Inter-Domain Routing
Internet-Draft
Updates: 4271 (if approved)
Intended status: Standards Track
Expires: 19 August 2026

Z. Li, Ed.
L. Song, Ed.
China Mobile
15 February 2026

BGP SR Policy Extensions for Performance-Aware Path Selection
draft-li-idr-sr-policy-metric-03

Abstract

To enable the headend node to do performance-aware path selection, this document proposes an extension to the BGP SR Policy protocol by defining a new optional Metric Sub-TLV within the BGP Tunnel Encapsulation Attribute. The introduced Metric Sub-TLV encodes performance parameters (such as latency, bandwidth, reliability, etc.) for SR Policy paths.

This specification also updates the BGP route selection procedures in RFC4271, modifying the Breaking Ties (Phase 2) logic to prioritize the metrics for SR Policy paths.

Key contributions include:

* Introduce Metric Sub-TLV in BGP SR Policy

* Update the tie-breaking procedure for BGP route selection

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Segment Routing (SR) [RFC8402] allows a headend node to steer a packet flow along a specific path. [RFC9256] further details the concepts of SR Policy and steering into an SR Policy. [RFC9830] specifies the use of BGP to distribute one or more of the candidate paths of an SR Policy to the headend node of that policy. Currently [RFC9830] lacks the capability to propagate performance metrics such as path latency, bandwidth, or reliability. This limitation prevents headend nodes from implementing policy selection based on path metrics when there are multiple paths reaching the same destination. Consequently, the headend nodes cannot dynamically elect performance-optimal path among multiple SR Policies.

[I-D.ietf-idr-sr-policy-metric] is intended to address this problem. However, it carries metrics within segment lists, and a single segment list may contain more than one metric. When the candidate path of a SR Policy comprises multiple segment lists, the headend node faces difficulties in determining the metric for that SR Policy. This situation becomes even more complex when each segment list itself includes multiple metrics.

Furthermore, merely extending the SR Policy protocol with a metric attribute cannot resolve the current issue. This is because, before the BGP process on the headend node considers multiple alternative SR Policies, it has already determined the next hop of the BGP route in accordance with the existing Tie-breaking Procedure. Subsequently, only the SR Policies leading to the selected next hop may be utilized. Even if there exist more optimal SR Policies that can reach the destination via other next hops, these SR Policies will not be employed. Therefore, based on the content currently provided in [I-D.ietf-idr-sr-policy-metric], it is still impossible to ensure that the optimal SR Policy is selected.

To address these limitations, this document extends the BGP SR Policy protocol [RFC9830] to carry performance metrics in the candidate path of a SR Policy and updates the BGP Tie-breaking Procedure [RFC4271] to prioritize metrics for SR Policy paths.

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

2. Use Case

As illustrated in Figure 1, the SR Policy Computation and Provisioning System, such as a SDN controller, collects real-time network state information (e.g., topology, link bandwidth) and performance metrics (e.g., link latency, jitter, packet loss rate) via BGP-LS [RFC7752], TWAMP [RFC5357]/Micro TWAMP [RFC9533], and Telemetry [RFC9232] etc.. Based on service or customer requirements (e.g., minimum latency), the system computes SR Policy paths between designated endpoints and distributes these paths to the headend nodes via the BGP SR Policy protocol [RFC9830].

For example:

The system provisions two low-latency policies to headend node PE3:

Policy1: Path via P1-->PE1, with a measured latency of 20 ms.

Policy2: Path via P2-->PE2, with a measured latency of 12 ms.

However, the current BGP SR Policy protocol [RFC9830] only propagates path definitions (e.g., segment lists) without embedding performance metrics. This forces headend nodes to select paths based solely on static criteria (e.g., administrative preferences), potentially leading to suboptimal traffic engineering decisions.

