Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 16, 2016

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

January 13, 2016

BGP Monitoring Protocol
draft-ietf-grow-bmp-17

Abstract

This document defines a protocol, BMP, that can be used to monitor BGP sessions. BMP is intended to provide a convenient interface for obtaining route views. Prior to 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 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 July 16, 2016.

Copyright Notice

Table of Contents

1. Introduction

Many researchers and network operators wish to have access to the contents of routers' BGP 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 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 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): 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.

- Peer Down Notification: A message sent to indicate 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: A message sent to indicate 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 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 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 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 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 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.)

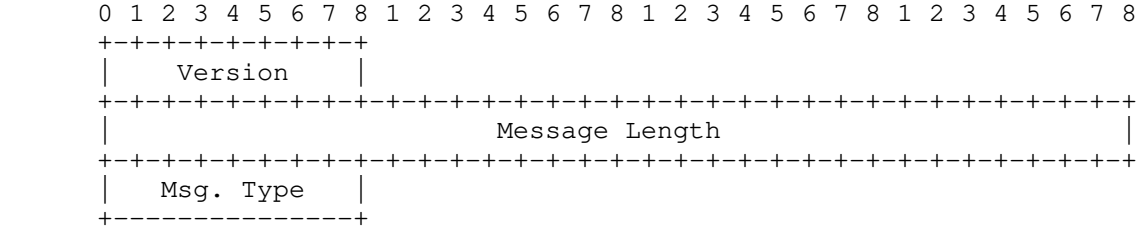
Following the initial table dump, the router sends incremental updates encapsulated in Route Monitoring messages. It MAY periodically send Stats Reports or even new Initiation messages, according to configuration. If any new monitored BGP peer becomes Established, a corresponding Peer Up message is sent. 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.



- 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).

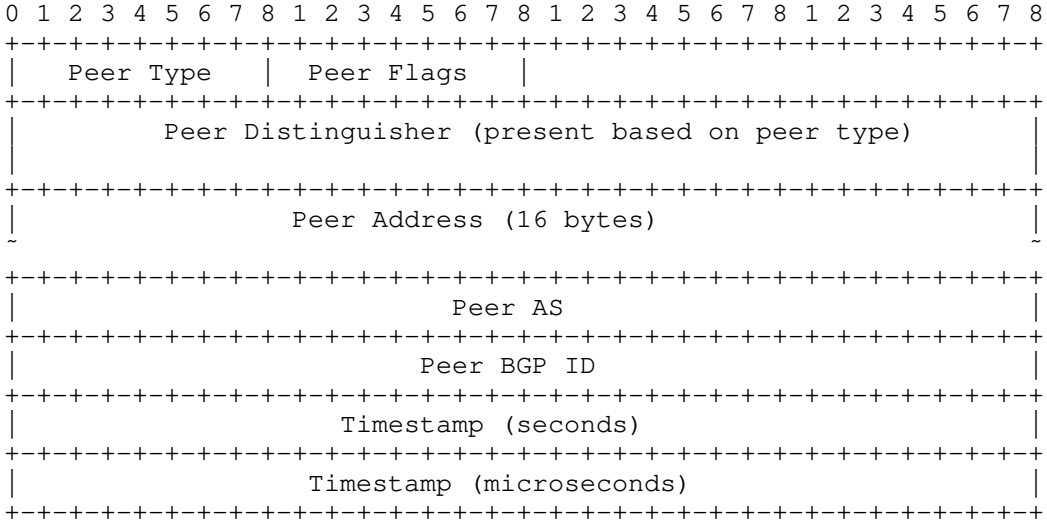
- Message Type (1 byte): This identifies the type of the BMP message. A BMP implementation MUST ignore unrecognized message types upon receipt.

  * 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
* Type = 6: Route Mirroring Message

4.2. Per-Peer Header

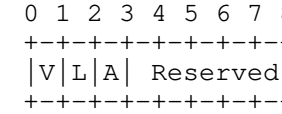
The per-peer header follows the common header for most BMP messages. The rest of the data in a BMP message is dependent on the "Message Type" field in the common header.



- Peer Type (1 byte): Identifies the type of the peer. Currently two types of peers are identified,

  * Peer Type = 0: Global Instance Peer
  * Peer Type = 1: RD Instance Peer
  * Peer Type = 2: Local 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 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 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. They MUST be transmitted as zero and their values MUST be ignored on receipt.

- Peer Distinguisher (8 bytes): Routers today can have multiple instances (example: 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.

- 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.

- 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.

- 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.

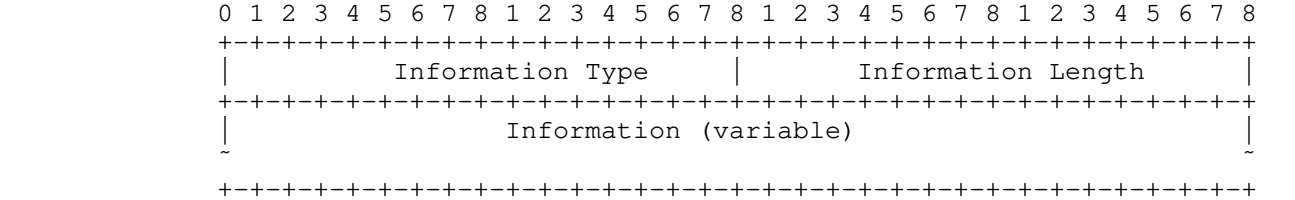
4.3. Initiation Message

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. 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.

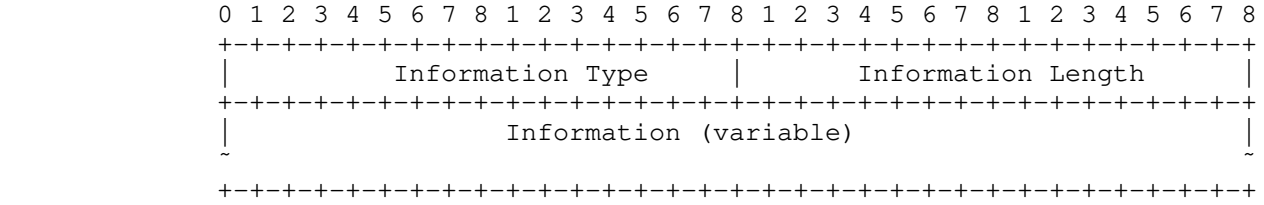
o Information Length (2 bytes): The length of the following Information field, in bytes.

