Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 20 March 2025

C. Cardona
P. Lucente
NTT
P. Francois
INSA-Lyon
Y. Gu
Huawei
T. Graf
Swisscom

16 September 2024

BMP Extension for Path Status TLV
draft-ietf-grow-bmp-path-marking-tlv-02

Abstract

The BGP Monitoring Protocol (BMP) provides an interface for obtaining BGP Path information. BGP Path Information is conveyed within BMP Route Monitoring (RM) messages. This document proposes an extension to BMP to convey the status of a path after being processed by the BGP process. This extension makes use of the TLV mechanims described in draft-ietf-grow-bmp-tlv [I-D.ietf-grow-bmp-tlv] and draft-ietf-grow-bmp-tlv-ebit [I-D.ietf-grow-bmp-tlv-ebit].

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 BCP 14 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

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 https://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 20 March 2025.

Table of Contents

1. Introduction

For a given prefix, multiple paths with different path status, e.g., the "best-path", "back-up path", "invalid", and so on, may co-exist in the BGP RIBs after being processed by the BGP decision process. The path status information is currently not carried in the BGP Update Message RFC4271 [RFC4271] or in the BMP Update Message RFC7854 [RFC7854].

External systems can use the path status for various applications. The path status is commonly ~~checked~~ used by operators when performing ~~Troubleshooting~~troubleshooting or verify redundancy. Having such status stored in a centralized system ~~can enable~~enables the development of tools that facilitate this process.

~~Optimisation~~ Optimization systems can include the path status in their process,

and ~~also~~ use the status as a validation source (since it can compare the calculated state to the actual outcome of the network, such as primary and backup path). As a final example, path status information can complement other centralized sources of data, for example, flow collectors.

This document defines a so-called Path Status TLV to convey the BGP path status to the BMP server. The BMP Path Status TLV is carried in the BMP Route Monitoring (RM) Message.

2. Path Status TLV

This document defines two types of Path Status TLVs: ~~one is the~~ IANA-registered Path Status TLV~~,~~ and ~~the other is the~~ Enterprise-specific Path Status TLV.

2.1. IANA-registered Path Status TLV

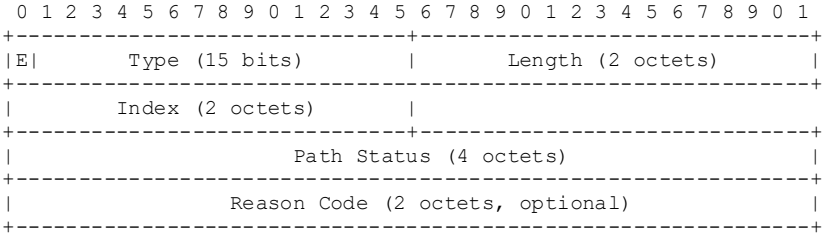

Figure 2: Encoding of IANA-Registered Path Status TLV

* E bit: For an IANA-registered TLV, the E bit MUST be set to 0 [I-D.ietf-grow-bmp-tlv-ebit].

* Type = TBD2 (15 Bits): indicates that it is the IANA-registered Path Status TLV.

* Length (2 Octets): indicates the length of the value field of the Path Status TLV. The value field further consists of the Path-Status field and Reason Code field.

* Index (2 Octets): indicates the prefix that this TLV is describing. Please see [I-D.ietf-grow-bmp-tlv] for details of the use of the index field to associate the path marking content with one or more NLRIs.

* Path Status (4 Octets): indicates the path status of the BGP Update PDU encapsulated in the RM Message. Currently 10 types of path status are defined, as shown in Table 1. All zeros are reserved.

* Reason Code (2 Octets, optional): indicates the reason of the path status indicated in the Path Status field. The reason code field is optional. If no reason code is carried, this field is empty. If a reason code is carried, the reason code is indicated by a 2-byte value, which is defined in Table 2.

Value	Path type
0x00000001	Invalid
0x00000002	Best
0x00000004	Non-selected
0x00000008	Primary
0x00000010	Backup
0x00000020	Non-installed
0x00000040	Best-external
0x00000080	Add-Path
0x00000100	Filtered in inbound policy
0x00000200	Filtered in outbound policy
0x00000400	Invalid ROV
0x00000800	Stale
0x00001000	Suppressed

Table 1: IANA-Registered Path Type

Figure 1

The Path Status field contains a bitmap where each bit encodes a specific role of the path. Multiple bits may be set when multiple path status ~~apply~~applies to a path.

* The best-path is defined in RFC4271 [RFC4271] and the best-external path is defined in draft-ietf-idr-best-external [I-D.ietf-idr-best-external].

* An invalid path is a route that does not enter the BGP decision process.

* A non-selected path is a route that is not selected in the BGP decision process. Back-up routes are considered non-selected, while the best and ECMP routes are not considered as non-selected.

* A primary path is a recursive or non-recursive path whose nexthop resolution ends with an adjacency draft-ietf-rtgwg-bgp-pic [I-D.ietf-rtgwg-bgp-pic]. A prefix can have more than one primary path ~~if~~ when multipath is configured draft-lapukhov-bgp-ecmp-considerations [I-D.lapukhov-bgp-ecmp-considerations]. A best-path is also considered as a primary path.

