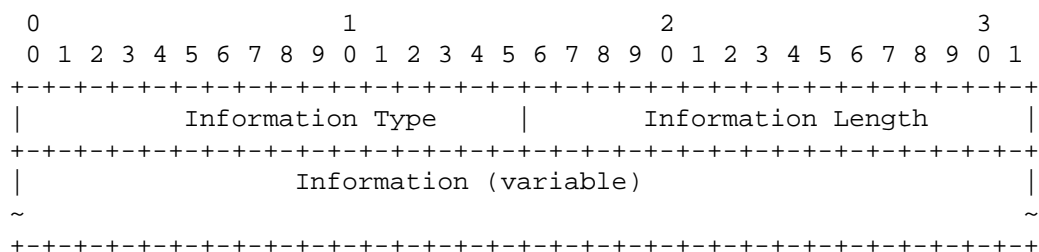
   o Information Type (2 bytes): Type of information provided. Defined
      types are:

      * Type = 0: String. The Information field contains a free-form
         UTF-8 string whose length is given by the Information Length
         field. 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 2-byte code
        indicating the reason that 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. Monitoring station should reduce
           (potentially to 0) 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 the
   ADJ-RIBs-In. They are also used for ongoing monitoring of the
   ADJ-RIB-In state. 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 the mirroring of messages that
   have been treated-as-withdraw [RFC7606], for debugging purposes.
   Mirrored messages may be sampled, or may be lossless. The Messages
   Lost Information code is provided to allow losses to be indicated.
   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 2-byte type code, a 2-byte length field, and a
variable-length value. Inclusion of any given TLV is OPTIONAL;
however, at least one TLV SHOULD be included, otherwise there's no
point in sending the message. Defined TLVs are as follows:

o 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.

o Type = 1: Information. A 2-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 [RFC7606]. 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
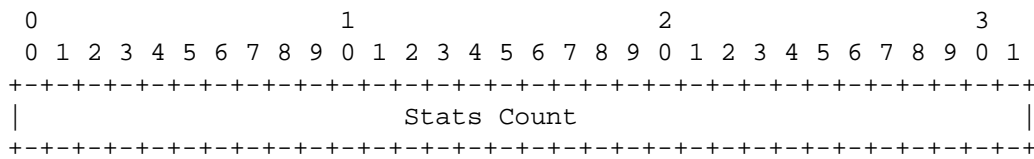   out of available buffer space to queue mirroring messages.

A Route Mirroring message may be sent any time it would be legal to
send a Route Monitoring message.

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.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Stats Count                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Each counter is encoded as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Stat Type            |             Stat Len          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Stat Data                          |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
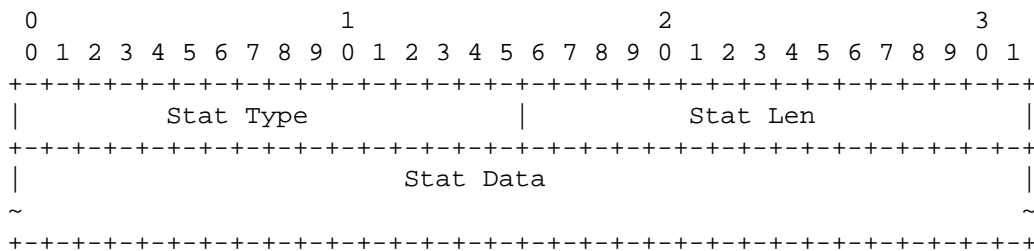
o Stat Type (2 bytes): Defines the type of the statistic carried in
   the Stat Data field.

o Stat Len (2 bytes): Defines the length of the Stat Data field.

This specification defines the following statistics. 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 in Section 3.2.3.3 of [RFC1155] and Section 4 of [RFC2856],
respectively:

32-bit Counter: A non-negative integer that monotonically increases
until it reaches a maximum value, when it wraps around and starts
increasing again from 0. It has a maximum value of $2^{32}-1$
(4294967295 decimal).

64-bit Gauge: A non-negative integer that may increase or decrease,
but shall never exceed a maximum value, nor fall below a minimum one.
The maximum value cannot be greater than $2^{64}-1$ (18446744073709551615
decimal), and the minimum value cannot 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 the maximum value (or increases above the minimum
value), the 64-bit Gauge also decreases (or increases).