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:



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 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 this connection is redundant with another one.

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

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

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

4.6. Route Monitoring

Route Monitoring messages are used for initial synchronization of ADJ-RIBs-In. They are also used for ongoing monitoring of 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 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 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 [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.

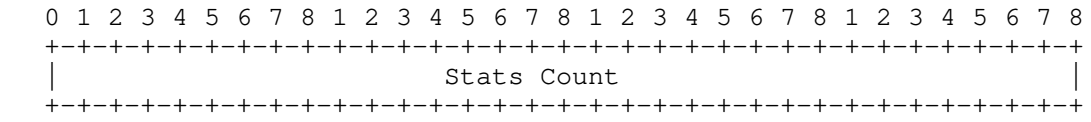
A Route Mirroring message may be sent any time it would be legal to send a Route Monitoring message.

4.8. Stats Reports

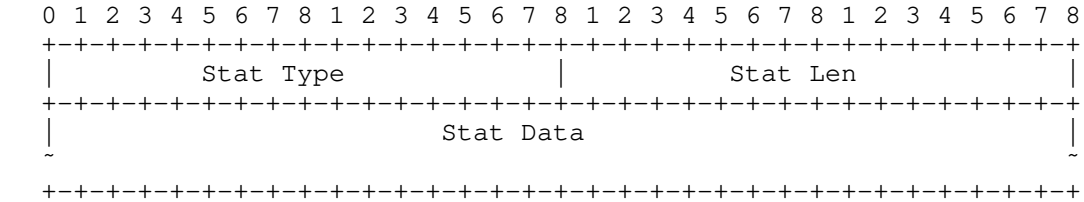
These messages contain information that could be used by the monitoring station to observe interesting events that occur on the router.

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.



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.

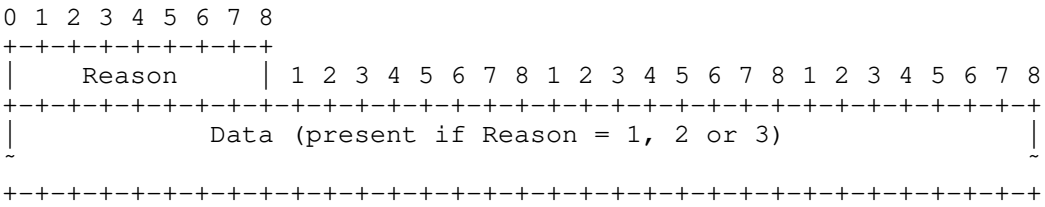
- Stat Type = 1: (32-bit Counter) Number of (known) duplicate prefix advertisements.

- 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 [RFC7606].
- Stat Type = 12: (32-bit Counter) Number of prefixes subjected to treat-as-withdraw treatment [RFC7606].
- 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 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 that caused the system to close the session (see Section 8.1 of [RFC4271]). Two bytes both set to zero are used to indicate 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.

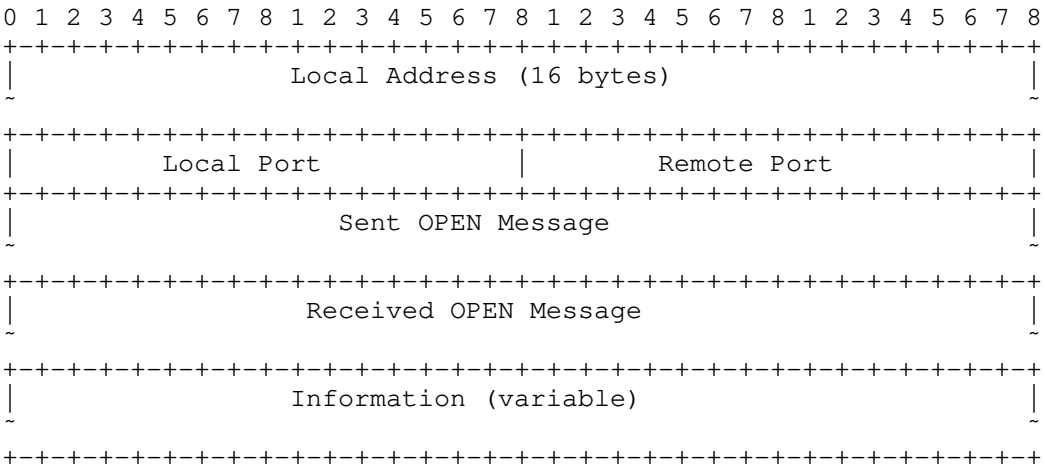
- Reason 4: The remote system closed the session without a notification message. This includes any unexpected termination of the transport session, so in some cases both the local and remote systems might consider this to apply.

- Reason 5: Information for this peer will no longer be sent to the monitoring station for configuration reasons. This does not, strictly speaking, indicate the peer has gone down, but it does indicate 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 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 the 12 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). (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 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 zero, indicating 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 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 flow-block BGP peers, and the slowing of routing convergence.

The second application for Route Mirroring is for error reporting and diagnosis. When [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 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). Different BMP instances SHOULD generate Initiation Messages that are distinct from one another, for example by using distinguishable sysNames or by 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 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 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 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 is requested to create registries for the following BMP parameters, to be 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):

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

Type values 0 through 128 MUST be assigned using the "Standards Action" policy, and values 129 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 two types of peers for purposes of interpreting the Peer Distinguisher field (Section 4.2):

- Peer Type = 0: Global Instance Peer.
- Peer Type = 1: RD Instance Peer.
- 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 zero (the high-order, or leftmost, bit) to seven (the low-order, or rightmost, bit):

- Flag 0: V flag.
- Flag 1: L flag.
- Flag 2: A flag.

Flags 3 through 7 MUST be assigned using the "Standards Action" policy.

10.4. BMP Statistics Types

This document defines fourteen 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 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):

- Type = 0: String.
- Type = 1: sysDescr.
- 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):

- Type = 0: String.
- 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,:

- 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 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):

- Type = 0 is reserved.
- 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 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):

- Type = 0: BGP Message.
- 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:

- Type = 0: Errored PDU.
- 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 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 preshared keys.

12. Acknowledgements

Thanks to Ebben Aries, Michael Axelrod, Serpil Bayraktar, Tim Evens, Pierre Francois, Jeffrey Haas, John ji Ioannidis, John Kemp, Mack McBride, Danny McPherson, David Meyer, Dimitri Papadimitriou, Tom Petch, Robert Raszuk, Erik Romijn, Peter Schoenmaker and the members of the GROW working group for their comments.

