Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 24, 2015

R. Raszuk
Individual

July 23, 2014

BGP Path Record Attribute
draft-raszuk-idr-bgp-pr-00

Abstract

BGP protocol today contains number of build in mechanisms which record critical for its loop free operation data along the path of BGP message propagation. Those are encoded in AS_PATH, CLUSTER_LIST or ORIGINATOR_ID attributes. However in the same time there is no provisioning to record other useful information along the path which can be helpful to the operator in order to enhance end to end visibility of BGP control plane.

In order to solve this problem this document proposes a new single BGP attribute designed as an generic and extensible container to carry number of new optional information corresponding to the BGP speakers given BGP advertisement (or withdraw) message traverses.

Requirements Language

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

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Ability to record various information from the mid points given control plane packets traverses seems as a useful tool in number of control plane protocols in use today. For some protocols such information is critical and mandatory for correct operation while in other cases can be used as a set of data for further processing on or offline.

Recently there have been two proposals discussing need to carry opaque to BGP operation data in the two new BGP Attributes proposed for such purposes. As it seems that there can be much more types of such data it seems much more efficient to define a single TLV based placeholder to carry all optional hop parameters along the path given BGP prefix traverses in the network.

To facilitate such transport this document proposed a definition of new BGP Attribute called BGP Path Record Attribute which can be used to to carry such new optional information.

Furthermore the new attribute next to set of predefined types of data to be carried can also accommodate local to the particular autonomous system set of information to be added at each hop per operator's local configuration at given set of BGP speakers in the control plane path.

2. Protocol Extensions

This document describes a new BGP attribute known as BGP Path Record Attribute, along with definition of new TLV and number of sub-TLVs which can be used to carry new type of information either defined below in this document or already identified in other IETF proposals which are included here for illustration purposes only.

The TLV is defined for both easy extensibility of other then BGP speaker related data which may be attached to the advertisements, but also due to the fact that such grouped information will be added by each capable and participating BGP node as one entity. The TLVs are appended by each participating BGP speaker in the ordered fashion. When storing BGP Path Record attribute the TLVs which it contained upon reception from peer MUST not be reordered.

The sub-TLVs on the other hand allow for very easy definition of new types of data which may be required to be carried both within this document as well as by new subsequent documents.

2.1. BGP Path Record Attribute

The BGP Path Record attribute is a new BGP optional transitive attribute. The attribute type code for the Path Record attribute is to be assigned by IANA. The value field of the Path Record attribute is defined as a set of one or more Path Record TLVs along with their sub-TLVs.

2.2. BGP Hop TLV

A Path Record type 1 TLV called BGP Hop TLV within a Path Record Attribute is defined as follows:

BGP Hop - Type 1 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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |              TYPE             |            LENGTH             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                    4 OCTET BGP ROUTER ID                      |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                      4 OCTET AS NUMBER                        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                             FLAGS                             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    ~                            sub-TLVs                           ~
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: BGP Hop Type 1 TLV

TYPE: Two octets encoding the Path Record TLV Type. Type 1 TLV is called BGP Hop TLV and contains information pertaining to BGP speaker given prefix has been received at and is advertised further from.

LENGTH: Two octets encoding the length in octets of the Path Record TLV, excluding the type and length fields. The Length is encoded as an unsigned binary integer.

4 OCTET BGP ROUTER ID: 4 octet BGP router ID assigned to a given BGP speaker processing prefix with path containing BGP Path Record Attribute.

4 OCTET AS NUMBER: 4 octet AS number or zero padded 2 octet AS number of the autonomous system BGP Hop belongs to.

FLAGS: Number of boolean flags describing BGP Hop basic characteristics. The following flags are defined for BGP Hop TLV:

Bit 0:
      NH - Next Hop - Values: 0 - next hop not set; 1 - next hop set
Bit 1:

RR - Route Reflector - Values: 0 - not a route reflector; 1 - route reflector

Bit 2:

RS - Route Server - Values: 0 - not route server; 1 - route server

Bit 3:

B - Beacon prefix - Values: 0 - not a special beacon prefix; 1 - special beacon prefix

Bits 4-31:

Reserved for future use

sub-TLVs: Variable length sub-TLVs describing various information pertaining to the TLV they are nested under. General format of sub-TLV is illustrated below:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              TYPE             |            LENGTH             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                             VALUE                             ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: sub-TLV format

TYPE: Two octets encoding of the sub-TLV type used to differentiate information types carried in the context of given TLV.

LENGTH: Two octets encoding the length in octets of the sub-TLV, excluding the type and length fields. The Length is encoded as an unsigned binary integer.

VALUE: Variable length field described in the context of each sub-TLV.

The following optional sub-TLVs are proposed in this document to be carried under BGP Hop TLV:

o Type 1 - Host Name sub-TLV
o Type 2 - Time Stamp sub-TLV
o Type 3 - Next Hop record sub-TLV
o Type 4 - Path Count sub-TLV
o Type 5 - Origin Validation sub-TLV
o Type 6 - Geo-location sub-TLV

- o Type 7 - BGP System Load sub-TLV

2.2.1. Host Name sub-TLV

```
   Type:
         1
   Length:
         Variable (TBD ... if we want to limit the size).
   Value:
         UTF-8 encoded BGP speaker hostname.
   Description:
         Useful for enhance display in number of direct or indirect show
         commands and operational logs.
```

2.2.2. Time Stamp sub-TLV

```
   Type:
         2
   Length:
         10 octets
   Value:
         8 octets - BGP UPDATE message receive timestamp,
         1 octet - flags; Bit 0 - T flag (synchronized to external
         clock) Bits 1-7 - reserved.
         1 octet - Sync type as described in [RFC5905],
   Description:
         For full details of this sub-TLV use case and architecture are
         described in separate document:
         [I-D.litkowski-idr-bgp-timestamp]
```