* A backup path is also installed in the RIB, but it is not used until some or all primary paths become unreachable. Backup paths are used for fast convergence in the event of primary path failures.

* A non-installed path refers to the route that is not installed into the IP routing table.

* For the advertisement of multiple paths for the same address prefix without the new paths implicitly replacing any previous ones, the add-path status is applied [RFC7911].

* Stale refers to a path which has been declared stale by the BGP Graceful Restart mechanism as described in Section 4.1 of [RFC4724].

* Suppressed refers to a path which has been declared suppressed by the BGP Route Flap Damping mechanism as described in Section 2.2 of [RFC2439].

The path status TLV does not force a BMP client to send any of these paths. It just provides a method to mark the paths that are available with their status.

Value	Reason code
[0x0001]	invalid for AS loop
[0x0002]	invalid for unresolvable nexthop
[0x0003]	not preferred for Local preference
[0x0004]	not preferred for AS Path Length
[0x0005]	not preferred for origin
[0x0006]	not preferred for MED
[0x0007]	not preferred for peer type
[0x0008]	not preferred for IGP cost
[0x0009]	not preferred for router ID
[0x000A]	not preferred for peer address
[0x000B]	not preferred for AIGP

Table 2: IANA-Registered Reason Code

Figure 2

2.2. Enterprise-specific Path Status TLV

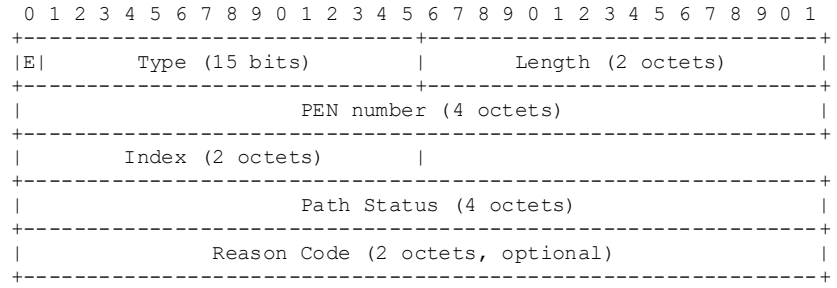

Figure 3: Encoding of Enterprise-specific Path Status TLV

* E bit: For an Enterprise-specific TLV, the E bit MUST be set to 1 [I-D.ietf-grow-bmp-tlv-ebit].

* Type = 1 (15 Bits): indicates that it's the Enterprise-specific Path Status TLV.

* Length (2 Octets): indicates the length of the value field of the Path Status TLV. The value field further consists of the Path-Status field and Reason Code field.

* Index (2 Octets): indicates the prefix that this TLV is describing. The index is the encapsulation order, starting from 0, of the prefix in the BGP Update PDU.

* PEN Number (4 octets): indicates the IANA enterprise number IANA-PEN.

* Path Status (4 Octets): indicates the enterprise-specific path status. The format is to be determined ~~w.r.t.~~with respect to each PEN number.

* Reason Code (2 octets, optional): indicates the reasons/ explanations of the path status indicated in the Path Status field. The format is to be determined ~~w.r.t.~~with respect to each PEN number.

3. Implementation notes

The BMP path marking TLV remains optional within BMP implementations.

An implementation of the BMP path marking TLV may not fully support marking of all status defined in table Figure 1 or any future extensions. Similarly, an implementation may choose to support the inclusion of the reason code (for which support is also optional), without necessarily incorporating any of the reason codes defined in table Figure 2 or future extensions.

This document refrains from defining mechanisms for signaling the status or reason codes an implementation supports. This could be established through external means (e.g. documentation) or potentially addressed in a subsequent document.

The remainder of this section encompasses additional points related to the implementation of the BMP Path marking TLV.

3.1. Configuration of BMP path marking

Implementations supporting the BMP path marking TLV SHOULD provide an option for enabling or disabling the Path Marking TLV over BMP sessions. Furthermore, the configuration options for this TLV SHOULD provide the means to enable and disable the transmission of reason codes, if the reason code are supported by the implementation.

Commented [TG1]: I suggest to add this capability in ietf-bmp.yang at draft-ietf-grow-bmp-yang. That leads to a normative reference in draft-ietf-grow-bmp-yang which is fine by me.

3.2. Paths with no markings

Some BGP routes might not require any type of status or reasons. For example, an unfiltered path obtained via the Adj-RIB-IN may fall under this category since there is really nothing to mark for that path. We suggest a couple of approaches for signaling that a path has no markings: (1) An implicit form of marking, achieved by abstaining from appending any BMP marking TLV pointing toward the route. (2) Alternatively, an explicit marking of the packet through a TLV containing no marked status and no associated reason code.

Commented [TG2]: This is not clear to me. Probably this meant to say that the best path decision process is not applied yet in this stage?