13. References

13.1. 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>.

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, 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>.

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 (editor)
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

Global Routing Operations
Internet-Draft
Intended status: Informational
Expires: November 6, 2016

K. Sriram
D. Montgomery
US NIST
D. McPherson
E. Osterweil
Verisign, Inc.
B. Dickson
May 5, 2016

Problem Definition and Classification of BGP Route Leaks
draft-ietf-grow-route-leak-problem-definition-06

Abstract

A systemic vulnerability of the Border Gateway Protocol routing system, known as 'route leaks', has received significant attention in recent years. Frequent incidents that result in significant disruptions to Internet routing are labeled "route leaks", but to date a common definition of the term has been lacking. This document provides a working definition of route leaks, keeping in mind the real occurrences that have received significant attention. Further, this document attempts to enumerate (though not exhaustively) different types of route leaks based on observed events on the Internet. The aim is to provide a taxonomy that covers several forms of route leaks that have been observed and are of concern to Internet user community as well as the network operator community.

Table of Contents

1. Introduction

Frequent incidents [Huston2012][Cowie2013][Toonk2015-A][Toonk2015-B][Cowie2010][Madory][Zmijewski][Paseka][LRL][Khare] that result in significant disruptions to Internet routing are commonly called "route leaks". Examination of the details of some of these incidents reveals that they vary in their form and technical details. In order to pursue solutions to "the route leak problem" it is important to first provide a clear, technical definition of the problem and enumerate its most common forms. Section 2 provides a working definition of route leaks, keeping in view many recent incidents that have received significant attention. Section 3 attempts to enumerate (though not exhaustively) different types of route leaks based on

observed events on the Internet. Further, Section 3 provides a taxonomy that covers several forms of route leaks that have been observed and are of concern to Internet user community as well as the network operator community. This document builds on and extends earlier work in the IETF [draft-dickson-sidr-route-leak-def][draft-dickson-sidr-route-leak-reqts].

2. Working Definition of Route Leaks

A proposed working definition of route leak is as follows:

A "route leak" is the propagation of routing announcement(s) beyond their intended scope. That is, an AS's announcement of a learned BGP route to another AS is in violation of the intended policies of the receiver, the sender and/or one of the ASes along the preceding AS path. The intended scope is usually defined by a set of local redistribution/filtering policies distributed among the ASes involved. Often, these intended policies are defined in terms of the pair-wise peering business relationship between ASes (e.g., customer, transit provider, peer). (For literature related to AS relationships and routing policies, see [Gao] [Luckie] [Gill]. For measurements of valley-free violations in Internet routing, see [Anwar] [Giotsas] [Wijchers].)

The result of a route leak can be redirection of traffic through an unintended path which may enable eavesdropping or traffic analysis, and may or may not result in an overload or black-hole. Route leaks can be accidental or malicious, but most often arise from accidental misconfigurations.

The above definition is not intended to be all encompassing. Our aim here is to have a working definition that fits enough observed incidents so that the IETF community has a basis for developing solutions for route leak detection and mitigation.

3. Classification of Route Leaks Based on Documented Events

As illustrated in Figure 1, a common form of route leak occurs when a multi-homed customer AS (such as AS3 in Figure 1) learns a prefix update from one transit provider (ISP1) and leaks the update to another transit provider (ISP2) in violation of intended routing policies, and further the second transit provider does not detect the leak and propagates the leaked update to its customers, peers, and transit ISPs.

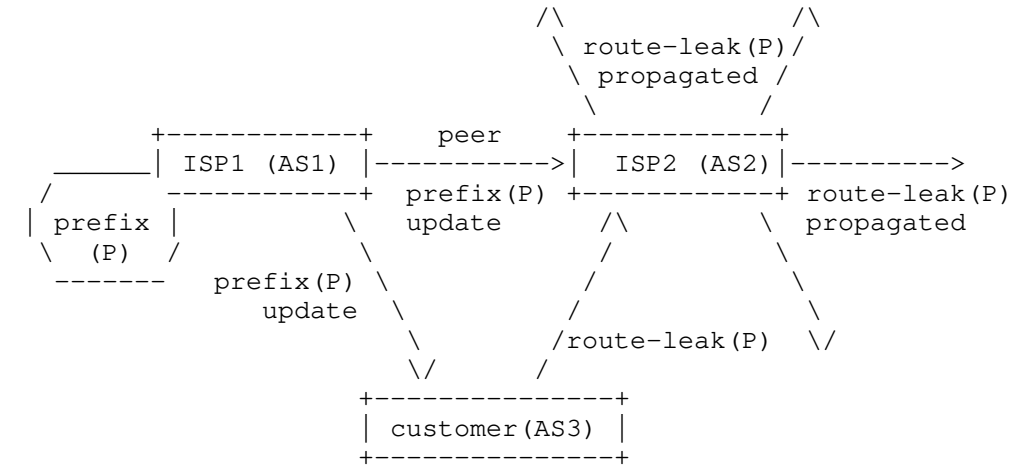


Figure 1: Illustration of the basic notion of a route leak.

This document proposes the following taxonomy to cover several types of observed route leaks, while acknowledging that the list is not meant to be exhaustive. In what follows, the AS that announces a route that is in violation of the intended policies is referred to as the "offending AS".

3.1. Type 1: Hairpin Turn with Full Prefix

Description: A multi-homed AS learns a route from one upstream ISP and simply propagates it to another upstream ISP (the turn essentially resembling a hairpin). Neither the prefix nor the AS path in the update is altered. This is similar to a straight forward path-poisoning attack [Kapela-Pilosov], but with full prefix. It should be noted that leaks of this type are often accidental (i.e. not malicious). The update basically makes a hairpin turn at the offending AS's multi-homed AS. The leak often succeeds (i.e. leaked update is accepted and propagated) because the second ISP prefers customer announcement over peer announcement of the same prefix. Data packets would reach the legitimate destination albeit via the offending AS, unless they are dropped at the offending AS due to its inability to handle resulting large volumes of traffic.

- Example incidents: Examples of Type 1 route-leak incidents are (1) the Dodo-Telstra incident in March 2012 [Huston2012], (2) the VolumeDrive-Atrato incident in September 2014 [Madory], and (3) the massive Telekom Malaysia route leak of about 179,000 prefixes, which in turn Level3 accepted and propagated [Toonk2015-B].