2.2.3. Next hop record sub-TLV

```
   Type:
         3
   Length:
         4 octets or 16 octets
   Value:
         IPv4 or IPv6 address of the next hop changed by current BGP Hop
   Description:
         For full details of this sub-TLV use case and architecture are
         described in separate document:
         [I-D.zhang-idr-nexthop-path-record]
```

2.2.4. Path count sub-TLV

```
   Type:
         4
   Length:
```

```
         2 octets
   Value:
         Integer indicating number of paths present for a given prefix
         in BGP speaker.
   Description:
         Enables easy visibility in validation of expected propagation
         model for multiple paths of given prefix.  Can validate
         effectiveness of various BGP mechanisms (best-external, bgp
         diverse path, bgp add-paths etc ...) (TBD .. should we include
         how many paths are marked as stale ?)
```

2.2.5. Origin Validation sub-TLV

```
   Type:
         5
   Length:
         10 octets
   Value:
         8 octets - Last time BGP Origin Validation database has been
         updated.
         1 octet - flags; Bit 0 - T flag (synchronized to external
         clock) Bits 1-7 - reserved.
         1 octet - Sync type as described in [RFC5905],
   Description:
         Allows to easily detect issues associated with possible lack of
         synchronization of Origin Validation local database.
```

2.2.6. Geo-location sub-TLV

```
   Type:
         6
   Length:
         16 octets
   Value:
         Proposed encoding follows IETF consensus for representation of
         coordinate based location.
```

```
        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
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |  LatUnc   |                  Latitude                         +
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       | Lat (cont'd)  |  LongUnc  |               Longitude           +
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |    Longitude (cont'd)         | AType |   AltUnc  |  Altitude +
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |               Altitude (cont'd)               |Ver| Res |Datum|
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: Geo-location encoding format

Description:
Allows to map BGP control plane hops taken by BGP advertisements to two or three dimensional geo location coordinates of participating BGP speakers. Encoding details are described in [RFC6225] (DHCP geo-location options document).

2.2.7. BGP System Load sub-TLV

Type:
7

Length:
2 octets

Value:
1 octet - avg last 15 min of CPU utilization in percent by BGP process/thread
1 octet - avg last 15 min of BGP process memory use to total available memory use in percent

Description:
Used as indicator of possible CPU or memory problems of any given participating BGP speaker along the BGP control plane path.

3. Operation

The proposed new BGP Path Record attribute is an opaque entity for BGP operation and as such there is no requirement for any direct modification to BGP operation or BGP state machine based on the information it contains. It is expected that such feedback loop will be performed by operator either by automated or manual process.

Operator should be able to allow or deny origination of BGP Path Record attribute or insertion of any TLV or sub-TLV into BGP UPDATE message. It is however recommended that given BGP implementation

adds available sub-TLVs to BGP Hop TLV when particular prefix has been received with BGP Path Record Attribute already containing such sub-TLVs.

BGP policy should be enhanced to allow for easy filtering of BGP Path Record attribute both on egress as well as ingress eBGP sessions.

BGP Path Record attribute can be used within any AFI/SAFI.

4. Deployment considerations

It needs to be recognized that some sub-TLVs of BGP Hop TLV of BGP Path Record attribute can break update packing. Therefor it is strongly recommended that sub-TLVs type 2, 4, 7 as defined above are to be used only for specific beacon prefixes injected into BGP control plane by operator and flagged with the "B" bit within the originating BGP Hop TLV.

Beacon prefixes due to the store-and-forward nature of P2MP BGP distribution for information correctness should be carefully injected and withdrawn from entire network before subsequent injection is to take place again.

5. IANA Considerations

This document defines a new BGP attribute known as a BGP Path Record Attribute. The code point for a new BGP Path Record attribute has to be assigned by IANA from the BGP Path Attributes registry.

This document requests IANA to define and maintain a new registry named: "BGP Path Record Attribute TLV types". The reserved pool of 0x0000-0xFFFF has been defined for its allocations. The allocations policy is on a first come first served basis. The recommended allocation of 0x0001 is to be allocated for BGP Hop TLV.

This document requests IANA to define and maintain a new registry named: "BGP Hop sub-TLV types". The reserved pool of 0x0000-0xFFFF has been defined for its allocations. The allocations policy is on a first come first served basis. The recommended allocation of first 7 sub-TLVs are indicated in section 2.2 titled: BGP Hop TLV.

6. Security considerations

No new security issues are introduced to the BGP protocol by this specification.

7. Acknowledgements

Authors would like to acknowledge ... for their valuable input, review and comments.

8. References

8.1. Normative References

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

[RFC4223] Savola, P., "Reclassification of RFC 1863 to Historic", RFC 4223, October 2005.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

8.2. Informative References

[I-D.litkowski-idr-bgp-timestamp]
Litkowski, S., Patel, K., and J. Haas, "Timestamp support for BGP paths", draft-litkowski-idr-bgp-timestamp-00 (work in progress), July 2014.

[I-D.zhang-idr-nexthop-path-record]
Li, Z., Zhang, L., and S. Hares, "NEXTHOP_PATH_RECORD ATTIBUTE for BGP", draft-zhang-idr-nexthop-path-record-00 (work in progress), July 2014.

[RFC5905] Mills, D., Martin, J., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, June 2010.

[RFC6225] Polk, J., Linsner, M., Thomson, M., and B. Aboba, "Dynamic Host Configuration Protocol Options for Coordinate-Based Location Configuration Information", RFC 6225, July 2011.

Author's Address

Robert Raszuk
Individual

Email: robert@raszuk.net