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

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

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

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

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

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

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

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

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

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

   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.

   o Stat Type = 11: (32-bit Counter) Number of updates subjected to
      treat-as-withdraw treatment [RFC7606].

   o Stat Type = 12: (32-bit Counter) Number of prefixes subjected to
      treat-as-withdraw treatment [RFC7606].

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

   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 that 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.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+
   |    Reason     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Data (present if Reason = 1, 2 or 3)               |
   ~                                                               ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
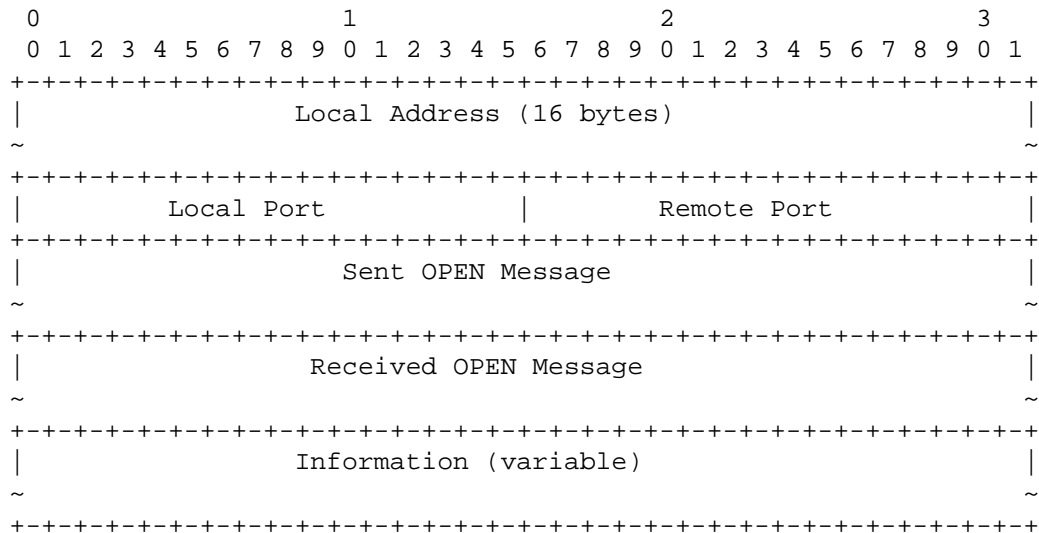
   Reason indicates why the session was closed. Defined values are:

   o 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.

   o Reason 2: The local system closed the session. No notification
      message was sent. Following the reason code is a 2-byte field
      containing the code corresponding to the Finite State Machine
      (FSM) Event that caused the system to close the session (see
      Section 8.1 of [RFC4271]). Two bytes both set to 0 are used to
      indicate that no relevant Event code is defined.

   o 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.

   o Reason 4: The remote system closed the session without a
      notification message. This includes any unexpected termination of
      the transport session, so in some cases both the local and remote
      systems might consider this to apply.

   o 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 were
   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 the Established state).
   Following the common BMP header and per-peer header is the following:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Local Address (16 bytes)                  |
   ~                                                              ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Local Port            |        Remote Port           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Sent OPEN Message                         |
   ~                                                              ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Received OPEN Message                     |
   ~                                                              ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Information (variable)                    |
   ~                                                              ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   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 the 12 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, or 0 if no TCP session actually exists (see Section 8.2).

   o Remote Port: The remote port number associated with the peering
      TCP session, or 0 if no TCP session actually exists (see
      Section 8.2). (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.

   o 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
   that 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
   0, 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 monitoring station 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
   monitoring station. 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. (Route mirroring (Section 6) is provided if
this is required.) While the router should attempt to generate
updates promptly, there is a finite time that could elapse between
reception of an update, 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. Route Mirroring 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. These tradeoffs include:
   router resource exhaustion, the potential to interfere with the
   transmission or reception of BGP UPDATE messages, and the slowing of
   routing convergence.

   The second application for Route Mirroring is for error reporting and
   diagnosis. When Revised Error Handling for BGP UPDATE messages
   [RFC7606] 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 that 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). Different BMP instances SHOULD
generate Initiation Messages that are distinct from one another, for
example, by using distinguishable sysNames or by the inclusion of
instance-identifying information in a string TLV.