3.2. Type 2: Lateral ISP-ISP-ISP Leak

Description: The term "lateral" here is synonymous with "non-transit" or "peer-to-peer". This type of route leak typically occurs when, for example, three sequential ISP peers (e.g. ISP-A, ISP-B, and ISP-C) are involved, and ISP-B receives a route from ISP-A and in turn leaks it to ISP-C. The typical routing policy between laterally (i.e. non-transit) peering ISPs is that they should only propagate to each other their respective customer prefixes.

- Example incidents: In [Mauch-nanog][Mauch], route leaks of this type are reported by monitoring updates in the global BGP system and finding three or more very large ISP ASNs in a sequence in a BGP update's AS path. [Mauch] observes that its detection algorithm detects for these anomalies and potentially route leaks because very large ISPs do not in general buy transit services from each other. However, it also notes that there are exceptions when one very large ISP does indeed buy transit from another very large ISP, and accordingly exceptions are made in its detection algorithm for known cases.

3.3. Type 3: Leak of Transit-Provider Prefixes to Peer

Description: This type of route leak occurs when an offending AS leaks routes learned from its transit provider to a lateral (i.e. non-transit) peer.

- Example incidents: The incidents reported in [Mauch] include the Type 3 leaks.

3.4. Type 4: Leak of Peer Prefixes to Transit Provider

Description: This type of route leak occurs when an offending AS leaks routes learned from a lateral (i.e. non-transit) peer to its (the AS's) own transit provider. These leaked routes typically originate from the customer cone of the lateral peer.

- Example incidents: Examples of Type 4 route-leak incidents are (1) the Axcelx-Hibernia route leak of Amazon Web Services (AWS) prefixes causing disruption of AWS and a variety of services that run on AWS [Kephart],(2) the Hathway-Airtel route leak of 336 Google prefixes causing widespread interruption of Google services in Europe and Asia [Toonk2015-A], (3) the Moratel-PCCW route leak of Google prefixes causing Google's services to go offline [Paseka], and (4) Some of the example incidents cited for Type 1 route leaks above are also inclusive of Type 4 route leaks. For instance, in the Dodo-Telstra incident [Huston2012], the leaked

routes from Dodo to Telstra included routes that Dodo learned from its transit providers as well as lateral peers.

3.5. Type 5: Prefix Re-Origination with Data Path to Legitimate Origin

Description: A multi-homed AS learns a route from one upstream ISP and announces the prefix to another upstream ISP as if it is being originated by it (i.e. strips the received AS path, and re-originates the prefix). This can be called re-origination or mis-origination. However, somehow a reverse path to the legitimate origination AS may be present and data packets reach the legitimate destination albeit via the offending AS. (Note: The presence of a reverse path here is not attributable to the use of path poisoning trick by the offending AS.) But sometimes the reverse path may not be present, and data packets destined for the leaked prefix may be simply discarded at the offending AS.

- Example incidents: Examples of Type 5 route leak include (1) the China Telecom incident in April 2010 [Hiran][Cowie2010][Labovitz], (2) the Belarusian GlobalOneBel route leak incidents in February-March 2013 and May 2013 [Cowie2013], (3) the Icelandic Opin Kerfi-Simmin route leak incidents in July-August 2013 [Cowie2013], and (4) the Indosat route leak incident in April 2014 [Zmijewski]. The reverse paths (i.e. data paths from the offending AS to the legitimate destinations) were present in incidents #1, #2 and #3 cited above, but not in incident #4. In incident #4, the misrouted data packets were dropped at Indosat's AS.

3.6. Type 6: Accidental Leak of Internal Prefixes and More Specific Prefixes

Description: An offending AS simply leaks its internal prefixes to one or more of its transit-provider ASes and/or ISP peers. The leaked internal prefixes are often more specific prefixes subsumed by an already announced less specific prefix. The more specific prefixes were not intended to be routed in eBGP. Further, the AS receiving those leaks fails to filter them. Typically, these leaked announcements are due to some transient failures within the AS; they are short-lived and typically withdrawn quickly following the announcements. However, these more specific prefixes may momentarily cause the routes to be preferred over other aggregate (i.e. less specific) route announcements, thus redirecting traffic from its normal best path.

- Example incidents: Leaks of internal routes occur frequently (e.g. multiple times in a week), and the number of prefixes leaked range from hundreds to thousands per incident. One highly conspicuous and widely disruptive leak of internal routes happened in August

2014 when AS701 and AS705 leaked about 22,000 more specifics of already announced aggregates [Huston2014][Toonk2014].

4. Additional Comments about the Classification

It is worth noting that Types 1 through 4 are similar in that a route is leaked in violation of policy in each case, but what varies is the context of the leaked-route source AS and destination AS roles.

Type 5 route leak (i.e. prefix mis-origination with data path to legitimate origin) can also happen in conjunction with the AS relationship contexts in Types 2, 3, and 4. While these possibilities are acknowledged, simply enumerating more types to consider all such special cases does not add value as far as solution development for route leaks is concerned. Hence, the special cases mentioned here are not included in enumerating route leak types.

5. Security Considerations

No security considerations apply since this is a problem definition document.

6. IANA Considerations

This document does not require an action from IANA.

7. Acknowledgements

The authors wish to thank Jared Mauch, Jeff Haas, Warren Kumari, Amogh Dhamdhere, Jakob Heitz, Geoff Huston, Randy Bush, Job Snijders, Ruediger Volk, Andrei Robachevsky, Charles van Niman, Chris Morrow, and Sandy Murphy for comments, suggestions, and critique. The authors are also thankful to Padma Krishnaswamy, Oliver Borchert, and Okhee Kim for their comments and review.

8. Informative References

[Anwar] Anwar, R., Niaz, H., Choffnes, D., Cunha, I., Gill, P., and N. Katz-Bassett, "Investigating Interdomain Routing Policies in the Wild", ACM Internet Measurement Conference (IMC), October 2015, <http://www.cs.usc.edu/assets/007/94928.pdf>.

[Cowie2010] Cowie, J., "China's 18 Minute Mystery", Dyn Research/Renesys Blog, November 2010, <http://research.dyn.com/2010/11/chinas-18-minute-mystery/>.

[Cowie2013]
Cowie, J., "The New Threat: Targeted Internet Traffic Misdirection", Dyn Research/Renesys Blog, November 2013, <http://research.dyn.com/2013/11/mitm-internet-hijacking/>.