To address this limitation, this proposal extends the BGP SR Policy protocol by introducing a new Performance Metrics Sub-TLV within the BGP Tunnel Encapsulation Attribute [RFC9012]. This Sub-TLV encodes key performance indicators (KPIs) such as latency, bandwidth, and reliability (see Section 3 for details). With this extension: The SR Policy Computation and Provisioning System can advertise SR Policies alongside their associated KPIs. Headend nodes leverage the enhanced BGP route selection logic (Section 6) to prioritize paths that meet dynamic performance requirements.

```
             +---------------------------------------------+
             |SR Policy Computation and Provisioning System|
             +---------------------------------------------+
                      *       *            *        *
                     *       *              *        *Extended BGP
                    *       *                *        *SR Policy
                   *       *                  *        *
               +---+  15ms*   +---+       5ms  *      +---+
              /|PE1| ----*----| P1|-------------*-----|PE3|\
             / +---+    *     +---+----\      /--*----+---+ \
            /   |      *        |       \    /    *     |    \
      +---+/    |     *         |        ----\     *    |     \+---+
      |CE1|0.5ms|    *          |0.5ms  /     \     *   |0.5ms |CE2|
      +---+\    |   *           |   ----       ---\  *  |     /+---+
            \   |  *            |  /1ms         5ms\  * |    /
             \ +---+          +---+                 --+---+ /
              \|PE2|----------|P2 |-------------------|PE4|/
               +---+   11ms   +---+       1ms         +---+
```

Figure 1: Use Case for Performance-Aware SR Policy Selection

3. Extensions to BGP SR Policy

This document extends the BGP SR Policy protocol [RFC9830] by introducing a new sub-TLV, Metric Sub-TLV, within the BGP Tunnel Encapsulation Attribute. The Extended BGP SR Policy Encoding structure is as follows:

```
SR Policy SAFI NLRl: <Distinguisher, Color, Endpoint>
Attributes:
  Tunnel Encapsulation Attribute(23)
    Tunnel Type: SR Policy(15)
      Binding SID
      Preference
      Priority
      Metric (This Document)
      SR Policy Name
      SR Policy Candidate Path Name
      Explicit NULL Label Policy (ENLP)
      Segment List
        Weight
        Segment
        Segment
        ...
      ...
```

Figure 2: Extended BGP SR Policy Encoding

Metric Sub-TLV is used to carry performance metrics such as latency, bandwidth, and reliability. The format of the Metric Sub-TLV is as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |    Length     |     Flags     |    Reserved   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Delay(8 octets, optional)                    |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Bandwidth(4 octets, optional)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             Reliability(4 octets, optional)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: Metric Sub-TLV

Where:

Type (1 octet): Indicates this sub-TLV is Metric, Specific values need to be assigned by IANA.

Length (1 octet): Indicates the length of the Metric Sub-TLV in bytes.

Flags (1 octet): Indicates the presence of specific performance metrics. Its definition is shown in Figure 4.

Reserved (1 octet): Reserved for future use. This field MUST be set to 0 when sending and ignored when receiving.

Delay(8 octets): Carries delay information. Its format depends on the D flag in the Flags Field:

* If D = 01: NTPv4 format delay

* If D = 10: IEEE 1588v2 PTP format delay

Bandwidth (4 octets): Carries bandwidth information in Mbps.

Reliability (4 octets): Carries reliability information, such as the maximum number of failures that have occurred on all links in the SR Policy path within the past month.