8.2. Locally Originated Routes

Some consideration is required for routes originated into BGP by the
local router, whether as a result of redistribution from 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 0. Clearly, the OPEN Message fields of the Peer Up message will
equally not have been transmitted physically, 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 that would have to
correlate and reflect the deletion of that route in its internal
state. This is an operation that 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 it can safely ignore.

10. IANA Considerations

   IANA has created registries for the following BMP parameters, which
   are organized in a new group "BGP Monitoring Protocol (BMP)
   Parameters".

10.1. BMP Message Types

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

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

   Type values 0 through 127 MUST be assigned using the "Standards
   Action" policy, and values 128 through 250 using the "Specification
   Required" policy defined in [RFC5226]. Values 251 through 254 are
   Experimental, and value 255 is Reserved.

10.2. BMP Peer Types

   This document defines three types of peers for purposes of
   interpreting the Peer Distinguisher field (Section 4.2):

   o Peer Type = 0: Global Instance Peer
   o Peer Type = 1: RD Instance Peer
   o Peer Type = 2: Local Instance Peer

   Peer Type values 0 through 127 MUST be assigned using the "Standards
   Action" policy, and values 128 through 250 using the "Specification
   Required" policy, defined in [RFC5226]. Values 251 through 254 are
   Experimental, and value 255 is Reserved.

10.3. BMP Peer Flags

   This document defines three bit flags in the Peer Flags field of the
   per-peer header (Section 4.2). The bits are numbered from 0 (the
   high-order, or leftmost, bit) to 7 (the low-order, or rightmost,
   bit):

   o Flag 0: V flag
   o Flag 1: L flag
   o Flag 2: A flag

   Flags 3 through 7 are unassigned. The registration procedure for the
   registry is "Standards Action".

10.4. BMP Statistics Types

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

   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 a loop found
      in AS_CONFED_SEQUENCE or AS_CONFED_SET
   o Stat Type = 7: Number of routes in Adj-RIBs-In
   o Stat Type = 8: Number of routes in Loc-RIB
   o Stat Type = 9: Number of routes in per-AFI/SAFI Adj-RIB-In
   o Stat Type = 10: Number of routes in per-AFI/SAFI Loc-RIB
   o Stat Type = 11: Number of updates subjected to treat-as-withdraw
   o Stat Type = 12: Number of prefixes subjected to treat-as-withdraw
   o Stat Type = 13: Number of duplicate update messages received

   Stat Type values 0 through 32767 MUST be assigned using the
   "Standards Action" policy, and values 32768 through 65530 using the
   "Specification Required" policy, defined in [RFC5226]. Values 65531
   through 65534 are Experimental, and value 65535 is Reserved.

10.5. BMP Initiation Message TLVs

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

   o Type = 0: String
   o Type = 1: sysDescr
   o Type = 2: sysName

   Information type values 0 through 32767 MUST be assigned using the
   "Standards Action" policy, and values 32768 through 65530 using the
   "Specification Required" policy, defined in [RFC5226]. Values 65531
   through 65534 are Experimental, and value 65535 is reserved.

10.6. BMP Termination Message TLVs

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

   o Type = 0: String
   o Type = 1: Reason

   Information type values 0 through 32767 MUST be assigned using the
   "Standards Action" policy, and values 32768 through 65530 using the
   "Specification Required" policy, defined in [RFC5226]. Values 65531
   through 65534 are Experimental, and value 65535 is Reserved.

10.7. BMP Termination Message Reason Codes

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

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

   Information type values 0 through 32767 MUST be assigned using the
   "Standards Action" policy, and values 32768 through 65530 using the
   "Specification Required" policy, defined in [RFC5226]. Values 65531
   through 65534 are Experimental, and value 65535 is Reserved.

10.8. BMP Peer Down Reason Codes

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

   o Type = 0 is Reserved.
   o Type = 1: Local system closed, NOTIFICATION PDU follows
   o Type = 2: Local system closed, FSM Event follows
   o Type = 3: Remote system closed, NOTIFICATION PDU follows
   o Type = 4: Remote system closed, no data
   o Type = 5: Peer de-configured

   Information type values 0 through 32767 MUST be assigned using the
   "Standards Action" policy, and values 32768 through 65530 using the
   "Specification Required" policy, defined in [RFC5226]. Values 65531
   through 65534 are Experimental, and values 0 and 65535 are Reserved.