[draft-dickson-sidr-route-leak-def]
Dickson, B., "Route Leaks -- Definitions", IETF Internet Draft (expired), October 2012, <https://tools.ietf.org/html/draft-dickson-sidr-route-leak-def-03>.

[draft-dickson-sidr-route-leak-reqts]
Dickson, B., "Route Leaks -- Requirements for Detection and Prevention thereof", IETF Internet Draft (expired), March 2012, <http://tools.ietf.org/html/draft-dickson-sidr-route-leak-reqts-02>.

[Gao] Gao, L. and J. Rexford, "Stable Internet routing without global coordination", IEEE/ACM Transactions on Networking, December 2001, <http://www.cs.princeton.edu/~jrex/papers/sigmetrics00.long.pdf>.

[Gill] Gill, P., Schapira, M., and S. Goldberg, "A Survey of Interdomain Routing Policies", ACM SIGCOMM Computer Communication Review, January 2014, <http://www.cs.bu.edu/~goldbe/papers/survey.pdf>.

[Giotsas] Giotsas, V. and S. Zhou, "Valley-free violation in Internet routing - Analysis based on BGP Community data", IEEE ICC 2012, June 2012.

[Hiran] Hiran, R., Carlsson, N., and P. Gill, "Characterizing Large-scale Routing Anomalies: A Case Study of the China Telecom Incident", PAM 2013, March 2013, <http://www3.cs.stonybrook.edu/~phillipa/papers/CTelecom.html>.

[Huston2012]
Huston, G., "Leaking Routes", March 2012, <http://labs.apnic.net/blabs/?p=139/>.

[Huston2014]
Huston, G., "What's so special about 512?", September 2014, <http://labs.apnic.net/blabs/?p=520/>.

[Kapela-Pilosov]
Pilosov, A. and T. Kapela, "Stealing the Internet: An Internet-Scale Man in the Middle Attack", DEFCON-16 Las Vegas, NV, USA, August 2008, <https://www.defcon.org/images/defcon-16/dc16-presentations/defcon-16-pilosov-kapela.pdf>.

[Kephart] Kephart, N., "Route Leak Causes Amazon and AWS Outage", ThousandEyes Blog, June 2015, <https://blog.thousandeyes.com/route-leak-causes-amazon-and-aws-outage>.

[Khare] Khare, V., Ju, Q., and B. Zhang, "Concurrent Prefix Hijacks: Occurrence and Impacts", IMC 2012, Boston, MA, November 2012, <http://www.cs.arizona.edu/~bzhang/paper/12-imc-hijack.pdf>.

[Labovitz]
Labovitz, C., "Additional Discussion of the April China BGP Hijack Incident", Arbor Networks IT Security Blog, November 2010, <http://www.arbornetworks.com/asert/2010/11/additional-discussion-of-the-april-china-bgp-hijack-incident/>.

[LRL] Khare, V., Ju, Q., and B. Zhang, "Large Route Leaks", Project web page, 2012, <http://nrl.cs.arizona.edu/projects/lsrl-events-from-2003-to-2009/>.

[Luckie] Luckie, M., Huffaker, B., Dhamdhere, A., Giotsas, V., and kc. claffy, "AS Relationships, Customer Cones, and Validation", IMC 2013, October 2013, <http://www.caida.org/~amogh/papers/asrank-IMC13.pdf>.

[Madory] Madory, D., "Why Far-Flung Parts of the Internet Broke Today", Dyn Research/Renesys Blog, September 2014, <http://research.dyn.com/2014/09/why-the-internet-broke-today/>.

[Mauch] Mauch, J., "BGP Routing Leak Detection System", Project web page, 2014, <http://puck.nether.net/bgp/leakinfo.cgi/>.

[Mauch-nanog]
Mauch, J., "Detecting Routing Leaks by Counting", NANOG-41 Albuquerque, NM, USA, October 2007, <https://www.nanog.org/meetings/nanog41/presentations/mauch-lightning.pdf>.

[Paseka] Paseka, T., "Why Google Went Offline Today and a Bit about How the Internet Works", CloudFare Blog, November 2012, <http://blog.cloudflare.com/why-google-went-offline-today-and-a-bit-about/>.

[Toonk2014] Toonk, A., "What caused today's Internet hiccup", August 2014, <http://www.bgpmon.net/what-caused-todays-internet-hiccup/>.

[Toonk2015-A] Toonk, A., "What caused the Google service interruption", March 2015, <http://www.bgpmon.net/what-caused-the-google-service-interruption/>.

[Toonk2015-B] Toonk, A., "Massive route leak causes Internet slowdown", June 2015, <http://www.bgpmon.net/massive-route-leak-cause-internet-slowdown/>.

[Wijchers] Wijchers, B. and B. Overeinder, "Quantitative Analysis of BGP Route Leaks", RIPE-69, November 2014, <http://ripe69.ripe.net/presentations/157-RIPE-69-Routing-WG.pdf>.

[Zmijewski] Zmijewski, E., "Indonesia Hijacks the World", Dyn Research/Renesys Blog, April 2014, <http://research.dyn.com/2014/04/indonesia-hijacks-world/>.

Authors' Addresses

Kotikalapudi Sriram
US NIST

Email: ksriram@nist.gov

Doug Montgomery
US NIST

Email: dougm@nist.gov

Danny McPherson
Verisign, Inc.

Email: dmcpherson@verisign.com

Eric Osterweil
Verisign, Inc.

Email: eosterweil@verisign.com

Brian Dickson

Email: brian.peter.dickson@gmail.com

Network Working Group
Internet-Draft
Updates: 4012 (if approved)
Intended status: Standards Track
Expires: November 24, 2015

J. Snijders
NTT

May 23, 2015

The "import-via" and "export-via" attributes in RPSL Policy Specifications
draft-ietf-grow-rpsl-via-01

Abstract

This document defines two attributes in the aut-num Class which can be used in RPSL policy specifications to publish desired routing policy regarding non-adjacent networks.

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 24, 2015.

Table of Contents

1. Introduction

The Routing Policy Specification Language [RFC4012] allows operators to specify routing policies regarding directly adjacent networks through various import and export attributes. These attributes only apply to directly adjacent networks.

This document proposes to extend RPSL according to the following goals and requirements:

- Provide a way for network (A) to describe what an adjacent network (B) could use as routing policy towards its adjacent networks (C, D, E .. N).

- The extension should be backward compatible with minimal impact on existing tools and processes, following Section 10.2 of [RFC2622].