3.3. Significance of status and origin RIBs

This document refrains from imposing any implementation to mark specific status from specific RIBs. We recognize the diversity among implementations; some might be able to mark some status over one RIB while other do it on others. For instance, some might be able to mark Adj-RIB-in filtered routes when obtained from the Adj-RIB-IN pre-policy, while other could do it only from the Adj-RIB-IN post-policy. To remove ambiguities in implementations, we recommend the meaning of status (and reason codes) to not depend on the origin RIB of a route.

3.4. Enterprise-specific status and reasons

Implementations introducing their own status and reason codes are advised to adhere to [I-D.ietf-grow-bmp-tlv-ebit] and use ebit and vendor specific status and reasons. Additionally, we recommend all implementations to provide comprehensive documentation for these codes.

For scenarios where a path state combines a standard status with an enterprise-specific reason code (or vice versa), the following alternatives are presented:

* Replication of the standard definitions within the enterprise-specific space, thus permitting direct marking within the same packet using the ebit.
* Assigning two TLVs to the same path(s): one containing the standard part and another housing the vendor-specific part.

3.5. Multiple TLVs assigned to the same route.

We advocate for the employment of TLV grouping wherever feasible. The inclusion of all marking information within a single message is recommended, except on the case described in ~~section~~ Section 3.4. In situations where multiple TLVs are associated with a single route, all markings will be applicable to that route.

3.6. Reason Code Applicability

Reason codes starting with 'invalid' are most likely to be applied to path type 'Invalid'. Describing the reason why they are invalid. Where reason codes starting with 'not preferred' are most likely being used with path types which are not 'Best'. For example reason codes 'not preferred for AS Path Length' or 'not preferred for Local preference' are likely applied to path type 'Backup' to describe why they are considered 'Backup' and not 'Best'.

4. Acknowledgments

We would like to thank Jeff Haas and Maxence Younsi for their valuable comments.

5. IANA Considerations

This document requests that IANA assign the following new parameters to the BMP parameters name space.

Type = TBD1 (15 Bits): indicates that it is the IANA-registered Path Status TLV.

6. Security Considerations

It is not believed that this document adds any additional security considerations.

7. Normative References

[I-D.ietf-grow-bmp-tlv]
Lucente, P. and Y. Gu, "BMP v4: TLV support for BMP Route Monitoring and Peer Down Messages", Work in Progress, Internet-Draft, draft-ietf-grow-bmp-tlv-14, 18 March 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-grow-bmp-tlv-14>.

[I-D.ietf-grow-bmp-tlv-ebit]
Lucente, P. and Y. Gu, "Support for Enterprise-specific TLVs in the BGP Monitoring Protocol", Work in Progress, Internet-Draft, draft-ietf-grow-bmp-tlv-ebit-05, 18 March 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-grow-bmp-tlv-ebit-05>.

[I-D.ietf-idr-best-external]
Marques, P., Fernando, R., Chen, E., Mohapatra, P., and H. Gredler, "Advertisement of the best external route in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-best-external-05, 3 January 2012, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-best-external-05>.

[I-D.ietf-rtgwg-bgp-pic]
Bashandy, A., Filsfils, C., and P. Mohapatra, "BGP Prefix Independent Convergence", Work in Progress, Internet-Draft, draft-ietf-rtgwg-bgp-pic-21, 7 July 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-rtgwg-bgp-pic-21>.

[I-D.lapukhov-bgp-ecmp-considerations]
Lapukhov, P. and J. Tantsura, "Equal-Cost Multipath Considerations for BGP", Work in Progress, Internet-Draft, draft-lapukhov-bgp-ecmp-considerations-12, 28 December 2023, <https://datatracker.ietf.org/doc/html/draft-lapukhov-bgp-ecmp-considerations-12>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2439] Villamizar, C., Chandra, R., and R. Govindan, "BGP Route Flap Damping", RFC 2439, DOI 10.17487/RFC2439, November 1998, <https://www.rfc-editor.org/info/rfc2439>.

[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, <https://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, <https://www.rfc-editor.org/info/rfc4724>.

[RFC7854] Scudder, J., Ed., Fernando, R., and S. Stuart, "BGP Monitoring Protocol (BMP)", RFC 7854, DOI 10.17487/RFC7854, June 2016, <https://www.rfc-editor.org/info/rfc7854>.

[RFC7911] Walton, D., Retana, A., Chen, E., and J. Scudder, "Advertisement of Multiple Paths in BGP", RFC 7911, DOI 10.17487/RFC7911, July 2016, <https://www.rfc-editor.org/info/rfc7911>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

Authors' Addresses

Camilo Cardona
NTT
164-168, Carrer de Numancia
08029 Barcelona
Spain
Email: camilo@ntt.net

Paolo Lucente
NTT
Siriusdreef 70-72
2132 Hoofddorp
Netherlands
Email: paolo@ntt.net

Pierre Francois
INSA-Lyon
Lyon
France
Email: Pierre.Francois@insa-lyon.fr

Yunan Gu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: guyunan@huawei.com

Thomas Graf
Swisscom
Binzring 17
CH-8045 Zurich
Switzerland
Email: thomas.graf@swisscom.com