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
| D |B|R|Reserved|
+-+-+-+-+-+-+-+-+
```

Figure 4: Flags Field for Metric Sub-TLV

Where, D Flag is for delay, B Flag is for Bandwidth and R Flag is for reliability, all the other bits are reserved. The detailed encodings of the three flags defined in this document are as follows:

Flag	Bits	Description
D	0-1	00: No delay 01: NTPv4 delay 10: PTP delay 11: Reserved
B	2	0: No bandwidth 1: Bandwidth
R	3	0: No reliability 1: Reliability

Figure 5: Flags for Metric Sub-TLV

Implementations SHOULD set only one flag (D, B, or R) at a time, as these metrics are typically not directly comparable. Network operators MAY configure which metric to prioritize based on service requirements.

4. Policy Computation and Provisioning System Behavior

The Policy Computation and Provisioning System is responsible for calculating Segment Routing (SR) policies based on network state and business requirements, and provisioning them to headend nodes. When provisioning SR policies that include performance metrics, the system should follow these steps:

Collect Network State Information: Gather real-time network topology, link utilization, and other relevant data.

Compute SR Policies: Calculate SR Policy paths that meet performance requirements based on service needs and network state.

Encapsulate Performance Metrics: Embed performance metrics such as latency, bandwidth, and reliability within the Metric Sub-TLV of the BGP Tunnel Encapsulation Attribute.

Provision BGP Update Messages: Include the SR policies with performance metrics in BGP update messages and send them to the appropriate headend nodes.

5. Headend Node Behavior

Upon receiving SR policies with performance metrics, headend nodes should process them as follows:

Parse BGP Update Messages: Extract SR policies and their associated performance metrics from the received BGP update messages.

Store Performance Metrics: Save the performance metrics in a local database for subsequent path selection.

Path Selection: Prioritize paths that meet dynamic performance requirements when multiple paths are available.

Update Routing Tables: Modify routing tables based on the selected paths to ensure traffic is forwarded along optimized routes.

6. Updated Tie-Breaking Procedure for BGP

Support for SR Policy metric introduced in this document involves several modifications to the tie-breaking procedures of the BGP "phase 2" decision described in [RFC4271], Section 9.1.2.2.

A new step, step e0, is inserted before step e of the tie-breaking (Phase 2) logic in [RFC4271].

d) If at least one of the candidate routes was received via EBGP, remove from consideration all routes that were received via IBGP.

e0) If any routes have the Color extended community attribute with identical values, remove from consideration all routes lacking the Color extended community attribute.

Compare the SR policies corresponding to the remaining routes. If any SR policies have the Metric Sub-TLV, remove from consideration all routes whose corresponding SR policies lack the Metric Sub-TLV.

Remove from consideration all routes whose Metric Sub-TLV are not the best. For the latency metric, the smallest value is considered the best; for the bandwidth metric, the largest value is the best; and for reliability, the smallest value is the best.

e) Remove from consideration any routes with less-preferred interior cost.

7. IANA Considerations

IANA is requested to assign the following code point from the "BGP Tunnel Encapsulation Attribute Sub-TLVs" Registry:

Code Point	Description	Reference
TBD	Metric Sub-TLV	This document

Table 1: Code Point for Metric Sub-TLV

8. Security Considerations

The security considerations specified in [RFC9830] apply to this document. This document introduces no new security considerations.

9. References

9.1. Normative References

[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>.

[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>.

[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>.

[RFC9012] Patel, K., Van de Velde, G., Sangli, S., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", RFC 9012, DOI 10.17487/RFC9012, April 2021, <https://www.rfc-editor.org/info/rfc9012>.

[RFC9830] Previdi, S., Filsfils, C., Talaulikar, K., Ed., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", RFC 9830, DOI 10.17487/RFC9830, September 2025, <https://www.rfc-editor.org/info/rfc9830>.

9.2. Informative References

[I-D.ietf-idr-sr-policy-metric] KaZhang, Dong, J., and K. Talaulikar, "BGP SR Policy Extensions for metric", Work in Progress, Internet-Draft, draft-ietf-idr-sr-policy-metric-04, 5 February 2026, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-metric-04>.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", RFC 5357, DOI 10.17487/RFC5357, October 2008, <https://www.rfc-editor.org/info/rfc5357>.

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC9232] Song, H., Qin, F., Martinez-Julia, P., Ciavaglia, L., and A. Wang, "Network Telemetry Framework", RFC 9232, DOI 10.17487/RFC9232, May 2022, <https://www.rfc-editor.org/info/rfc9232>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

[RFC9533] Li, Z., Zhou, T., Guo, J., Mirsky, G., and R. Gandhi, "One-Way and Two-Way Active Measurement Protocol Extensions for Performance Measurement on a Link Aggregation Group", RFC 9533, DOI 10.17487/RFC9533, January 2024, <https://www.rfc-editor.org/info/rfc9533>.

Acknowledgements

The authors would like to acknowledge the support from Cheng Chang and Bo Liu.

Yao Liu from ZTE, Changwang Lin from H3C, Jie Dong from Huawei reviewed this document and provided valuable comments.

Authors’ Addresses

Zhenqiang Li (editor)
China Mobile
China
Email: lizhenqiang@chinamobile.com

Liyan Song (editor)
China Mobile
China
Email: songliyan@chinamobile.com