The addition of the "import-via" and "export-via" attributes in the aut-num Class will especially help participants of Multi-Lateral Peering services to inform the intermediate autonomous system what routing policy should be applied towards other participants.

2. Notational Conventions

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 [RFC2119].

3. Background

The via keyword specifically benefits operators who were assigned a 32 bit AS Number and transit providers when participating in Multi-Lateral Peering Agreements facilitated by a Route Server.

Often Route Server operators overload BGP Communities [RFC1997]) to facilitate signaling of desired routing policy between the participants and the Route Server. Because BGP Communities have a length of 32 bit, it is not possible to signal a 32 bit AS Number coupled with an action. In practise this means Route Server participants who use a 32 bit AS Number cannot specifically be included or excluded during path distribution calculations on the Route Server unless a mapping system is applied.

Transit providers often have a routing policy which states that the transit provider does not want to exchange paths with its downstream customers through Route Servers via public Internet Exchanges. The import-via and export-via attributes allow transit providers to participate in Multi-Lateral Peering services while instructing Route Server operators through a simple routing policy specification that paths should not be distributed to downstream customers and reducing the likelihood of Path Hiding on the Route Server.

4. Import and Export Via Syntax and Semantics

The two attributes can be used within the aut-num class.

import-via:

export-via:

The syntax for these attributes is as follows:

```
Attribute  Value                                                  Type
import-via [protocol <protocol-1>] [into <protocol-2>]   optional,
           [afi <afi-list>]                              multi-valued
           <mp-peering-1>
           from <mp-peering-2> [action <action-1>; ... <action-N>;]
           . . .
           <mp-peering-3>
           from <mp-peering-M> [action <action-1>; ... <action-N>;]
           accept <mp-filter> [;]

export-via [protocol <protocol-1>] [into <protocol-2>]   optional,
           [afi <afi-list>]                              multi-valued
           <mp-peering-1>
           to <mp-peering-2> [action <action-1>; ... <action-N>;]
           . . .
           <mp-peering-3>
           to <mp-peering-M> [action <action-1>; ... <action-N>;]
           announce <mp-filter> [;]
```

Figure 1

The import-via and export-via attributes are optional, and should be ignored by implementations which do not support interpretation of those attributes. The syntax closely mimics the mp-import and mp-export attributes described in Section 2.5 of [RFC4012], with the exception that before the "from" and "to" keywords an <mp-peering> is defined to indicate the common AS between two non-adjacent networks.

In the above example <peering-1> and <peering-3> are directly adjacent networks, for instance a Multi-Lateral Peering service. <peering-2> is a non-adjacent network.

5. Example Usage

Putting it all together:

```
aut-num: AS15562
import-via: AS6777
             from AS15562
             action pref = 2;
             accept AS-SNIJDERS
export-via: AS6777
             to AS15562 action community.={15562:40};
             announce AS-SNIJDERS
import-via: AS15562:AS-ROUTESERVERS
             from AS15562:AS-CUSTOMERS
             accept NOT ANY
export-via: AS15562:AS-ROUTESERVERS
             to AS15562:AS-CUSTOMERS announce NOT ANY
import-via: AS6777
             from AS4247483647
             accept AS4247483647
export-via: AS6777
             to AS4247483647 action community.={15562:40};
             announce AS-SNIJDERS
```

Figure 2

In the above examples AS15562 and AS15562 are Route Server participants. AS4247483647 is a participant who has been assigned a 32 bit AS Number. AS6777 functions as a Route Server [I-D.ietf-idr-ix-bgp-route-server] and AS-SET AS15562:AS-ROUTESERVERS contains a list of Route Server AS Numbers. AS-SET AS15562:AS-CUSTOMERS contains a list of downstream transit customers from AS15562.

The intention of the above policy would be to enable the exchange of NLRI's through AS6777 with two parties: AS15562 and AS4247483647, yet prevent the Route Server from distributing NLRI's announced by AS15562 towards customers of said network. Publishing the policy that AS15562 will not accept customer routes through the Route Server can help counteract the "path hiding" phenomenon as described in Section 2.3.1 of [I-D.ietf-idr-ix-bgp-route-server], as the Route Server now is informed which NLRI's should not be considered in the best path selection process.

6. Ambiguity Resolution

The same peering can be covered by more than one "via" policy attribute or by a combination of multi-protocol policy attributes, or multi-protocol policy attributes (when specifying IPv4 unicast policy) and the previously defined IPv4 unicast policy attributes.

In these cases, implementations should follow the specification-order rule as defined in Section 6.4 of RFC 2622 [1]. Operators should take note that in order to break ambiguity, the action corresponding to the first peering specification is used.

Consider the following example regarding ambiguity resolution.

```
aut-num: AS15562
export:     to AS6777 195.69.144.255 announce AS15562
export-via: AS6777 195.69.144.255 to AS-AMS-IX-RS announce AS-SNIJDERS
```

Figure 3

As both policy specifications cover the same peering with AS6777, specification-order rule is used and Route Server AS6777 195.69.144.255 should only accept AS15562, instead of AS-SNIJDERS, even though the export-via specification is more specific. If the intended policy was to announce all routes which can be resolved through AS-SNIJDERS on this particular peering, the operator should have specified:

```
aut-num: AS15562
export-via: AS6777 195.69.144.255 to AS-AMS-IX-RS announce AS-SNIJDERS
export:     to AS6777 195.69.144.255 announce AS15562
```

Figure 4

7. Security Considerations

There are no security considerations for this specification.

8. IANA Considerations

This document has no IANA actions.

9. Acknowledgments

The author would like to thank Remko van Mook for confirming that "via" is a better keyword than 'through' or 'thru', Nick Hilliard for his unparalleled support, Jeffrey Haas for providing historic perspective, David Croft and Martin Pels for nitpicking.

10. References

10.1. Normative References

[RFC1997] Chandrasekeran, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, August 1996.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2622] Alaettinoglu, C., Villamizar, C., Gerich, E., Kessens, D., Meyer, D., Bates, T., Karrenberg, D., and M. Terpstra, "Routing Policy Specification Language (RPSL)", RFC 2622, June 1999.

[RFC4012] Blunk, L., Damas, J., Parent, F., and A. Robachevsky, "Routing Policy Specification Language next generation (RPSLng)", RFC 4012, March 2005.

10.2. Informative References

[I-D.ietf-idr-ix-bgp-route-server] Jasinska, E., Hilliard, N., Raszuk, R., and N. Bakker, "Internet Exchange Route Server", draft-ietf-idr-ix-bgp-route-server-02 (work in progress), February 2013.