10.9. Route Mirroring TLVs

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

   o Type = 0: BGP Message
   o Type = 1: Information

   Information type values 0 through 32767 MUST be assigned using the
   "Standards Action" policy, and values 32768 through 65530 using the
   "Specification Required" policy, defined in [RFC5226]. Values 65531
   through 65534 are Experimental, and value 65535 is Reserved.

10.10. BMP Route Mirroring Information Codes

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

   o Type = 0: Errored PDU
   o Type = 1: Messages Lost

   Information type values 0 through 32767 MUST be assigned using the
   "Standards Action" policy, and values 32768 through 65530 using the
   "Specification Required" policy, defined in [RFC5226]. Values 65531
   through 65534 are Experimental, and value 65535 is Reserved.

11. Security Considerations

   This document defines a mechanism to obtain a full dump or provide
   continuous monitoring of a BGP speaker's BGP routes, including
   received BGP messages. This capability could allow an outside party
   to obtain information not otherwise obtainable. For example,
   although it's hard to consider the content of BGP routes in the
   public Internet to be confidential, BGP is used in private contexts
   as well, for example, for L3VPN [RFC4364]. As another example, a
   clever attacker might be able to infer the content of the monitored
   router's import policy by comparing the pre-policy routes exposed by
   BMP, to post-policy routes exported in BGP.

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

   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, or they could
   masquerade as a monitoring station and gain unauthorized access to
   BMP data. Unless a transport that provides confidentiality is used,
   a passive or active attacker could gain access to, or tamper with,
   the BMP data in flight.

   Where the security considerations outlined above are a concern, users
   of this protocol should use IPsec [RFC4303] in tunnel mode with pre-
   shared keys.

12. References

12.1. Normative References

   [RFC1213] McCloghrie, K. and M. Rose, "Management Information Base
              for Network Management of TCP/IP-based internets: MIB-II",
              STD 17, RFC 1213, DOI 10.17487/RFC1213, March 1991,
              <http://www.rfc-editor.org/info/rfc1213>.

   [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>.

   [RFC4724] Sangli, S., Chen, E., Fernando, R., Scudder, J., and Y.
              Rekhter, "Graceful Restart Mechanism for BGP", RFC 4724,
              DOI 10.17487/RFC4724, January 2007,
              <http://www.rfc-editor.org/info/rfc4724>.

   [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an
              IANA Considerations Section in RFCs", BCP 26, RFC 5226,
              DOI 10.17487/RFC5226, May 2008,
              <http://www.rfc-editor.org/info/rfc5226>.

   [RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet
              Autonomous System (AS) Number Space", RFC 6793,
              DOI 10.17487/RFC6793, December 2012,
              <http://www.rfc-editor.org/info/rfc6793>.

   [RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K.
              Patel, "Revised Error Handling for BGP UPDATE Messages",
              RFC 7606, DOI 10.17487/RFC7606, August 2015,
              <http://www.rfc-editor.org/info/rfc7606>.

12.2. Informative References

   [RFC1155] Rose, M. and K. McCloghrie, "Structure and identification
              of management information for TCP/IP-based internets",
              STD 16, RFC 1155, DOI 10.17487/RFC1155, May 1990,
              <http://www.rfc-editor.org/info/rfc1155>.

   [RFC2856] Bierman, A., McCloghrie, K., and R. Presuhn, "Textual
              Conventions for Additional High Capacity Data Types",
              RFC 2856, DOI 10.17487/RFC2856, June 2000,
              <http://www.rfc-editor.org/info/rfc2856>.

   [RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)",
              RFC 4303, DOI 10.17487/RFC4303, December 2005,
              <http://www.rfc-editor.org/info/rfc4303>.

   [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private
              Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February
              2006, <http://www.rfc-editor.org/info/rfc4364>.

Authors' Addresses

   John Scudder (editor)
   Juniper Networks
   1194 N. Mathilda Ave.
   Sunnyvale, CA 94089
   United States

   Email: jgs@juniper.net

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

   Email: rex@cisco.com

   Stephen Stuart
   Google
   1600 Amphitheatre Parkway
   Mountain View, CA 94043
   United States

   Email: sstuart@google.com