Appendix A. Grammar Rules

Note that only 'via' specific grammar rules have been listed. Currently two routing registry whois daemons have support for the 'via' attributes: IRRd 3.0.7 and RIPE Whois Server 1.71.

```
%type <string> line_autnum
%type <string> attr_autnum
%type <string> attr_import_via
%type <string> attr_export_via

%token T_IV_KEY // *** import-via: ***
%token T_EV_KEY // *** export-via: ***
%token T_AFI T_PROTOCOL T_WORD T_INTO T_EXCEPT T_REFINE
%token T_ACCEPT T_ANNOUNCE T_TO T_FROM T_PRNGNAME

line_autnum: attr_autnum
    | attr_import_via
    | attr_export_via

attr_import_via: T_IV_KEY attr_import_syntax

attr_import_syntax: opt_protocol_from opt_protocol_into import_simple
```

```
    | opt_protocol_from opt_protocol_into afi_import_exp

attr_export_via: T_EV_KEY attr_export_syntax

attr_export_syntax: opt_protocol_from opt_protocol_into export_simple
    | opt_protocol_from opt_protocol_into afi_export_exp

opt_afi_specification:
    | T_AFI afi_list

afi_list: afi_token
    | afi_list ',' afi_token

afi_token: afi_name

opt_protocol_from:
    | T_PROTOCOL T_WORD

opt_protocol_into:
    | T_INTO T_WORD

import_simple: opt_afi_specification import_factor

export_simple: opt_afi_specification export_factor

afi_import_exp: opt_afi_specification import_exp

afi_export_exp: opt_afi_specification export_exp

import_exp: import_term
    | import_term T_EXCEPT afi_import_exp
    | import_term T_REFINE afi_import_exp

import_factor_list: import_factor ';'
    | import_factor_list import_factor ';'

import_term: import_factor ';'
    | '{' import_factor_list '}'

export_exp: export_term
    | export_term T_EXCEPT afi_export_exp
    | export_term T_REFINE afi_export_exp

export_factor_list: export_factor ';'
    | export_factor_list export_factor ';'

export_term: export_factor ';'
    | '{' export_factor_list '}'
```

```
import_factor: import_peering_action_list T_ACCEPT filter

export_factor: export_peering_action_list T_ANNOUNCE filter

peering: as_expression opt_router_expression opt_router_expression_with_at
    | T_PRNGNAME

// Below are two grammar rules that actually differ
// from mp-import: + mp-export:

import_peering_action_list: peering T_FROM peering opt_action
    | import_peering_action_list peering T_FROM peering opt_action

export_peering_action_list: peering T_TO peering opt_action
    | export_peering_action_list peering T_TO peering opt_action
```

Figure 5

Appendix B. TODO

(RFC Editor - this Appendix can be removed upon publication as RFC)

1. Add python parser example based on Grako EBNF.

Appendix C. Document Change Log

(RFC Editor - this Appendix can be removed upon publication as RFC)

1. Initial document.
2. Changes to draft-snijders-rpsl-via-01.txt
    A. Moved from adding a new RPSL keyword to a new RPSL attribute to improve backwards compatibility.
3. Changes to draft-snijders-rpsl-via-02.txt
    A. Added grammar appendix.
    B. Added section about Ambiguity Resolution.
4. Changes to draft-snijders-rpsl-via-03.txt
    A. Updated current IRR implementations.
5. Changes to draft-grow-rpsl-via-01.txt

A. Bump version - add TODO.

Author's Address

Job Snijders
NTT
Theodorus Majofskistraat 100
Amsterdam 1065 SZ
NL

Email: job@ntt.net

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 30, 2016

T. King
C. Dietzel
DE-CIX Management GmbH
J. Snijders
NTT
G. Doering
SpaceNet AG
G. Hankins
Alcatel-Lucent

July 29, 2015

BLACKHOLE BGP Community for Blackholing
draft-ymbk-grow-blackholing-01

Abstract

This document describes the use of a well-known Border Gateway Protocol (BGP) community for blackholing at IP networks and Internet Exchange Points (IXP). This well-known advisory transitive BGP community, namely BLACKHOLE, allows an origin AS to specify that a neighboring IP network or IXP should blackhole a specific IP prefix.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in [RFC2119] only when they appear in all upper case. They may also appear in lower or mixed case as English words, without normative meaning.

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 January 30, 2016.

Table of Contents

1. Introduction

The network infrastructure has been getting hammered by DDoS attacks for years. In order to block DDoS attacks, IP networks have offered BGP blackholing to neighboring networks (iBGP scenarios [RFC3882] and RTBH filtering [RFC5635]), much like some IXPs have recently started to do.

DDoS attacks targeting a certain IP network may cause congestion of links used to connect to other networks. In order to limit the impact of such a scenario on legitimate traffic, IP networks and IXPs adopted a mechanism called BGP blackholing. A network that wants to trigger blackholing needs to understand the triggering mechanism adopted by its neighboring IP networks and IXPs. Different IP networks and IXPs provide different BGP mechanism to trigger

blackholing including pre-defined blackhole next- hop IP addresses and pre-defined BGP communities.

Having several different mechanisms to trigger blackholing at different IP networks and IXPs makes it an unnecessarily complex, error-prone and cumbersome task for network operators. Therefore a well-known BGP community [RFC1997] is defined for operational ease.

Having such a well-known BGP community for blackholing also supports IP networks and IXPs as

- o implementing and monitoring blackholing gets easier if implementation and operational guides do not cover many options to trigger blackholing
- o the amount of support requests from customers about how to trigger blackholing at a particular IP network or IXP will be reduced as the mechanism is unified

Making it considerably easier for network operators to utilize blackholing makes operations easier.

2. BLACKHOLE Attribute

This document defines the use a new well-known BGP transitive community, BLACKHOLE.

The semantics of this attribute is to allow a network to interpret the presence of this community as an advisory qualification to drop any traffic being sent towards this prefix.

3. Operational Recommendations

3.1. IP Prefix Announcements with BLACKHOLE Community Attached

When an IP network is under DDoS duress, it MAY announce an IP prefix covering the victim's IP address(es) for the purpose of signaling to neighboring IP networks or IXPs that any traffic destinated for these IP address(es) should be discarded. In such a scenario, the network operator SHOULD attach BLACKHOLE BGP community.

3.2. Local Scope of Blackholes

A BGP speaker receiving a BGP announcement tagged with the BLACKHOLE BGP community SHOULD add a NO_ADVERTISE, NO_EXPORT or similar communities to prevent propagation of this route outside the local AS.

Unintentional leaking of more specific IP prefixes to neighboring networks can have adverse effects. Extreme caution should be used when purposefully propagating IP prefixes tagged with the BLACKHOLE BGP community outside the local routing domain.

3.3. Accepting Blackholed IP Prefixes

It has been observed announcements of IP prefixes larger than /24 for IPv4 and /48 for IPv6 are usually not accepted on the Internet (see section 6.1.3 [RFC7454]). However, blackhole routes should be as small as possible in order to limit the impact of discarding traffic for adjacent IP space that is not under DDoS duress. Typically, the blackhole route's prefix length is as specific as /32 for IPv4 and /128 for IPv6.

BGP speakers SHOULD only accept and honor BGP announcements carrying the BLACKHOLE community if the announced prefix is covered by a shorter prefix for which the neighboring network is authorized to advertise.

3.4. IXPs: Peering at Route Servers

Many IXPs provide the so-called policy control feature as part of their route servers [I-D.ietf-idr-ix-bgp-route-server] (see e.g. the LINX website [1]). Policy control allows members to specify by using BGP communities which ASNs connected to the route server receive a particular BGP announcement.

Combined usage of the BGP communities for blackholing and policy control allows a fine-grained control of a blackhole.

In some implementations of blackholing at IXPs, the route server after receiving a BGP announcement tagged with the BLACKHOLE BGP community rewrites the next-hop IP address to the pre-defined blackholing IP address before redistributing the announcement.

4. IANA Considerations

The IANA is requested to register BLACKHOLE as a well-known BGP community with global significance:

BLACKHOLE (= 0xFFFF029A)

The low-order two octets in decimal are 666, amongst IP network operators a value commonly associated with BGP blackholing.

5. Security Considerations

BGP contains no specific mechanism to prevent the unauthorized modification of information by the forwarding agent. This allows routing information to be modified, removed, or false information to be added by forwarding agents. Recipients of routing information are not able to detect this modification. Also, RPKI [RFC6810] and BGPSec [I-D.ietf-sidr-bgpsec-overview] do not fully resolve this situation. For instance, BGP communities can still be added or altered by a forwarding agent even if RPKI and BGPSec are in place.

The BLACKHOLE BGP community does not alter this situation.

A new additional attack vector is introduced into BGP by using the BLACKHOLE BGP community: denial of service attacks for IP prefixes.

Unauthorized addition of the BLACKHOLE BGP community to an IP prefix by a forwarding agent may cause a denial of service attack based on denial of reachability. The denial of service will happen if an IP network or IXP offering blackholing is traversed. However, denial of service attack vectors to BGP are not new as the injection of false routing information is already possible.

In order to further limit the impact of unauthorized BGP announcements carrying the BLACKHOLE BGP community the receiving BGP speaker SHOULD verify by applying strict filtering (see section 6.2.1.1.2. [RFC7454]) that the peer announcing the prefix is authorized to do so. If not, the BGP announcement should be filtered out.

The presence of this BLACKHOLE BGP community may introduce a resource exhaustion attack to BGP speakers. If a BGP speaker receives many IP prefixes containing the BLACKHOLE BGP community its internal resources such as CPU power and/or memory might get consumed, especially if usual prefix sanity checks (e.g. such as IP prefix length or number of prefixes) are disabled (see Section 3.3).

6. References

6.1. Normative References

[RFC1997] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, DOI 10.17487/RFC1997, August 1996, <http://www.rfc-editor.org/info/rfc1997>.

[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>.

6.2. Informative References

[I-D.ietf-idr-ix-bgp-route-server] Jasinska, E., Hilliard, N., Raszuk, R., and N. Bakker, "Internet Exchange BGP Route Server", draft-ietf-idr-ix-bgp-route-server-07 (work in progress), June 2015.

[I-D.ietf-sidr-bgpsec-overview] Lepinski, M., "An Overview of BGPsec", draft-ietf-sidr-bgpsec-overview-07 (work in progress), June 2015.

[RFC3882] Turk, D., "Configuring BGP to Block Denial-of-Service Attacks", RFC 3882, DOI 10.17487/RFC3882, September 2004, <http://www.rfc-editor.org/info/rfc3882>.

[RFC5635] Kumari, W. and D. McPherson, "Remote Triggered Black Hole Filtering with Unicast Reverse Path Forwarding (uRPF)", RFC 5635, DOI 10.17487/RFC5635, August 2009, <http://www.rfc-editor.org/info/rfc5635>.

[RFC6810] Bush, R. and R. Austein, "The Resource Public Key Infrastructure (RPKI) to Router Protocol", RFC 6810, DOI 10.17487/RFC6810, January 2013, <http://www.rfc-editor.org/info/rfc6810>.

[RFC7454] Durand, J., Pepelnjak, I., and G. Doering, "BGP Operations and Security", BCP 194, RFC 7454, DOI 10.17487/RFC7454, February 2015, <http://www.rfc-editor.org/info/rfc7454>.

6.3. URIs

[1] https://www.linx.net/members/support/route-servers.html

Appendix A. Acknowledgements

The authors gratefully acknowledges the contributions of:

- Petr Jiran, NIX.CZ, Milesovska 1136/5, Praha 130 00, Czech Republic, Email: pj@nix.cz
- Yordan Kritski, NetIX Ltd., 3 Grigorii Gorbatenko Str., Sofia 1784, Bulgaria, Email: ykritski@netix.net
- Christian Seitz, STRATO AG, Pascalstr. 10, Berlin 10587, Germany, Email: seitz@strato.de

Authors' Addresses

Thomas King
DE-CIX Management GmbH
Lichtstrasse 43i
Cologne 50825
Germany

Email: thomas.king@de-cix.net

Christoph Dietzel
DE-CIX Management GmbH
Lichtstrasse 43i
Cologne 50825
Germany

Email: christoph.dietzel@de-cix.net

Job Snijders
NTT Communications, Inc.
Theodorus Majofskistraat 100
Amsterdam 1065 SZ
NL

Email: job@ntt.net

Gert Doering
SpaceNet AG
Joseph-Dollinger-Bogen 14
Munich 80807
Germany

Email: gert@space.net

Greg Hankins
Alcatel-Lucent
777 E. Middlefield Road
Mountain View, CA 94043
USA

Email: greg.hankins@alcatel-lucent.com